

First Person VisionのためのInside-Outカメラの提案

清水 彰一[†] 西尾 和晃[†] 木村 誠^{††} 藤吉 弘亘[†]

[†] 中部大学工学部 〒487-8501 愛知県春日井市松本町 1200

^{††} 産業技術総合研究所 デジタルヒューマン研究センター 〒135-0064 東京都江東区青海 2-41-6

E-mail: [†]{shiyou,xeno,hf}@vision.cs.chubu.ac.jp, ^{††}makoto.kimura@aist.go.jp

あらまし 人の行動意図を認識しようとする First Person Vision では、人の状態とその人が何を見ているのかという情報が必要となる。そこで我々は、人の眼球と人の視界映像を同時に取得する Inside-Out カメラを提案し、そのカメラの構成を生かした注視点の推定法も提案する。Inside-Out カメラはハーフミラーを介して眼球を正面から、視界映像を眼球と同等の位置から撮影することができる。Inside-Out カメラでは、眼球を撮影した画像から得られる視線ベクトルと視界を撮影した画像から得られる注視点位置の関係を変換式で表すことが可能である。そのため、変換式のパラメータをあらかじめ推定することにより、視線ベクトルから注視点を推定する。評価実験では、ランドマークを注視した際の両眼、両視界映像を撮影し、視界画像のランドマーク位置を真値として視線ベクトルから推定された注視点との誤差を算出した。その結果より、約 15 pixel の平均誤差で注視点を推定可能であることを確認した。

キーワード First Person Vision, Inside-Out カメラ, 視線検出, ステレオ視

Proposal of Inside-Out Camera for First Person Vision

Shoichi SHIMIZU[†], Kazuaki NISHIO[†], Makoto KIMURA^{††}, and Hironobu FUJIYOSHI[†]

[†] Dept. of Computer Science, Chubu University Matsumoto-cho 1200, Kasugai-shi, Aichi, 487-8501 Japan

^{††} AIST Digital Human Research Center, Aomi 2-41-6, Koutou-ku, Tokyo, 135-0064 Japan

E-mail: [†]{shiyou,xeno,hf}@vision.cs.chubu.ac.jp, ^{††}makoto.kimura@aist.go.jp

Abstract First Person Vision, which attempts to understand a person's behavioral intention, requires information on the person's state and on what the person is looking at. We propose an inside-out camera that simultaneously obtains image of one of the person's eyeballs and image of that person's visual field, and propose a method for estimating the person's gaze point based on the configuration of the camera. The inside-out camera uses a half mirror to capture the eyeball from the front and the person's visual field from the same position as the eyeball. The relationship between the gaze vector obtained from the eyeball video and the gaze point obtained from the scene video is expressed by a conversion equation, and the gaze point is estimated from the gaze vector by estimating beforehand the parameters of this conversion equation. In an evaluation experiment, we took scene video for both eyes when gazing at a certain landmark as evaluation data, and calculated the error between the position of the landmark in the scene video as the true value and the gaze point estimated from the gaze vector. It was found that the gaze point could be estimated with an average error of about 15 pixels.

Key words First Person Vision, Inside-Out Camera, Gaze Estimation, Stereo Vision

1. はじめに

First Person Vision では、人の頭部にカメラを装着し、人の視界と同じような画像を得ると同時に他方のカメラで眼球の観測を行うことにより、人の行動意図を観測しようとする試みである [1].

従来、人の行動理解を行うためのセンサとして、ウェアラブ

ルカメラが用いられている。久保田らは、頭部に装着したカメラから個人行動記録システムのために注目シーンの検出を提案している [2]。この注目シーンは、画像中の移動量を観測することにより検出される。また、青木らは、人物の動作特徴を抽出し、動作特徴を自然言語の概念に対応付けて、装着者の動作を自然言語テキストで表現する手法を提案している [3]。カメラは、人の移動量を安定して取得するために胸元に装着し、人物

の移動量から人の移動方向認識を行い、手領域を検出することにより手の動きを認識している。このように、ウェアラブルカメラを用いて人の行動を認識する手法が提案されているが、人の行動意図までを観測・認識することができない。First Person Vision では、人の行動意図を観測することを目的とするため、人がどのように動いているかに加え、人が何を注目しているのかという情報も重要となる。そのため、First Person Vision では、視線を観測するカメラを使用し、人の行動意図の観測を行う。

