

# インタラクティブな情報提示システムのための 非装着・非拘束な視線推定

佐竹 純二<sup>†</sup> 小林 亮博<sup>†</sup> 川嶋 宏彰<sup>‡</sup> 平山 高嗣<sup>‡</sup> 水口 充<sup>†</sup> 小嶋 秀樹<sup>†</sup> 松山 隆司<sup>†,‡</sup>

<sup>†</sup> 情報通信研究機構 <sup>‡</sup> 京都大学

**概要** 本研究では、大型ディスプレイを用いた情報提示システムにおいて、ユーザの顔や視線の方向を推定し、ユーザの興味や反応を認識することで、コンテンツをインタラクティブに制御することを目指している。一般に、日常生活環境での視線推定を精度良く行うためには、立ち位置や顔向きの変化に対してロバストな推定が必要になるが、カメラの高解像度化だけでは十分には対応できない。そこで、1.カメラの首振り機構の導入による広範囲撮影、2.視点の異なる複数台カメラの導入による広範囲撮影と多様な顔向きへの対応を実現した。さらに、複数台カメラの推定結果を統合することで、情報提示システムに必要な十分な精度での視線推定を実現した。

## Non-contact and Non-constraint Gaze Estimation for Interactive Information Display

Junji SATAKE<sup>†</sup> Akihiko KOBAYASHI<sup>†</sup> Hiroaki KAWASHIMA<sup>‡</sup> Takatsugu HIRAYAMA<sup>‡</sup>  
Mitsuru MINAKUCHI<sup>†</sup> Hideki KOZIMA<sup>†</sup> Takashi MATSUYAMA<sup>†,‡</sup>

<sup>†</sup> National Institute of Information and Communications Technology <sup>‡</sup> Kyoto University

**Abstract** We develop an interactive information display that estimates face and gaze directions of a user in real-time, and interactively controls contents by recognizing the user's interest and reaction. In a real environment, it is necessary to estimate the gaze direction robustly for a variety of standing positions and face directions. In this paper, we report on the system that takes user's face images using pan-tilt cameras, and estimates gaze direction by the image processing. To estimate robustly, it integrates the results of multiple cameras.

### 1. はじめに

本研究では、ユーザの非言語情報（視線や頷きなど）からユーザの興味や反応を認識し、インタラクティブなやり取りを行う対話システムの開発を目指している。例えば、大型ディスプレイを用いた情報提示システムにおいて、ユーザの注意（顔向き、視線）や動作（頷き、首傾げ、指差しなど）からユーザの興味や反応を自動的に認識することで、ユーザの好みに合わせた旅行プランニングシステムやユーザの興味を先取りして説明するガイダンスシステムなどに応用できる。

これまでに数多くの視線推定手法[1-8]が提案されているが、推定誤差1度前後といった高精度な視線推定を実現するために、ユーザに計測機器を装着させたり、ユーザの顔を固定して高解像度な正面向きの目領域画像を用いるものが多かった。しかし、街角にシステムを設置した時、

訪れたユーザに計測機器を装着させたり、立ち位置を指示することは非現実的であると考えられる。そこで、本研究では実環境を想定し、設置したカメラの向きを制御することで自由な立ち位置での顔画像を取得し、画像処理によってユーザの視線（ディスプレイ上の注目場所）をリアルタイムかつ、情報提示システムに必要な十分な精度で推定することを目的とする。この時、大型ディスプレイの周囲に設置したカメラでは顔向きの変化によってセルフオクルージョンが発生しやすいため、複数台のカメラを用い、それぞれ独立に処理した推定結果を統合することで顔向きの変化にロバストな視線推定の実現を目指す。

### 2. システムの概要

#### 2.1 問題設定と必要条件

本研究では、大型ディスプレイを用いた情報

提示システムにおいて、ユーザの顔と視線の方向を推定することを目的とする。できる限り自然な状況を想定し、以下のような条件を設ける。

- ユーザに計測機器を装着させない。
- 立ち位置はユーザの自由とする。身長に個人差もある。
- カメラはディスプレイより外側に設置し、入力顔画像を正面向きに限定しない。
- オンラインで実時間推定を行う。

ただし、今回はユーザ数を一人に限定し、複数のユーザが同時にシステムの前に立つことは考えない。また、照明条件は一定とする。その他、ユーザの顔や体の動作速度・動作範囲などについても自然な状況を想定する。

なお、ユーザが50インチディスプレイから距離1mの場所に立った時、ディスプレイ上の注目場所を±10cmで求めるために必要な視線推定精度は±約6度である。これを本システムにおける現段階での目標精度とする。

## 2.2 従来の視線推定手法

視線推定の手法はこれまでも数多く提案されており、ユーザに計測機器を装着させるものや、赤外光源など非接触な機器を用いるものなどが存在したが、近年では処理ハードウェアの進歩と装置構成の手軽さから通常のビデオカメラのみを用いた視線推定手法が一般的になってきている。そして、カメラのみを用いた視線推定手法には、大きく分類して頭部姿勢を必要とするものと必要としないものの2つが存在する。

