

実時間注視判定システム

石見 太郎[†], 三宅 哲夫[†], 章 忠[†],

[†] 豊橋技術科学大学

近年、誰もが容易に扱うことができるヒューマンインタフェースの必要性が高まっている。そこで我々は視線を用いたヒューマンインタフェースの開発を行ってきた。本システムは特殊な光源や装置を用いずに、単一のカメラ画像を用いてカメラを注視したか否か(以下、注視判定)をリアルタイムで判定する。従来は顔面にマークを貼り付けて注視判定を行っていたが、新たにマークの代わりとなる仮想参照点 (VRP: Virtual Reference Point) を用いた注視判定法を提案する。本手法は顔の3次元モデルは用いずに、顔の特徴点から VRP の位置推定を行う。本稿では、VRP の有効性について述べる。

Real-Time Eye-Gaze Estimation System

Taro IWAMI[†], Tetsuo MIYAKE[†], Zhong Zhang[†]

[†] Department of Production Systems Engineering Toyohashi University of Technology

A human interface that uses a line of sight has been studied. In this paper we propose a new eye-gaze input method. The method uses only a single facial image acquired under ordinary lighting condition, and it requires no laborious camera calibration. A state of eye-gaze is estimated by a virtual reference point, which is determined based on a motion of feature points on a face. We demonstrate the validity of the new method by simulation and experiments.

1 はじめに

近年、高齢者・障害者への支援、介護が重要な課題となっている。また、ペット型ロボットや様々な作業をこなす産業機械が人間社会に浸透し、人間と機械の共存性が高まるなか、機械を操作するインタフェースとして、誰もが容易に扱うことができるヒューマンインタフェースの必要性が高まっている。目の運動機能は、最後まで残るとされており、人の意図をよく表す視線を用いたヒューマンインタフェースは非常に有用である。

これまで我々は視線の追跡を行うことで、利用者が CCD カメラのレンズを注視した時にシステムが応答するヒューマンインタフェースの開発を行ってきた。本システムは、1台のビデオカメラで撮影した顔画像を基に、利用者がカメラを見ているか否かをリアルタイムで判定する(以下、注視判定)。本システムの注視判定法は、特殊な装置や光源やカメラキャリブレーションを必要とせず、頭部の動きに関する制約がない。従来は、注視状態の判定(以下、注視判定)に、顔面に貼付した一対のマークを用いていたが、マークの代わりとなる仮想参照点 (VRP: Virtual Reference Point) で注視判定を行う手法を開発した。VRP は目尻、目頭、口

元といった顔の特徴点を用いて推定されるが、顔の3次元モデルを用いないため誰でも容易に扱うことが可能である。これによりマークを用いずに、注視判定を行うことが可能となり、より実用的なヒューマンインタフェースに近づいた。

顔のトラッキングは特徴点を用いて行い、ロバストな追跡手法の一つであるパーティクルフィルタ⁴⁾を用いることで、オクルージョンが生じた場合でも、安定した追跡が可能である。システムの利用者がメガネを装着している場合も、特別な認識処理を必要とせず注視判定を行えることが特徴である。

2 注視判定原理

Fig.1 に示す P, Q, L は、それぞれ左右の眼球中心の中点(以下、VRP)を画像面に投影した位置ベクトル、左右黒目中心の中点の位置ベクトル、左右の目尻間距離を表す。このとき、黒目変位ベクトル W を式(1)により定義し、注視判定に用いる黒目変位量 d を式(2)により定義する。ここで W_0 は、カメラを注視している正面顔画像(以下、基本画像)の黒目変位ベクトルを表す。