視線を検出するためのカメラシステムは古くから製品化され、現在では人の視界を撮影するカメラを備えた製品が存在する[4]～[8]。また、視線を検出する手法も多く提案されている[9]～[12]。製品化されたカメラは、眼球を撮影するカメラと視界を撮影するカメラを有するが、カメラの設置構造上の問題により、眼球は正面から撮影されておらず、視界も実際の視界と異なる。視線計測では、眼球を正面から撮影することにより、視線検出を容易にすることができる。また、視界は実際の視界と同等の位置から撮ることにより、オクルージョンの発生による見えの違いを軽減することが可能となる。

そこで我々は、ハーフミラーを介することにより眼球を正面から、視界を眼球と同等の位置から撮影する Inside-Out カメラを提案する。また、Inside-Out カメラの構造上の特長を生かした注視点の算出法についても提案する。Inside-Out カメラは、合計 4 台のカメラがあり、左右の眼球を撮影し 2 台の Eye カメラ、視界を撮影する 2 台の Scene カメラにより構成される。注視点の算出には、あらかじめ Eye カメラの画像から推定した視線ベクトルから Scene カメラの画像に存在する注目点へ変換する変換式のパラメータを推定する。このパラメータを用いることにより注視点移動ベクトルから算出することが可能となる。

2. 従来の視線計測装置

本章では、従来の視線計測装置として、頭部装着型とスタンドアロン型について述べ、その長所と短所を示す。

2.1 頭部装着型

頭部装着型の視線計測装置[4]～[6]は、両眼と視界の撮影を行うことが可能である。眼球を撮影するカメラは 2 台、視界を撮影するカメラは 1 台で構成される。スポーツ選手の視線解析やヘッドマウントディスプレイを装着することでコンピュータゲーム等に利用されている。

2.2 スタンドアロン型

スタンドアロン型の視線計測装置[7],[8]は、非接触であるため、自然な状態で被験者の視線を計測することが可能であるが、視界の映像を撮影することはできない。被験者への負担が少なく、子供と大人に関係なく視線計測を容易に行うことができる。

2.3 両タイプの比較

両タイプの長所と短所を表 1 にまとめる。頭部装着型は、頭部に装着することにより被験者が自由に移動できるため、スポーツ選手の視線解析等に利用されている。しかし、人体に装着する装置は被験者に負担がかかるため、計測される視線が日常のものとは異なる可能性がある。スタンドアロン型は、視界

映像を撮影できず、被験者の行動は装置から眼球が撮影できる範囲に限られるが、被験者に「計測されている」という意識を与えることなく視線を計測できるため、自然なデータを得る事ができるメリットがある。

表 1 両タイプの比較

	頭部装着型	スタンドアロン型
被験者の移動	制限なし	制限あり
被験者の負担	中	低
精度	高	中
視界映像の撮影	可能	不可能

2.4 従来の視線計測装置の問題点

従来の視線計測装置は、図 1(a) に示すように、視界を撮影する Scene カメラが眼球とは異なる位置に設置されているため、実際の人の視界と撮影される映像に視差が存在する。複数カメラのステレオ視と同様、被写物体までの距離が短い近距離映像になると、人の視界とカメラ映像の視差は増大する。そのため、従来の視線計測装置は、注視対象をある程度離れた対象に設定する、あるいは注視対象を既知の平面上に限定するなど、注視点に関して何らかの拘束条件を設定することが一般的であった。