まず、ユーザに機器を装着させる手法としては、頭部に電極を取り付けて角膜-網膜の電位変化から眼球運動を検出するEOG法や、コイルを組み込んだコンタクトレンズを装着するサーチコイル法などが存在し、1度前後の高精度な測定が可能ではあるが、ユーザに大きな負担がかかってしまう。また、ヘルメットや眼鏡のような装着型のアイカメラを用いる手法も存在するが、同様にユーザに負担を強いる必要がある。

非接触な視線推定手法としては、近赤外の点光源で目を照射し、角膜で反射されたプルキニエ像と瞳孔の位置から視線方向を推定する角膜反射法が一般的によく用いられており、据え置き型の計測装置も市販されている。Hennesseyら[1]の手法では単眼カメラ（焦点距離 $f=32\text{mm}$ 、解像度 $1024\times 768$ ）と複数光源を用い、推定誤差が1度、頭部の可動範囲は $W14\times H12\times D20\text{cm}$ である。また、FreeGaze[2]でも単眼カメラ（ $640\times 480$ ）と単一光源を用い、推定誤差が1度、頭部可動範囲は $W4\times H4\times D1\text{cm}$ （カメラとの距離は $60\text{cm}$ ）

である。このように角膜反射法は高精度な計測が可能ではあるが、高解像度な目領域画像を必要とし、光源距離やカメラ画角などから頭部位置が制限されるため、ユーザがディスプレイの前に座った状態を想定したものが多い。さらに、コンタクトレンズや眼鏡によって推定精度が低下すること、赤外照明装置が必要であること、太陽光の影響を受けることなどの問題点もある。同様に、角膜（黒目）と強膜（白目）の反射率の差を利用したリンバストラッキング法（強膜反射法）も赤外光を必要とし、上下方向の推定精度が低いという欠点もある。

頭部姿勢を用いずにカメラ画像から視線推定を行う手法では、高解像度な目領域画像を使用する。one-circle法[3]は虹彩の楕円形状から視線方向を推定する手法であり、精度1度を実現するためには虹彩の輪郭を約350pixel必要とし、目の幅180pixel、虹彩直径70pixel程度の高解像度画像（ $f=55\text{mm}$ 、 $640\times 480$ ）を使用している。画角が狭いために頭部の動きが制限され、正面向きでは虹彩形状が円に近くなるために楕円パラメータを用いた推定精度が低下するという欠点もある。また、ノイズにロバストな虹彩検出法としてLmedSを用いた方法[4]（虹彩直径60pixel程度）が提案されており、候補点の50%が正しければ楕円推定が可能である。ハフ変換を用いた虹彩検出手法も存在するが、楕円の検出には計算コストがかかり、現段階では実時間処理に至っていない。その他、個人差を考慮した高精細なアイモデル[5]や、正面向き画像限定ではあるが手動で設定が必要であったその個人性パラメータを自動学習する手法[6]なども提案されている。

画像から頭部姿勢を用いて視線推定を行う手法として、Zelinskyら[7]は2台のステレオ画像を用いることで、低解像度な画像（目の幅30pixel、虹彩は直径10pixel程度の円）から3.5度の精度で視線を推定している。同様に、松本ら[8]も虹彩検出に二値化とハフ変換を用い、 $320\times 240$ のステレオ画像から誤差5度程度で視線を推定している。これらの方法ではまず頭部姿勢から眼球中心位置を求めており、頭部姿勢を推定する手法にはパーティクルフィルタを用いた手法[9]や、AAMを基にした3次元顔モデル[10]、表情変化と個人差を分離した多重線形顔形状モデル[11]など、数多くの手法が提案されている。また、山添ら[12]は任意方向を注目した複数枚の顔画像から不特定ユーザの顔と眼球の3次元形状モデルを作成し、単眼画像（ $320\times 240$ ）から誤差約5度で視線を推定する手法を提案している。以上

の方法は眼球半径が小さいために誤差の影響が大きく、また、元々人間の眼球が1点で回転する構造になっていないなどの問題もあり、まだ高い精度は得られていない。しかし、装置構成の手軽さやユーザに負担をかけないなどの大きな利点があり、また、ユーザが何を見ているかを推定するには頭部や目の3次元位置情報が必要であるため、本研究ではこの方法を採用する。

### 2.3 システム構成

試作したシステムの外観を図1に示す。50インチプラズマディスプレイの右、左、下に計3台のカメラを設置し、左右のカメラにはパン・チルト制御機構を設けた(図2)。使用したカメラはPoint Grey Research社製のDragonfly2 (1024×768, 30fps, 8bitグレー画像, 1/3インチCCD)であり、レンズはFUJINON社製のHF12.5HA-1B (f=12.5mm)とHF9HA-1B (f=9mm)を用いた。撮影画像の例を図3に示す。また、虹彩を検出するには照度が不十分だったため、ディスプレイの両脇に写真電気工業社製のライトRIFA-F (50×50cm)を配置した。