注視判定は、ある時点より過去 0.6 秒間に得た画像の 6 割以上において、黒目変位量 d が予め設

定した閾値以下である時，利用者はカメラを注視していると判定する^{1) 2)}。

$$\mathbf{W} = \frac{\mathbf{P} - \mathbf{Q}}{L} \quad (1)$$

$$d = \|\mathbf{W} - \mathbf{W}_0\| \quad (2)$$

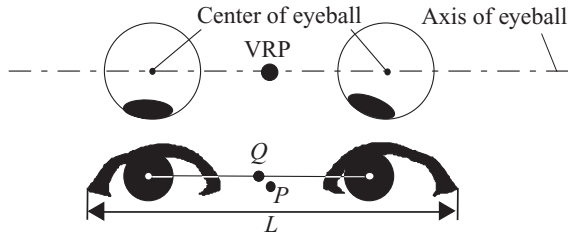


Fig. 1 Displacement of irises.

3 処理の流れ

Fig.2 に注視判定処理の流れを示す．基本情報取得部は，特徴点座標群，基本参照点，黒目変位量を取得する．判定処理部は，特徴点の追跡や瞬きの検出を行いながら参照点を仮想的に推定する．モノクロ 256 階調，512 × 418[pixel] の画像を，1[frame]あたり約 45[msec] で実時間処理を行う．

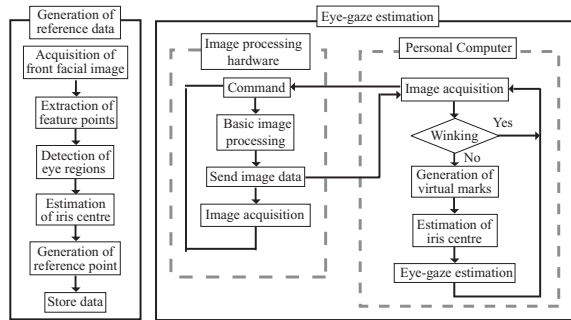


Fig. 2 Flow chart of eye-gaze estimation.

4 VRP 推定法

4.1 VRP の導出

Fig.3 に判定画像上の VRP を推定するための流れを示す．Fig.3 の左図と右図はそれぞれ基本画像と判定画像を表し，図中の (+) と (○) は特徴点と VRP を表す．特徴点群が同一平面上に存在した時，両画像の特徴点群の間には式 (3) に示すアフィン変換の関係が成り立つ． $\tilde{\mathbf{s}}$ と $\tilde{\mathbf{s}}'$ は，それぞれ基本画像上と判定画像上での共通する特徴点齊次座標を表し，これらを用いて最小二乗法からアフィン行列 A を推定する．判定画像上の VRP の位置座標は，基本画像上で予め求めた VRP 座標を，

推定したアフィン行列 A でアフィン変換して導出する．このとき基本画像上の VRP は左右の黒目中心の中点を用いる．

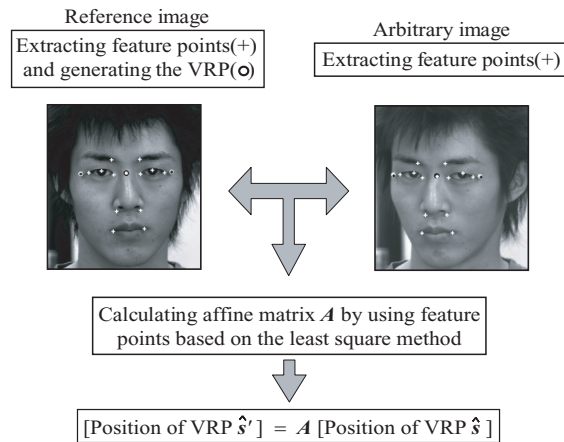


Fig. 3 Determination of the position of the VRP in an arbitrary image.

$$\tilde{\mathbf{s}}' = A\tilde{\mathbf{s}}$$

$$\tilde{\mathbf{s}} = (u, v, 1)^T, \quad \tilde{\mathbf{s}}' = (u', v', 1)^T \quad (3)$$

$$A = \begin{bmatrix} a_{11} & a_{12} & t_u \\ a_{21} & a_{22} & t_v \\ 0 & 0 & 1 \end{bmatrix}$$