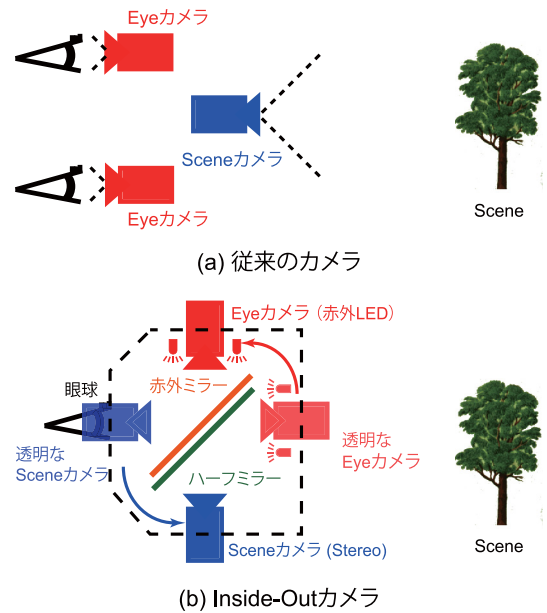


図 1 従来の視線推定装置との比較

3. Inside-Out カメラ

本章では、提案する Inside-Out カメラについて述べる。

3.1 装置構成

これまでに試作した装置は、図 2 に示すように、ゴーグル上部に配置した眼球を撮影する Eye カメラ 2 台と、下部に配置した視界を撮影する Scene カメラ 2 台で構成されたゴーグル型の視線計測装置である。寸法は $W160 \times H80 \times D100$ mm であり、重量は約 200 g である。材質は木を使用しており、カメラが発する熱を被験者に伝えにくいため、長時間の計測が可能である。Inside-Out カメラは図 1(b) に示すように、透明なカメ

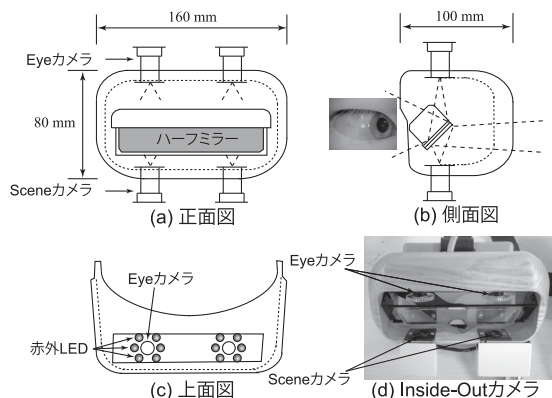


図 2 Inside-Out カメラの機構

ラが存在するような光学配置を実現している．その詳細は以降で述べる．

3.2 Eye カメラ

Eye カメラは、赤外ミラー、左右の眼球を撮影する 2 台の赤外カメラとカメラの周りに 6 個ずつ配置された赤外 LED で構成されている．赤外カメラは赤外ミラーを介して、人物の眼球を目の前から撮影した近赤外画像を 640×480 pixel の解像度で取得する．Eye カメラより取得した画像を図 3 に示す．LED はカメラの周囲に配置されており $750 \text{ nm} \sim 900 \text{ nm}$ の近赤外光を発する．赤外光は不可視であるため、視覚刺激を与えることなく眼球を撮影することが可能である．

3.3 Scene カメラ

Scene カメラは左右の視界を撮影する小型 CCD カメラ 2 台と 50 % ハーフミラーで構成されている．CCD カメラの画角は約 80 度、焦点距離は約 4 mm である．50 % ハーフミラーは入射光の 50 % を反射し、残りを通過させる性質を持つ鏡である．ハーフミラーを介すことで、透明なカメラで光学的に人物視点とほぼ同一の位置からの映像を撮影することができる．また、ステレオカメラであるため、Tsai のモデル [13] や Zhang のモデル [14] を用いてキャリブレーションを容易に行うことができ、視界内における注視点の 3 次元位置を推定することが可能である．Scene カメラから得られる画像を図 3 に示す．

3.4 Eye カメラと Scene カメラの関係

Eye カメラと Scene カメラは、ハーフミラーを挟んで向かい合うように設置してある．したがって、各カメラで構成される画像平面は平行になる．Eye カメラにより観測される物体と Scene カメラで観測される物体があり、同様に移動したとすると、各画像で観測される物体の移動距離には相関関係があることは明白である．したがって、各カメラ間の関係を容易に扱うことが可能である．

4. Inside-Out カメラを用いた 3 次元注視点推定

一般に、注視点を推定するためには視線ベクトルが必要であり、眼球回転中心や角膜曲率中心、瞳孔中心より推定する手法が提案されている [9]～[12]．この視線ベクトルは、注視するランドマークに依存し動くため、視線ベクトルとランドマーク間に相関関係があることは明白である．提案する Inside-Out カ