現在のシステムのモジュール構成を図4に示す。各カメラはそれぞれ別のPC(視線推定サーバ)に接続し、顔の向きによって奥側の目や眉が隠れるなどのセルフオクルージョンが生じやすいため、単眼で顔向き(頭部姿勢)と視線の推定を行う。そして、統合サーバで情報を統合し、インタラクションサーバがユーザの興味を認識しながらコンテンツの制御を行う。なお、将来的には複数のカメラを用いて複数のユーザにそれぞれカメラを割り当てることや、レンズの異なるカメラを併用して高解像度な目領域画像を獲得することなどを検討している。

## 3. 認識処理の内容

### 3.1 各座標系の定義

各座標系を図5のように定義する。座標系は全て右手系とし、ワールド座標系の原点をディスプレイの中央正面1mの所に置く。ディスプレイの法線方向を $Z_w$ とし、水平方向に $X_w$ 、垂直方向に $Y_w$ を定義する。各カメラ座標系はレンズ中心を原点、光軸方向を $Z_c$ とする。また、ワールド座標系と画像座標系の関係は以下の式で与えられる。

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$



図1: 試作システムの外観



図2: パン・チルト制御カメラ



(a) 右カメラ f=12.5mm (b) 下カメラ f=9mm

図3: 撮影画像例

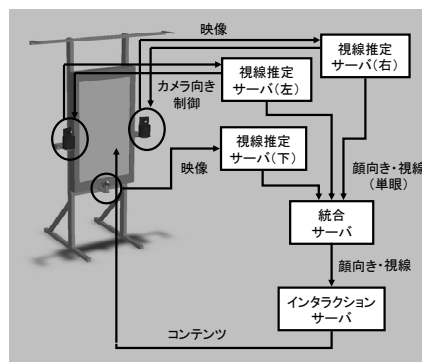


図4: モジュール構成図

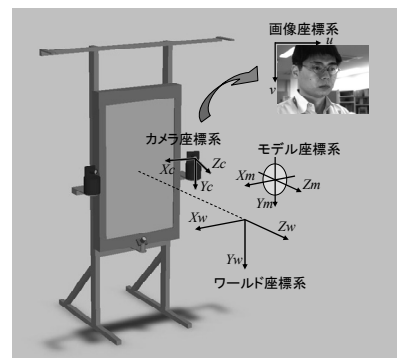


図5: 各座標系の定義

ただし、 $P$  は射影行列であり、カメラキャリブレーションとレンズ歪みの補正は OpenCV の関数を用いて行った。また、カメラ向きを制御した場合には、その都度、回転量  $pan$ ,  $tilt$  に応じて以下のように射影行列を算出する。

$$P = AR_x(tilt)R_y(pan)R_x(-tilt_0)[R \ T]$$

ここで、 $A$ ,  $R$ ,  $T$  はカメラキャリブレーションを行った際（基準向き）のカメラパラメータ行列であり、 $R_x$ ,  $R_y$  はそれぞれ  $x$  軸,  $y$  軸まわりの回転を表す行列である。また、基準向きを  $pan=0$ ,  $tilt=tilt_0$ , カメラが制御部に垂直な状態を  $tilt=0$  とし、回転中心とカメラの投影中心はほぼ一致していると仮定している。

なお、ユーザの移動に合わせて動的にカメラ向きを制御するにはまだ撮影との同期や向き制御の応答速度が不十分であるため今後の課題とし、今回は手動で指定したタイミングでのみ、複数カメラを用いて求めたユーザの顔位置を統合サーバから受け取り、その方向にカメラを向けるように実装した。

### 3.2 特徴点抽出

顔の特徴点の抽出には Active Appearance Model (AAM)[13,14]を用いる。AAM では顔部品の輪郭を抽出することができ、将来は眉や口の輪郭を用いた表情の認識や、発話タイミングの検出などにも利用できると考えられる。

AAM は、顔向きや表情の変化における特徴点の位置関係と、その際の照明の影響を固有空間で表現するモデルである。このモデルは各特徴点の  $u, v$  座標を並べたものを主成分分析により圧縮したベクトル  $b_s$  と、特徴点を頂点とする三角領域内の画素の輝度値を並べたものを主成分分析により圧縮したベクトル  $b_g$  を組み合わせ、再度、主成分分析することで作成される。

$$b = \begin{bmatrix} Wb_s \\ b_g \end{bmatrix}$$

ここで、 $W$  は重みを表す対角行列である。今回使用した学習モデルは、複数の人物について、ディスプレイ上の色々な場所を見ている複数枚の画像を用い、特徴点には両目、両眉、鼻、口の輪郭、計 45 点を用いた。作成した AAM 学習モデルの例を図 6 に示す。