4.2 VRP の補正

VRP と特徴点の位置関係を Fig.4 に示す．図中の A 面と B 面は，VRP が存在する面と特徴点が存在する面を表し，奥行き方向に D_z の差が生じている．特徴点は左右の目頭，目尻，眉頭および，鼻，口の両端の合計 10 点である．全ての特徴点は同一平面 B 上に存在すると仮定しているが，実際のところ特徴点は B 面を中心として奥行き方向に分布している．しかし，その距離はカメラと B 面までの距離 C_z に比べて微小であるため，全ての特徴点が同一平面上に存在していると仮定する．

B 面上の特徴点群から推定されたアフィン行列を用いて初期 VRP を変換すると，VRP は B 面上に生成される．そのため，式 (3) で得られた VRP は頭部の回転 (θ_x, θ_y) に伴い A 面上の真の VRP との誤差を生じる．頭部の回転によって生じる誤差は，幾何学的に求めた式 (4) と (5) の誤差関数を用いて補正する．式中の f はカメラの焦点距離を表す．なお (fD_z/C_z) は特徴点から得られる距離を用いて，1 個の値として求めることが出来る． θ_x と θ_y は，アフィン行列から求められる縦方向と横方向の回転角度であり，導出法は次章で述べる．

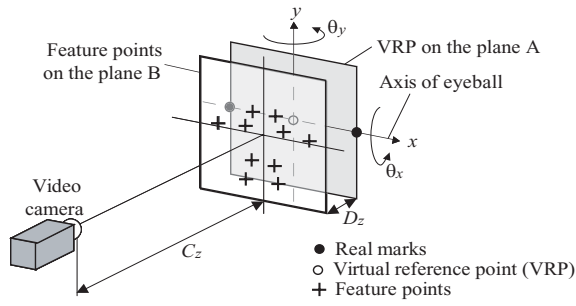


Fig. 4 Relation between the real marks and the VRP.

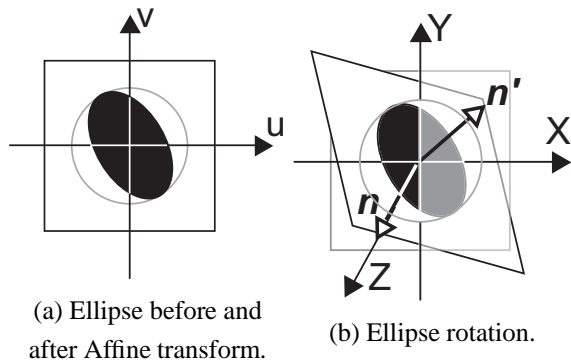
$$\Delta u = \left(\frac{f}{C_z} D_z\right) \cos\theta_x \sin\theta_y \quad (4)$$

$$\Delta v = -\left(\frac{f}{C_z} D_z\right) \sin\theta_x \quad (5)$$

5 頭部の回転角度の推定法

円をアフィン変換した楕円を用いて、頭部の回転角度 $(\theta_x, \theta_y, \theta_z)$ を推定する。アフィン変換前の円とアフィン変換後の楕円を Fig.5(a) に示す。Fig.5(b) に示す円は xy 平面上に描かれた単位円であり、楕円はその単位円を y 軸、 x 軸、 z 軸の順で回転し、 xy 平面に投影した像を表す。(a) のアフィン変換と、(b) の各軸の回転で得られた像を xy 面に投影する変換は、等価の変換であり、アフィン変換で得られた楕円を用いて $(\theta_x, \theta_y, \theta_z)$ を推定できる。

推定方法を以下に述べる。 y 軸回転、 x 軸回転、 z 軸回転の回転行列はそれぞれ式 (6), (7), (8) で表される。また x, y, z 軸方向の単位ベクトルをそれぞれ l, m, n 、回転後の x, y, z 軸方向の単位ベクトルをそれぞれ l', m', n' とする。



(a) Ellipse before and after Affine transform.