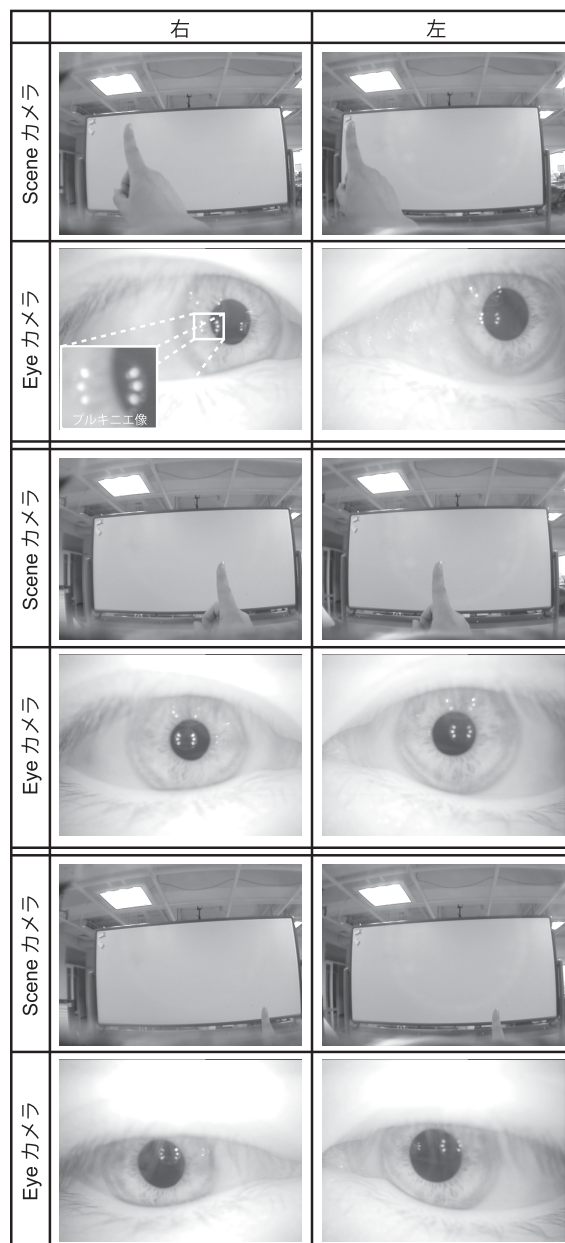


図 3 Inside-Out カメラにより得られる画像

メラでは、3.4 で述べたように、この相関関係を容易に扱うことが可能である．実際に、図 3 からわかるように、注視点が左に移動するとプルキニエ像と瞳孔中心から推定される視線ベクトルも左方向を向き、右に移動すると、右方向を向いている．したがって、注視点と視線ベクトルには相関関係があることから、視線ベクトルを推定することにより、Scene 画像の注視点を推定することが可能になる．以下に注視点推定の流れを示す．

(1) 視線ベクトルの推定

瞳孔中心とプルキニエ像から視線ベクトルを推定

(2) 注視点の算出

変換式により視線ベクトルを注視点に変換

(3) 3 次元注視点の算出

両眼の注視点を用いてステレオ視により 3 次元注視点を算出

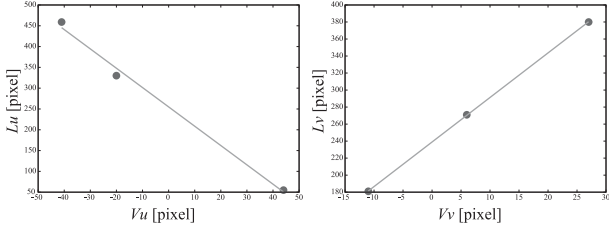


図 4 注視点と視線ベクトルの関係

4.1 視線ベクトルの推定

大野らにより提案された視線検出法[15]は、プルキニエ像から角膜曲率中心を算出し、その角膜曲率中心と瞳孔中心を結ぶ直線を視線ベクトルとする方法である。眼球回転中心と瞳孔中心を用いた手法より精度よく視線ベクトルが推定可能であると報告されている。そこで本手法においても、角膜曲率中心と瞳孔中心から視線ベクトルを推定する。

まず、視線ベクトルを推定するために角膜曲率中心をプルキニエ像を用いて算出する。大野ら[15]の光源は一つであるため、プルキニエ像をカメラの光軸中心になるように補正を行うことで、角膜曲率中心を算出している。一方、我々の提案する Inside-Out カメラは、カメラの周りに 6 個の光源があるため、その中心がカメラの光軸と仮定することができる。そのため、プルキニエ像群の中心を角膜曲率中心とすることが可能になる。プルキニエ像は、文献[12],[15],[16]により提案されている手法により抽出可能である。このようにして求めたプルキニエ像群から画像上の角膜曲率中心 $P_C = [u_C, v_C]^T$ を推定することができる。