### 3.3 顔向き推定

顔向きのパラメータ  $\xi$  は以下の 6 変数（回転 3 自由度、平行移動 3 自由度）で定義する。

$$\xi = [\phi \ \theta \ \psi \ t_x \ t_y \ t_z]^T$$

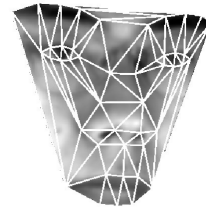


図 6： AAM 学習モデルの例

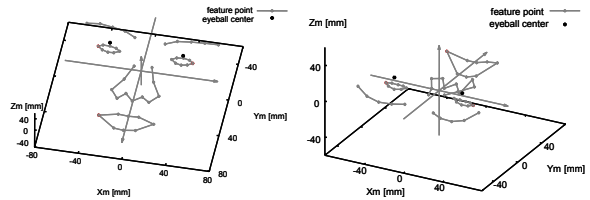


図 7： 3次元顔形状モデル

この時、 $i$  番目の特徴点のモデル座標  $(X_{mi}, Y_{mi}, Z_{mi})$  からワールド座標  $(X_{wi}, Y_{wi}, Z_{wi})$  への変換は以下の式で与えられる。

$$\begin{bmatrix} X_{wi} \\ Y_{wi} \\ Z_{wi} \end{bmatrix} = R_z(\phi)R_y(\theta)R_x(\psi) \begin{bmatrix} X_{mi} \\ Y_{mi} \\ Z_{mi} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$$

顔向きの推定には最急降下法を用い、評価式  $f$  は AAM で抽出した特徴点座標  $(u_i, v_i)$  と、顔向き  $\xi$  から再投影した座標  $(u_i(\xi), v_i(\xi))$  との誤差で定義する。

$$\xi_{k+1} = \xi_k - \alpha \cdot \frac{\partial f}{\partial \xi}(\xi_k)$$

$$f(\xi) = \sum_i w_i \sqrt{\{u_i(\xi) - u_i\}^2 + \{v_i(\xi) - v_i\}^2}$$

ここで、 $w_i$  は各特徴点の重みであり、目頭や鼻など、表情変動の影響を受けにくい特徴点の重みを大きく設定した。

なお、前処理として、3次元顔形状モデル（図 7）をステレオ情報によりあらかじめ作成しておき、ディスプレイ中央を見ている状態を基準姿勢とする。また、ディスプレイ上に表示したマーカーを注視した画像を用い、ステレオ情報によって求めた虹彩の 3次元位置から、顔形状モデル中の眼球中心位置も学習しておく。

### 3.4 虹彩検出

虹彩の検出には、単純な楕円探索を用いる。まず、3.2 で求めた目領域に対して大津の二値化[15]を行うことで、白色領域・黒色領域・目領域外の三つに分類する（図 8(b)）。次に、虹彩を直径 11mm の円、視線と顔の向きは近いものと仮定し、3.3 で求めた顔向きをもとに画像上

での虹彩の大きさと形状を算出する(図8(c)). 内側の黒い楕円領域Bが算出した虹彩の形状であり, 外側の白い領域Wは一回り大きな楕円との中間領域である(経験的に虹彩の1.2倍とした). この楕円テンプレートを三値画像に重ね合わせ, 領域B内の各画素  $i=1, \dots, N_B$  について黒色領域を+, 白色領域を-でカウントし, また, 領域W内の各画素  $j=1, \dots, N_W$  について白色領域を+, 黒色領域を-でカウントする. そして, 三値画像上を探索して最も評価値  $E_{iris}$  の高い位置を虹彩中心とする.

$$v_{B_i} = \begin{cases} 1 & \text{(黒色領域)} \\ -1 & \text{(白色領域)} \\ 0 & \text{(目領域外)} \end{cases}, v_{W_j} = \begin{cases} 1 & \text{(白色領域)} \\ -1 & \text{(黒色領域)} \\ 0 & \text{(目領域外)} \end{cases}$$

$$E_{iris} = \frac{1}{N_B + N_W} \left( \sum_{i=1}^{N_B} v_{B_i} + \sum_{j=1}^{N_W} v_{W_j} \right)$$

### 3.5 視線推定

視線は眼球中心と虹彩中心を結ぶ直線と定義し, 視線とディスプレイ平面の交点を注目座標として求める. 眼球中心の3次元座標  $(X_{center}, Y_{center}, Z_{center})$  は, 3.3で推定した顔向きから求められる. 虹彩中心の3次元座標  $(X_{iris}, Y_{iris}, Z_{iris})$  は, 眼球半径  $r=13\text{mm}$  と仮定し, 虹彩の画像上座標  $(u_{iris}, v_{iris})$  と眼球中心の3次元座標から以下の連立方程式を解くことで求められる(図9参照).