(b) Ellipse rotation.

Fig. 5 Geometric relation of Affine transform.

$$R_y = \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \quad (6)$$

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \quad (7)$$

$$R_z = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ -\sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Fig.5(a) の楕円の短径 L_s と長径 L_m の比は、回転前の n と回転後の n' のなす角 ϕ で表される (式 (9))。 n と n' の内積は θ_x, θ_y を用いて式 (10) で表される。式 (9) に式 (10) を代入すると、式 (11) を得る。この式は頭部の回転角度 θ_x および θ_y を用いて、楕円の短径と長径の比で表されることを示している。

$$\frac{L_s}{L_m} = \cos\phi = n \cdot n' \quad (9)$$

$$\begin{aligned} n \cdot n' &= n \cdot (R_y R_x R_z n) \\ &= \cos\theta_x \cos\theta_y \end{aligned} \quad (10)$$

$$\cos\theta_x \cos\theta_y = \frac{L_s}{L_m} \quad (11)$$

l と R_x, R_y, R_z から導出した l' を式 (12) に示す。ここで、 xy 平面に投影した楕円の形状は z 軸回転の有無に関係なく一定である。そこで式 (12) に $\theta_z = 0$ を代入すると、式 (13) に示すベクトル l'_{xy} を得る。

l'_{xy} は y 軸成分を持たないので、 l'_{xy} を xy 平面に投影すると、 x 軸成分のみのベクトルとなる。この x 軸成分は、Fig.5(a) に示すアフィン変換で得られた楕円の $v = 0$ の値に相当し、交点 $(u', 0)$ は、式 (15) で与えられる。このとき、単位円の極座標方程式 (14) をアフィン変換し、そのときの v 座標が 0 となる ζ を求める。 (u', v') と l'_{xy} を xy 平面に投影したベクトルは等価となり、 θ_y は式 (16) で表される。

$$\begin{aligned} l' &= R_y R_x R_z l \\ &= \begin{bmatrix} \cos\theta_y \cos\theta_z + \sin\theta_x \sin\theta_y \sin\theta_z \\ \cos\theta_x \sin\theta_z \\ -\sin\theta_y \cos\theta_z + \sin\theta_x \sin\theta_y \sin\theta_z \end{bmatrix} \end{aligned} \quad (12)$$

$$l'_{xy} = [\cos\theta_y, 0, -\sin\theta_y] \quad (13)$$

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos(\zeta) \\ \sin(\zeta) \end{bmatrix} \quad (14)$$

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} a_{11}\cos(\zeta) + a_{12}\sin(\zeta) \\ 0 \end{bmatrix} \quad (15)$$

$$\zeta = -\tan^{-1} \frac{a_{21}}{a_{22}}$$

$$\begin{aligned} \cos\theta_y &= \frac{u'}{L} \\ &= \frac{a_{11}\cos(\zeta) + a_{12}\sin(\zeta)}{L} \end{aligned} \quad (16)$$

式(16)を式(11)に代入することで θ_x が導出できる．また、 l' をxy平面に投影したときのベクトルは、uv座標上のベクトル(1,0)をアフィン変換した時のベクトルと一致する．これら二つのベクトルを用いて式(17)が導かれ、 θ_z が推定できる．

$$\tan\theta_z = \frac{a_{21}\cos\theta_y}{\cos\theta_x(a - \tan\theta_x\sin\theta_y)} \quad (17)$$

6 特徴点の追跡

各々の特徴点を独立して追跡すると、頭部の回転や障害物によってオクルージョンが生じた時、4.1で述べた最小二乗法から安定したアフィン行列を推定することは困難となる．オクルージョンが起こった場合でもロバストな追跡ができる手法の一つとして、パーティクルフィルタを用いた特徴点追跡方法があり、岡らによってステレオ計測での追跡について報告されている³⁾．以下に、パーティクルフィルタを用いて運動する特徴点群を追跡する手法について説明する．