次に、瞳孔中心を推定する。瞳孔中心 $P_p = [u_p, v_p]^T$ の算出には、坂下らにより提案された手法[17]を用いる。

以上より推定した角膜曲率中心 $P_C = [u_C, v_C]^T$ と瞳孔中心 $P_p = [u_p, v_p]^T$ から視線ベクトルを算出する。視線ベクトル $V = [V_u, V_v]^T$ は、角膜曲率中心を基準とするため、以下の式により算出できる。

$$V = P_p - P_C \quad (1)$$

4.2 変換式を用いた注視点の算出

4.1 により推定した視線ベクトルを用いて、Scene 画像上の注視点を推定する。図 3 から注視点 $L = [L_u, L_v]^T$ と視線ベクトル $V = [V_u, V_v]^T$ の各 u, v 成分を抽出し、その点に対して最小 2 乗法により直線を算出した例を図 4 に示す。このように、注視点と視線ベクトルの各成分における関係は、比例関係として近似できることがわかる。したがって、あらかじめ数点のキャリブレーションデータから、直線の方程式の傾き $a = [a_u, a_v]^T$ と切片 $b = [b_u, b_v]^T$ を算出しておくことにより、視線ベクトルを入力とした際の注視点 $L = [L_u, L_v]^T$ を算出することが可能になる。

$$L_u = a_u V_u + b_u \quad (2)$$

$$L_v = a_v V_v + b_v \quad (3)$$

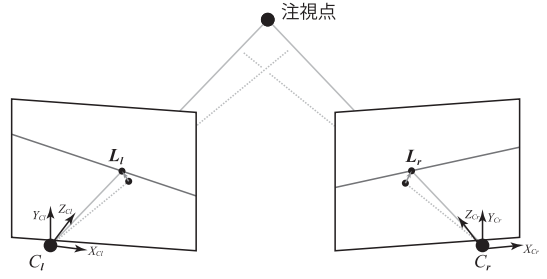


図 5 注視点の最適補正

4.3 注視点の最適補正

4.2 により算出した両眼の注視点は、推定した視線ベクトルに誤差を含むため、注視点にもずれが生じる。そのため、Scene 画像上の注視点にもずれが生じる。そこで、文献[18]にて提案されている最適補正を用いて、図 5 に示すように、Scene 画像上の注視点同士が対応するように位置の補正を行う。

左側 Scene 画像中の点 L_l と右側 Scene 画像中の点 L_r が対応しているとき、それぞれの点を通る視線同士が交差する条件は、各点を同次座標 $m_l = [L_{u_l}/f_0, L_{v_l}/f_0, 1]^T$ $m_r = [L_{u_r}/f_0, L_{v_r}/f_0, 1]^T$ (注1) で表現すると、式 (4) を満たす。

$$m_r^T F m_l = 0 \quad (4)$$

ここで、 F は基礎行列を表し、左右の画像の幾何学的関係を示す。

最適補正では、視線同士が交差するように、Scene 画像上の注視点を移動させ、その移動量が最小になるように定める。左側 Scene カメラの最適補正後の Scene 画像座標 $\hat{m}_l$ と、右側 Scene 画像座標 $\hat{m}_r$ はそれぞれ式 (5)、(6) を用いて算出する。

$$\hat{m}_l = m_l - \frac{(u\xi)P_k F m_r}{(uV_0[\xi]u)} \quad (5)$$

$$\hat{m}_r = m_r - \frac{(u\xi)P_k F^T m_l}{(uV_0[\xi]u)} \quad (6)$$

ここで、 P_k は第 3 成分を 0 にする行列である。

$$P_k = \text{diag}(1, 1, 0) \quad (7)$$

u と ξ は、それぞれ基礎行列 F とデータ m を 9 次元ベクトルに書き直したものである。

$$u = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T \quad (8)$$

$$\xi = [L_{u_l} L_{u_r}, L_{u_l} L_{v_r}, f_0 L_{u_l}, L_{v_l} L_{u_r}, L_{v_l} L_{v_r}, f_0 L_{v_l}, f_0 L_{u_r}, f_0 L_{v_r}, f_0^2]^T \quad (9)$$