$$\begin{cases} \lambda \begin{bmatrix} u_{iris} \\ v_{iris} \\ 1 \end{bmatrix} = \mathbf{P} \cdot \begin{bmatrix} X_{iris} \\ Y_{iris} \\ Z_{iris} \\ 1 \end{bmatrix} \\ \begin{bmatrix} X_{iris} \\ Y_{iris} \\ Z_{iris} \end{bmatrix} - \begin{bmatrix} X_{center} \\ Y_{center} \\ Z_{center} \end{bmatrix} = r \end{cases}$$

ただし, 解が2つ求まるため, ディスプレイに近い方を正解の虹彩中心  $(X_{iris}, Y_{iris}, Z_{iris})$  とする. そして, ディスプレイ上の注目座標  $X_{gaze}, Y_{gaze}$  は, それぞれのカメラ, 左右それぞれの目について, 以下の式から求められる.

$$\begin{bmatrix} X_{gaze} \\ Y_{gaze} \\ Z_{gaze} \end{bmatrix} - \begin{bmatrix} X_{center} \\ Y_{center} \\ Z_{center} \end{bmatrix} = \lambda \left( \begin{bmatrix} X_{iris} \\ Y_{iris} \\ Z_{iris} \end{bmatrix} - \begin{bmatrix} X_{center} \\ Y_{center} \\ Z_{center} \end{bmatrix} \right)$$

ただし, 現在のシステムではワールド座標系の原点をディスプレイの前方1mに置いているため,  $Z_{gaze} = -1000$  としている.



(a) 入力 (b) 三値化 (c) 探索形状

図8: 虹彩検出

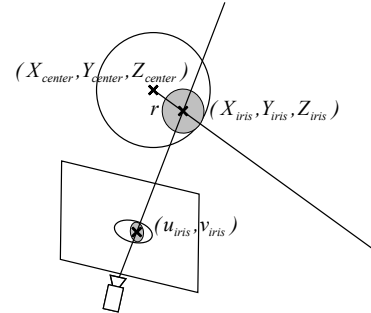


図9: 虹彩中心と視線の単眼推定

### 3.6 単眼視線推定結果の統合

視線推定サーバはそれぞれ単眼で推定した視線情報を統合サーバに送り, 統合サーバでそれらの情報を統合する. 統合サーバは合計6つの視線情報(カメラ3台×左右の目)の内, 顔の検出に失敗したものや虹彩検出の評価が低いものを除き, 残った視線情報の重み付き平均によってディスプレイ上の注目座標  $\bar{X}_{gaze}, \bar{Y}_{gaze}$  を求める. 今回の実装では, この重みには虹彩検出の評価値  $E_{iris}$  を使い, 虹彩検出成功の閾値を経験的に  $E_{iris} > 0.4$  とした.

$$\bar{X}_{gaze} = \frac{\sum E_{iris} X_{gaze}}{\sum E_{iris}}, \bar{Y}_{gaze} = \frac{\sum E_{iris} Y_{gaze}}{\sum E_{iris}}$$

なお, 統合サーバは視線情報だけでなく, 顔向きのパラメータなども管理し, カメラ向きの制御の際には複数カメラを用いて求めた顔位置情報を各視線推定サーバに伝える.

## 4. 実験

### 4.1 頭部可動範囲とカメラ向き制御精度

ディスプレイからの距離が約1mの時の頭部可動範囲は, カメラ向き固定の場合では左右カメラ W30×H20×D30cm, 下カメラ W40×H30×D25cm であり, 画角が範囲制限の原因である. 一方, カメラ向きを制御した場合には W100×H50×D40cm (距離90cm~1.3m) の範囲において1台以上のカメラで顔向きと視線を推定することができた. 自然な状況での立ち位置や身長個人差にほぼ対応できていると考えられる.

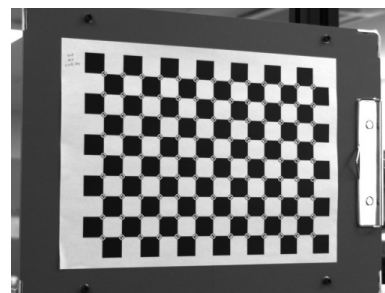
この場合の範囲制限の主な原因は、顔の向きや照明との位置関係によって顔の見え方が変化し、特徴点抽出が不安定になってしまうことである。特徴点抽出を安定化することで、頭部可動範囲を広げる事ができる。ただし、カメラの向きを自動制御するためには頭部位置を求める必要がある、より広角なカメラを追加する必要がある。

カメラ向き制御の精度計測の様子を図 10 に示す。ボードの格子の大きさは 18.8mm であり、灰色の円が計算で求めた射影行列を用いて格子位置を再投影したものである。カメラ向きを下と左に 5 度ずつ回転させた時の投影誤差は 2mm であり、原因は回転中心と投影中心のずれやギヤの遊びによるものである。この誤差の影響は、推定したディスプレイ上の注目場所が 2mm ずれるだけであり、顔向きや視線の推定誤差に比べて小さく、無視できると考えられる。なお、4.2, 4.3 の精度評価ではこのカメラ向き制御による誤差の影響を排除するため、カメラの向きを固定してカメラキャリブレーションを行い、その向きで撮影した動画像を用いてオフライン処理を行っている。