6.1 状態ベクトル

パーティクルフィルタは、各時刻における状態ベクトルを逐次的に推定できる手法である．アフィン行列の成分式(18)は、並進(T_u, T_v)、回転(θ, ϕ)、拡大または縮小(α, β)に分解でき、これらを状態ベクトル p としてアフィン行列を逐次的に推定する．

$$A = \begin{bmatrix} \alpha\cos\theta & -\beta\sin\phi & T_u \\ \alpha\sin\theta & \beta\cos\phi & T_v \\ 0 & 0 & 1 \end{bmatrix} \quad (18)$$

$$p = [\theta, \phi, \alpha, \beta, T_u, T_v]^T \quad (19)$$

6.2 パーティクルフィルタ

パーティクルフィルタは、予測、観測、推定という3つの操作によって行われる．パーティクル数を N とし、状態ベクトルと状態ベクトルに対応する重みを式(20)で表す．重みの求め方は観測段階で述べる．

$$q^{(n)} = [p^{(n)}, \pi^{(n)}] \quad (20)$$

予測

予測段階は、1サンプル前の $q^{(n)}$ から重み $\pi^{(n)}$ に基づいて状態ベクトルを選択し、式(21)を用いて予測する．ここで、 Δp は状態ベクトルの速度であり、隣り合う画像フレーム間では等速直線運動が成り立つと仮定している．予測した状態ベクトル p' に式(22)を用いてガウス性のシステム雑音 $p_{noise}^{(j)}$ を加え、ガウス雑音を含んだ状態ベクトルが重み $\pi^{(n)}$ に比例した数だけ生成される．このときユーザーの突発的な運動にも対応するため、システム雑音は速度に比例して適応的に変化する．生成された状態ベクトル数が N 個になるまで、操作を繰り返す．

$$p' = p^{(n)} + \Delta t_{i-1} \times \Delta p, \quad \Delta p = \frac{p_{i-1} - p_{i-2}}{\Delta t_{i-2}} \quad (21)$$

$$p^{(j)} = p' + p_{noise}^{(j)} \quad (22)$$

観測

観測段階は、画像間の相互相関を用いて式(22)で得られた状態ベクトルの重みを求める．式(23)を用いて、各アフィン行列 $A^{(n)}$ に対する特徴点座標 $\hat{s}_k^{(n)}$ を K 個求める． \hat{s}_k は基本画像における特徴点座標を表す．なお $A^{(n)}$ は、新しく生成された状態ベクトル $p^{(n)}$ を用いて式(18)から求める．

ある状態ベクトルにおける K 個の特徴点に関する相互相関値の合計値 $W^{(n)}$ を求める(式(24))．このとき $\mathcal{N}()$ は K 個の特徴点近傍の画像同士の相互相関関数を表し、相関値は -1 から $+1$ の値をとる．

$$\hat{s}_k^{(n)} = A^{(n)} \hat{s}_k, \quad (k = 1, \dots, K) \quad (23)$$

$$W^{(n)} = \sum_{k=1}^K \mathcal{N}(\hat{s}_k, \hat{s}_k^{(n)}) \quad (24)$$

推定

最後の推定段階では、複数の相互相関合計値 $W^{(n)}$ は正規分布に従うものとして、式(25)を用いて正規化された重み $\pi^{(n)}$ を求め、この重みを用いて式(26)により p_i を推定する．

$$\pi^{(n)} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(K - W^{(n)})^2}{2\sigma^2}\right) \quad (25)$$

$$p_i = \frac{\sum_{n=1}^N \pi^{(n)} p^{(n)}}{\sum_{n=1}^N \pi^{(n)}} \quad (26)$$