ここで、行列 $V_0[\xi]$ は 9×9 の行列であり、文献[18]にて定義されている式である。

算出した $\hat{m}$ を m に代入し、式 (5)、(6) を再度適用し、収束するまで繰り返す。

(注1)： f_0 は任意のスケール定数であり、数値計算を安定化させるためのパラメータである。

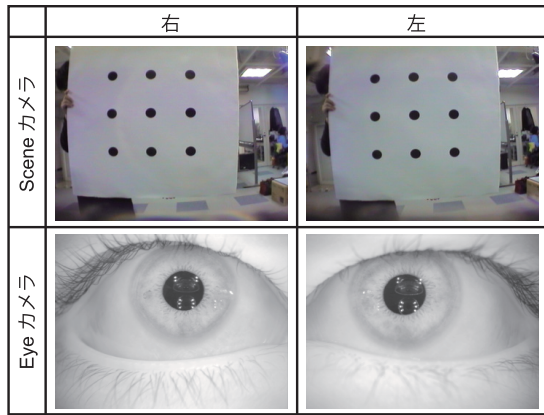


図 6 実験データの一例

4.4 3次元注視点の算出

両眼の注視点を用いて、3次元注視点を算出する。注視点は、2台のScene画像上の点であるため、ステレオ視の問題として3次元の注視点を算出することができる。人間の視野角は2度と報告されており[19]、両方の注視点の領域からステレオ視により3次元位置を求める。

5. 評価実験

本章では、提案手法を評価するために、実際にランドマークを注視したデータを採取し、ランドマークの真値と推定した注視点との誤差を算出する。また、3次元注視点を推定した例も示す。

5.1 実験概要

採取したデータは、図6に示すように、9つのランドマークを用意し、左上から順番に注視する作業を奥行きを1mから5mまでを50cm間隔でボードを移動させるごとに行い、1つの奥行きで54点、全486組のシーン画像と眼球画像を得たものである。評価人数は、5人とする。

5.2 誤差推定結果と考察

提案手法では、あらかじめ変換式のパラメータを算出する必要があり、そのパラメータの精度は、キャリブレーションデータに依存することは明らかである。そこで、キャリブレーションに使用するデータ数を変化させた際の誤差についての検討を行う。キャリブレーションに使用するデータは、1mのときのデータ54組からランダムに2組から9組までを選択し、100回の試行とする。人間の視野角は2度と報告されているため[19]、2度の領域をScene画像に投影した際の直径約20pixelを誤差範囲として評価を行う。

図7にキャリブレーションデータ数に対する平均誤差と100回の試行の最小平均誤差を示す。図7からデータ数を4組以上使用することにより、平均誤差を15pixel以下とすることができ、誤差範囲である20pixel以内に抑えることができることを確認した。また、平均誤差は減少傾向にあるため、組数を増やすことにより安定した注視点推定を可能にした。

次に、図7の最小平均誤差が約10pixelであることから、少ないキャリブレーションデータにおいても平均誤差が低く、精

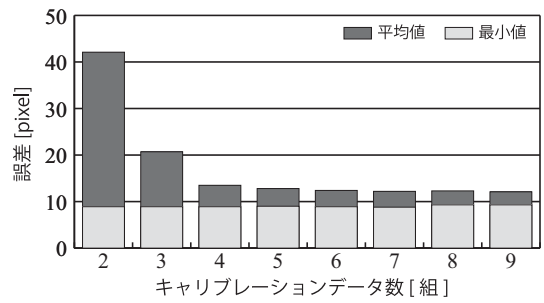


図 7 キャリブレーションに使用するデータ数に対する誤差 [pixel]

表 2 誤差におけるキャリブレーションデータの標準偏差 [pixel]

	左目		右目	
	u	v	u	v
最小	85.1	74.5	84.8	75.7
最大	18.3	64.7	17.8	65.9

度の高い推定結果が得られることがわかる。これらのキャリブレーションデータを解析することにより、キャリブレーションに有効なデータ採取方法の手助けになると考えられる。

視線ベクトル方向やランドマーク位置に偏りの少ないデータが有効なキャリブレーションデータであると考えられる。そのため、データの分布について調査を行う。表2にキャリブレーションデータを2組とした際の、最大・最小誤差における視線ベクトルの標準偏差を示す。表から v 軸に対する標準偏差は最小と最大誤差で約10pixel違い、 u 軸に対する標準偏差は約70pixelの差がある。このことから、視線ベクトルの標準偏差が大きくなるようなデータを採取することが重要であり、 v 軸のばらつきよりも u 軸のばらつきを考慮した方が、よりよいキャリブレーションデータを取得できるということがわかる。