#### 4.2 顔向き推定精度

顔向きの推定精度を計測するため、図 11(a) のように顔の側面（眼鏡）にレーザーポインタを固定し、ポインタがディスプレイ上に表示したマーカ（図 11(b)）を指すように顔を動かした動画像を用いて顔向き推定の誤差を計測した。ただし、レーザーポインタが顔の正面を向いていないため、この実験では事前にディスプレイ中央にポインタを向けた画像を用いて基準の顔向きを学習し直している。なお、3 次元磁気センサ（FASTRACK）を用いて顔向きの正解値を求める方法も試してみたが、プラズマディスプレイなどの影響により正しい値を得る事はできなかった。

既知の人物 A, 未知の人物 B の 2 名について、ディスプレイ上に表示した 15 個のマーカへ順に顔を向けた動画像を用いて顔向き推定を行った。各マーカにおける推定誤差（各 1 フレーム）を表 1 に示す。人物 A の推定には AAM, 顔形状とともに本人の学習用動画像から作成したモデルを用いており、ノイズ（表情変化を含む）の影響により 60mm 程度の推定誤差が生じている。人物 B の推定には本人を含まない 5 人の動画像から作成した AAM 学習モデルと、人物 A の顔形状モデルを使用しており、100mm 前後の推定誤差が生じている。また、表中の下線は、3 台のカメラの内、特徴点抽出の評価値が最も良か



(a) 基準時

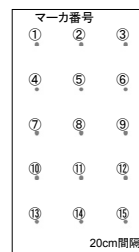


(b) 下 5 度, 左 5 度

図 10: カメラ向き制御の精度測定



(a) 実験の様子



(b) マーカ情報

図 11: 顔向き推定の精度測定

表 1: 顔向き推定の誤差 [mm]

マーカ番号	人物A (学習モデル本人)			人物B (未知のユーザ)		
	右	左	下	右	左	下
①	98.13	<u>93.94</u>	85.53	116.44	<u>142.49</u>	50.71
②	71.96	<u>90.10</u>	44.35	108.57	149.55	<u>19.55</u>
③	<u>74.70</u>	82.82	33.41	<u>97.29</u>	210.52	32.60
④	70.85	<u>47.88</u>	27.15	101.43	86.52	<u>38.67</u>
⑤	11.41	33.79	<u>31.93</u>	39.58	127.13	<u>23.47</u>
⑥	<u>45.08</u>	61.28	25.00	<u>57.57</u>	204.33	32.33
⑦	54.75	<u>10.94</u>	8.57	104.70	35.81	<u>8.64</u>
⑧	18.08	16.44	<u>6.54</u>	30.57	17.44	<u>18.00</u>
⑨	71.90	57.42	<u>30.92</u>	31.72	150.54	<u>27.24</u>
⑩	64.19	<u>4.23</u>	28.55	175.77	79.15	<u>25.54</u>
⑪	41.26	37.57	<u>40.65</u>	125.20	96.58	<u>6.65</u>
⑫	106.03	126.49	<u>59.93</u>	62.14	389.58	51.51
⑬	64.88	<u>34.52</u>	57.78	180.74	<u>164.28</u>	66.83
⑭	77.11	65.40	<u>85.90</u>	124.73	158.67	<u>43.79</u>
⑮	127.97	126.42	<u>75.39</u>	<u>157.71</u>	207.01	125.90
平均	66.55	59.28	42.77	100.94	147.97	38.09

ったものであり、正面向きに近いカメラの評価値が良くなる傾向が見られる。しかし、評価値最良のカメラが必ずしも推定誤差最小にはなっ

ておらず、表情変動や顔形状の個人差、眼鏡の影響やオクルージョン、曲線部分に対する特徴点抽出や顔向き推定の不安定性などによるものと考えられ、今後、より安定な評価値の算出方法を検討する必要がある。

### 4.3 視線推定精度

4.2と同様に既知の人物A、未知の人物Bの2名について、顔向きと視線を自由に動かし、自然な状態でマーカを注視した動画像を用いて視線推定を行った。

まず、各マーカを注目している連続50フレーム中、虹彩の検出に成功したフレーム数を表2、表3に示す。顔の向きによってセルフオクルージョンが生じるため、特に右カメラにおける左目、左カメラにおける右目、下を向いた時の左右カメラ、上を向いた時の下カメラで虹彩検出に失敗しやすい傾向があることが確認できた。一方、3台のカメラの推定結果を統合した場合には、まばたきによる失敗(人物Bのマーカ⑤)を除き、全てのフレームをカバーできている。