7 実験方法

シミュレーションから従来法の実マークを用いた黒目変位量とVRPの黒目変位量を比較して、VRPの検証を行う．Fig.6に示すシステムを用いて実験

を行った．被験者とカメラの距離は 2.4[m] で固定し，指標は視線角 10[Deg] と 20[Deg] の位置に計 16 個，配置する．被験者は自由に指標注視し，ときどき口元や鼻を手で数秒間隠し，カメラを注視したときは手元のスイッチを押す．被験者のメガネの有無は問わないが，今回はメガネを装着した状態で実験を行った．なお，追跡に用いたパーティクル数は 500 とした．

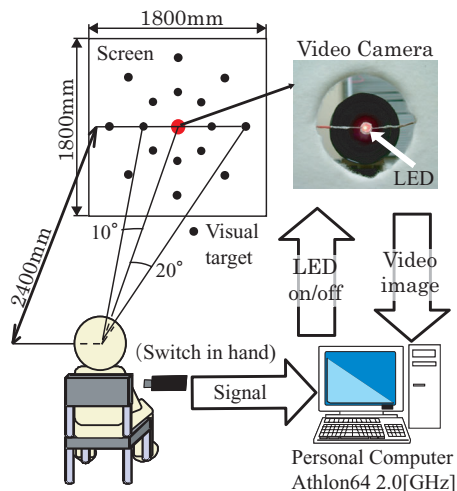


Fig. 6 System config.

8 結果と考察

8.1 シミュレーション

VRP の有効性を確認するために，Fig.7 に示す眼球モデルを用いてシミュレーションを行った．図中の黒丸 (●) は実マークを表し，画像面に投影された左右の実マーク中心の中点を用いて実マークの黒目変位量を求める．視線をカメラの方向に固定したまま，顔の正面方向とカメラの光軸のなす角が 10,20[Deg] になるように θ_x と θ_y の割合を変化させ，そのときの黒目変位量の変化を求めた．

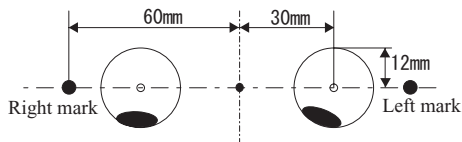


Fig. 7 Simulation Model.

Fig.8 にシミュレーション結果を示す．グラフの横軸と縦軸は，それぞれ縦方向と横方向の黒目変位量を表し，白丸 (○) と黒丸 (●) は VRP と実マークを用いたときの黒目変位量を表す．VRP による黒目変位量は頭部の回転のため円周上に分布するが，この原因は VRP の補正関数の影響である．

カメラ注視状態における黒目変位量は理想的にはゼロとなるが，Fig.8 からわかるとおり実マークを用いる場合でも誤差があり，誤差量は仮想マークを用いる場合と実マークを用いる場合とで同程度である．したがって，VRP を用いても注視判定には影響はないと考えられる．

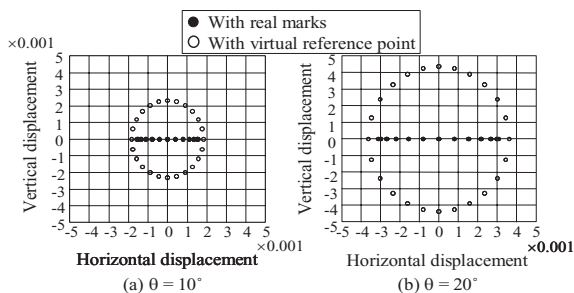


Fig. 8 Distribution of the displacement of irises by simulation.