5.3 3次元注視点の推定

算出した注視点を用いて、3次元注視点を推定した例を図8に示す。図8(a)は、直方体の箱の縁をなぞるように視線を移動させた3次元注視点である。箱の縁に沿って3次元注視点が見られていることがわかる。図8(b)は、①標識、②人間、③箱の順番に注視対象を変化させたときの3次元注視点である。3次元注視点が見られるため、注視対象の大まかな位置を把握することが可能となり、3次元ScanPathを取得することができる。図8(c)は、螺旋状にランドマークを配置し、それを注視した例である。これも図8(b)と同様に3次元ScanPathが見られていることがわかる。以上より、Inside-Outカメラでは3次元の注視点を推定可能であり、3次元空間中のどこを見ているかという情報を知ることが可能になる。そのため、人の行動意図を理解する際に、人がどこを見ているのかという情報を取得することが可能になる。

6. まとめ

本稿では、First Person VisionのためのInside-Outカメラとそのカメラの構造を生かした注視点の推定法を提案した。Inside-Outカメラはハーフミラーを介して眼球を正面から、視界映像を眼球と同等の位置から撮影することを可能とした。注視点をあらかじめ算出した変換式のパラメータを用いて、視

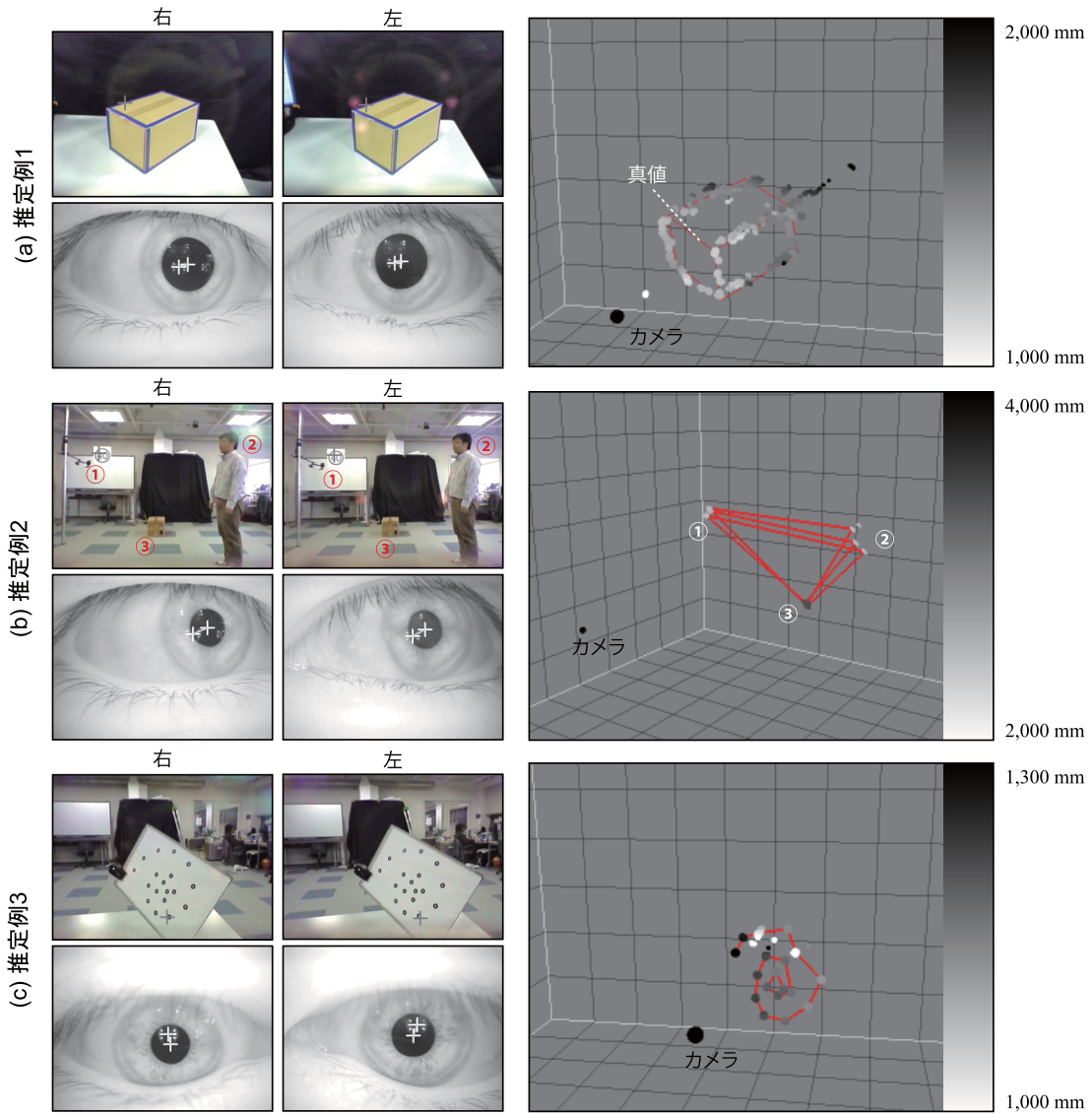


図 8 注視点の 3 次元位置推定

線ベクトルから算出する手法を提案した。評価実験から約 15 pixel の平均誤差で注視点を推定可能であることを確認した。

文 献

- [1] T. Kanade, "First-person, inside-out vision," IEEE Workshop on Egocentric Vision, CVPR, 2009.
- [2] 久保田敏司, 中村裕一, 大田友一, "個人行動記録システムにおける注目シーンの検出: 注目シーン検出の高精度化と環境カメラの利用," 電子情報通信学会技術研究報告. PRMU, vol.102, no.216, pp.47–52, 2002.
- [3] 青木茂樹, 高橋昌史, 大西正輝, 小島篤博, 福永邦雄, "ウェアラブルカメラによる動作の認識とテキスト表現," 電子情報通信学会技術研究報告. PRMU, vol.103, no.208, pp.59–64, 2003.
- [4] ナックイメーテクノロジー, "アイマークレコーダー EMR-9".
- [5] Arrington Research, "アイトラッカー ViewPoint".
- [6] ISCAN, "AA-ETL-500B".
- [7] ナックイメーテクノロジー, "アイマークレコーダー EMR-AT VOXER".
- [8] Tobii, "Tobii XL".