次に、各マーカ50フレームの推定誤差の平均を表4に示す。それぞれのカメラでの推定誤差は既知の人物Aが118~154mm、未知の人物Bが141~183mmであったのに対し、3台の推定結果を統合した場合の平均誤差は人物Aが96mm、人物Bが116mmに抑えられていることが確認できた。図12のように各カメラの推定結果が散らばっている場合でも、統合処理によって正面向きに近いカメラがうまく選択され、より正解に近い値が得られている。

1フレーム当たりの処理時間は、Intel Core2 Duo 2.66GHzのPCで、特徴点抽出が20ms、顔向き推定が5ms、虹彩検出と視線推定が5msであり、30fpsを実現できている。

### 4.4 考察

カメラ向きを制御することで立ち位置の違いや身長の個人差に対応でき、また、3台のカメラの推定結果を統合することで顔向きの変化にロバストな推定を実現できていることが確認できた。現在の推定精度はディスプレイ上で誤差100mm程度であり、ユーザがディスプレイのどの辺りを見ているのかを推定することができる。例えば、図2のような4つのコンテンツの中からどれに興味を持っているのかを判断し、類似するコンテンツを提案したり、より詳細な情報を提示するなどの利用が考えられる。

現在のシステムでは、右カメラ(図3参照)の虹彩の直径が約33pixelであり、虹彩直径が11mm、眼球半径が13mmの時、虹彩検出の誤

表2: 虹彩検出の成功フレーム数(人物A)

マーカ番号	右カメラ			左カメラ			下カメラ			統合
	右目	左目	統合	右目	左目	統合	右目	左目	統合	
①	50	50	50	50	50	50	50	50	50	50
②	50	50	50	50	50	50	37	49	49	50
③	50	50	50	2	50	50	49	50	50	50
④	50	41	50	50	50	50	50	34	50	50
⑤	50	50	50	50	50	50	40	42	49	50
⑥	50	50	50	0	50	50	26	49	49	50
⑦	50	4	50	50	50	50	50	50	50	50
⑧	23	0	23	50	50	50	35	47	48	50
⑨	50	50	50	0	50	50	31	50	50	50
⑩	50	0	50	44	50	50	49	50	50	50
⑪	32	50	50	19	50	50	48	50	50	50
⑫	48	47	50	0	16	16	48	50	50	50
⑬	0	0	0	0	0	0	50	50	50	50
⑭	0	0	0	0	0	0	50	50	50	50
⑮	0	0	0	0	0	0	50	50	50	50
合計	553	442	573	365	566	566	663	721	745	750

表3: 虹彩検出の成功フレーム数(人物B)

マーカ番号	右カメラ			左カメラ			下カメラ			統合
	右目	左目	統合	右目	左目	統合	右目	左目	統合	
①	50	50	50	50	50	50	10	28	30	50
②	50	50	50	50	50	50	1	7	8	50
③	50	50	50	0	50	50	0	1	1	50
④	50	50	50	50	50	50	47	50	50	50
⑤	43	44	44	43	45	45	31	40	40	45
⑥	50	50	50	0	50	50	42	41	45	50
⑦	31	50	50	50	50	50	49	50	50	50
⑧	50	50	50	50	50	50	50	50	50	50
⑨	50	50	50	0	50	50	49	50	50	50
⑩	0	0	0	50	50	50	50	50	50	50
⑪	21	50	50	0	50	50	50	50	50	50
⑫	50	50	50	0	0	0	50	50	50	50
⑬	0	34	34	27	33	41	2	50	50	50
⑭	0	27	27	15	42	48	50	50	50	50
⑮	0	0	0	0	0	0	49	24	50	50
合計	495	605	605	385	620	634	530	591	624	745

表4: 視線推定の誤差 [mm]

マーカ番号	人物A(学習モデル本人)				人物B(未知のユーザ)			
	右	左	下	統合	右	左	下	統合
①	184.17	90.58	174.43	79.51	63.74	135.98	296.66	111.38
②	109.13	103.37	109.46	31.17	100.78	56.21	202.52	57.14
③	106.63	225.78	65.48	38.41	84.31	128.12	111.70	70.59
④	120.82	47.75	104.39	37.71	27.35	60.38	96.71	24.58
⑤	103.20	59.95	214.02	103.67	103.25	85.21	125.09	51.79
⑥	192.42	125.68	99.28	79.81	185.09	159.25	151.93	79.66
⑦	60.89	117.85	112.19	83.31	127.03	134.24	75.80	85.84
⑧	72.56	144.10	113.19	112.25	108.15	186.89	63.32	84.23
⑨	235.74	155.74	114.06	144.14	216.95	57.72	70.84	73.83
⑩	67.66	178.17	62.34	71.68	-	221.14	185.98	195.31
⑪	188.56	210.26	104.58	143.47	155.59	250.69	130.43	135.23
⑫	411.48	201.66	125.80	149.83	260.75	-	137.21	89.52
⑬	-	-	99.49	99.49	483.15	356.53	95.90	198.00
⑭	-	-	150.54	150.54	461.70	379.71	52.30	167.11
⑮	-	-	114.75	114.75	-	-	312.11	312.11
平均	154.44	138.41	117.60	95.98	182.91	170.16	140.57	115.75