8.2 実験

Fig.9 は 200 フレーム間隔の処理結果画像を示す．No.0800 に示すオクルージョンが生じても特徴点を追跡でき，メガネフレームを特に認識することなく注視判定が行える．

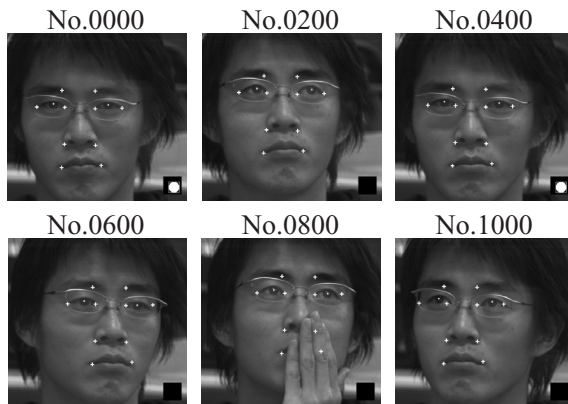


Fig. 9 The reference image and the video images picked out every 200 frames.

Fig.10 に Fig.9 の被験者における正味変移量の変化と，それに対応するシステムの応答状況を示す．実線波形が正味変移量を，黒色の矩形波が被験者の注視意思表示スイッチの ON/OFF，灰色の矩形波がシステムの注視判定結果である．破線は注視判定閾値を表し，閾値は 0.015 とする．被験者が注視すると，システムは必ず応答するが，判定の立ち上がりと立下りにばらつきが生じ，4 箇所の誤判定が生じた．理由は，いずれも頭部の回転角度に誤推定が生じ，VRP を正確に補正できなかった

たためである．頭部の回転角度が正しく推定できない理由として，頭部の回転が正面画像に対して5[Deg]以内のとき，回転による特徴点の変化が画像上で捉えられないためと考えられる．システムの正解率は，スイッチのON/OFFと注視状態の推定結果の一致率であり，89.4[%]となった．

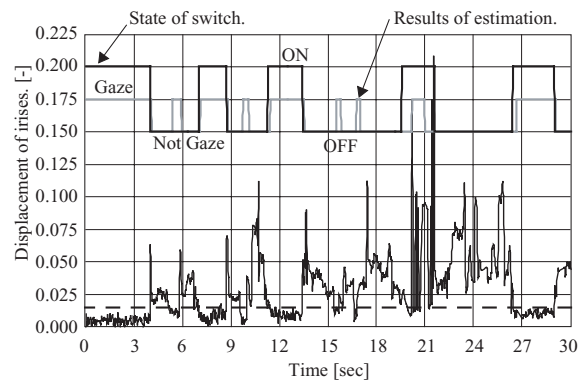


Fig. 10 Experimental results of real-time eye-gaze estimation.

9 おわりに

仮想参照点 (VRP) を提案し注視判定を行った．注視判定はリアルタイム (45[msec/frame]) で実行され，その正解率は89.4[%]であった．パーティクルフィルタを用いることで，特徴点の追跡中にオクルージョンが生じた場合でも安定した追跡が可能となった．また，メガネを装着している場合でも，同様にオクルージョンの問題が回避できることが実験により確かめられた．

提案した VRP を用いることにより，利用者がマークを取り付ける不快感や煩わしさを感じることなく，注視判定を行うことが可能となり，より多くのニーズに応えることができると考えられる．今後は，頭部の回転角度を正確に捉える手法を考え，誤判定の軽減を目指す．

参考文献

- 1) 三宅 哲夫, 春田 誠司, 堀畑 聡, "顔の向きに依存しない特徴量を用いた注視判定法", 電子情報通信学会, Vol.j86-D- , No.12, pp.1737-1744, Decenber,2003.
- 2) Tetsuo Miyake, Taro Iwami, Satoshi Horihata, Zhong Zhang , "Real time eye gaze estimation by using a virtual reference point", Proc. of the 5th WSEAS International Conference on SPCGAV, pp.123-127, Sep. 2005.

- 3) 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", 電子情報通信学会, Vol.j88-D- , No.8, pp.1601-1613, August,2005.
- 4) M. Isard and A. Blake, "Condensation - conditional density propagation for visual trackong", Int. J. Comput. Vis., Vol.29, no.1, pp.5-28, 1998.