- [9] R. Stiefelhagen, J. Yang, and A. Waibel, "Tracking eyes and monitoring eye gaze," Proceedings of the Workshop on Perceptual User Interfaces, pp.98–100, 1997.
- [10] 竹上健, 後藤敏行, "角膜反射像と虹彩輪郭情報を併用した視線検出法," 電子情報通信学会論文誌. D-I, 情報・システム, I-情報処理, vol.82, no.10, pp.1295–1303, 1999.
- [11] 松田圭司, 永見武司, 山根茂, "視線位置計測システムの開発," 電子情報通信学会技術研究報告. TL, vol.100, no.47, pp.9–16, 2000.
- [12] 田中宏明, 疋田真一, 笠井健, 竹田仰, "二光源を用いた角膜曲率中心位置計測による視線検出法," 電子情報通信学会技術研究報告. MBE, vol.108, no.479, pp.177–180, 2009.
- [13] R.Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," IEEE Journal of Robotics and Automation, vol.3, no.4, pp.323–344, 1922.
- [14] Z. Zhang, "A flexible new technique for camera calibration," PAMI, vol.22, no.11, pp.1330–1334, 2000.
- [15] 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," 情報処理学会論文誌, vol.44, no.4, pp.1136–1149, 2003.
- [16] 大野健彦, 武川直樹, 吉川厚, "眼球形状モデルに基づく視線測定法," 第 8 回画像センシングシンポジウム, pp.307–312, 2002.
- [17] 坂下祐輔, 藤古弘亘, 平田豊, 高丸尚教, 深谷直樹, "高速楕円検出に基づく眼球回旋運動のリアルタイム計測システム," 電気学会論文誌, vol.127-C, pp.591–598, 2007.
- [18] 金谷健一, 菅谷保之, 新妻弘崇, "2 画像からの三角測量: Hartley-sturm vs. 最適補正," 情報処理学会研究報告. CVIM, vol.2008, no.27, pp.335–342, 2008.
- [19] 大野健彦, "視線を用いた高速なメニュー選択作業," 情報処理学会論文誌, vol.40, no.2, pp.602–612, 1999.