差1pixelによって1m先のディスプレイ上の注目座標は約25.6mmずれる。また、カメラキャリブレーションの誤差や特徴点抽出の誤差、表情の変化や顔形状の個人差なども存在し、眼球

の大きさを考慮すると、顔向き推定（眼球中心位置）の誤差の影響は非常に大きい。そのため、視線の誤差1度といった高精度な推定を行うためには、目を大きく撮影できる高解像度なカメラを併用する必要があると考えられる。

その他、推定精度を向上させるためには、未知の人物や照明変動に対する特徴手抽出の安定化、他カメラや前フレームまでの結果を利用した再推定、統合アルゴリズムの改良、効果的なカメラ配置の検討なども必要と考えられる。

## 5. まとめ

本稿では、実環境下に設置した情報提示ディスプレイを想定し、リアルタイムに視線を推定するシステムを構築した。カメラ向きを制御することで、W100×H50×D40cmの範囲において視線推定が可能となり、立ち位置の違いや身長個人差に対応できた。また、3台のカメラを独立に処理し、推定結果を統合することでセルフオクルージョンにロバストなシステムを構築した。注目場所の推定誤差はディスプレイ上で約100mmであり、ユーザの注目場所を大まかに推定できると考えられる。

今後の課題としては、視線推定精度を向上させるため、特に高解像度カメラを併用する方法を検討する必要がある。また、カメラ向きを自動的に制御する方法や、インタラクションの具体的な内容について検討する必要がある。

**謝辞** 本研究の一部は文部科学省科学研究費補助金特定領域研究 No.18049046 の補助を受けて行った。

## 参考文献

[1] C. Hennessey, B. Noureddin, and P. Lawrence, "A Single Camera Eye-Gaze Tracking System with Free Head Motion", *ETRA*, pp.87-94, 2006.

[2] 大野健彦, 武川直樹, 吉川厚, "2点補正による簡易キャリブレーションを実現した視線測定システム", *情処学論*, vol.44, no.4, pp.1136-1149, 2003.

[3] J.G. Wang, E. Sung, and R. Venkateswarlu, "Estimating the eye gaze from one eye", *CVIU*, vol.98, pp.83-103, 2005.

[4] 辻徳生, 柴田真吾, 長谷川勉, 倉爪亮, "視線計測のためのLMedSを用いた虹彩検出法", *MIRU*, pp.1-684-689, 2004.

[5] 森山剛, J. Xiao, J. F. Cohn, 金出武雄, "高精細な目のモデル化と顔画像解析への応用", *MIRU*, pp.830-835, 2006.

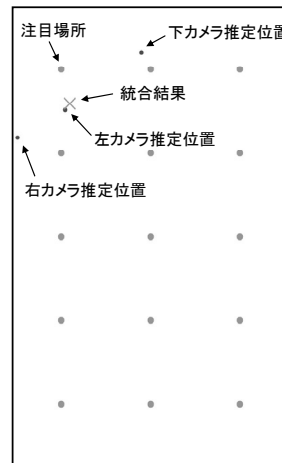


図12: 視線推定結果の例

[6] 北川洋介, 呉海元, 和田俊和, 加藤丈和, "自動視線推定のためのアイモデルの個人適用法", *信学技法 PRMU2006-192*, pp.55-60, 2007.

[7] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, "Real-Time Stereo Tracking for Head Pose and Gaze Estimation", *FG*, pp.122-128, 2000.

[8] 松本吉央, 怡土順一, 竹村憲太郎, 小笠原司, "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", *情処学論 (CVIM)*, vol.47, no.SIG15, pp.10-21, 2006.

[9] 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", *信学論*, vol.88-D-II, no.8, pp.1601-1613, 2005.

[10] F. Dornaika and F. Davoine, "Head and Facial Animation Tracking using Appearance- Adaptive Models and Particle Filters", *CVPR*, 2004.

[11] 菅野裕介, 佐藤洋一, "顔変形を伴う3次元頭部姿勢の単眼推定", *MIRU*, pp.311-317, 2007.

[12] 山添大丈, 内海章, 安部伸治, "顔特徴点の追跡を用いた単眼姿勢推定", *MIRU*, pp.319-324, 2007.

[13] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models", *PAMI*, vol.23, no.6, pp.681-684, 2001.

[14] M. B. Stegmann, B. K. Ersbøll, R. Larsen, "FAME - A Flexible Appearance Modeling Environment", *IEEE Transactions on Medical Imaging*, vol.22, no.10, pp.1319-1331, 2003.

[15] 大津展之, "判別および最小2乗規準に基づく自動しきい値選定法", *信学論*, vol.63-D, no.4, pp.349-356, 1980.