

角膜表面反射と全天球画像を用いた全周型注視点推定

小川 太士^{1,a)} 中澤 篤志^{1,b)} 西田 豊明^{1,c)}

概要: 本研究では、現実空間に対して注視点推定を行うべく、角膜表面反射画像と全天球画像を用いて、今の自分の位置から360度、全方向を対象に注視点推定を行う手法を提案する。注視点推定を行うために、全天球画像を複数方向の透視投影に変換し、角膜表面反射画像と各透視投影の位置合わせを行い、適切な透視投影と、座標変換関数を出力する。実験では、位置合わせの精度を確認するとともに、個人誤差のキャリブレーションにより、どの程度の精度で中心推定ができるかを検証し、角膜表面反射による中心点推定の有用性を確認した。

Point of Gaze Estimation Using Corneal Surface Reflection and 360° Spherical Image

TAISHI OGAWA^{1,a)} ATSUSHI NAKAZAWA^{1,b)} TOYOAKI NISHIDA^{1,c)}

1. はじめに

生体情報のなかでも、特に視線には多くの情報が含まれている。例えば我々がある対象を注視したとき、その対象に興味がある、注意しているといった情報が読み取れる。このように、視線の滞在時間、および視線を向ける頻度は、人間の興味と直結していると考えられ、商業的、学術的に利用価値が高い。ディスプレイに対しての注視点推定は、アイトラッキングなどの技術を用いて、現在多方面において開発、使用されているが、推定できる範囲はその範囲内のみである。そこで本研究では、現実空間に対して注視点推定を行うべく、シーン画像として全天球画像(図1)を使用する。これにより、従来の2次元平面上の視線推定にはない、全周型の注視点推定が可能になる。そして、この注視点推定を行うにあたり、角膜イメージングカメラを用いて撮影した、角膜表面反射(図2)を用いる。本研究では、全天球画像に対する注視点推定を角膜表面反射画像のみから行う手法を提案する。周囲360度に対する注視点推定を、角膜表面反射画像のみでシンプルに行うことが有用



図1 全天球画像, および注視点のヒートマップ



図2 角膜表面反射画像および、角膜イメージングカメラ

性のひとつである。さらに今自分のいる位置から全ての方向に対して注視点推定ができ、注視点マップが生成できるという点が今回の研究の大きな意義である。これを用いることにより、視線の解析の幅が広がり、利用価値がさらに高まることも期待される。

2. 関連研究

[角膜表面反射] 角膜表面反射は近年研究が進んでいる。

¹ 京都大学大学院情報学研究科
Graduate school of Informatics, Kyoto University

a) ogawa@ii.ist.i.kyoto-u.ac.jp

b) nakazawa.atsushi@i.kyoto-u.ac.jp

c) nishida@i.kyoto-u.ac.jp

角膜表面反射に関する研究として、Nishino ら [7] は、照度変化中における顔認識において、1人の被験者に対して、複数のサンプル画像が必要だった従来手法とは異なり、1枚の角膜表面反射画像のみを用いて、光の反射を逆投影することでシーンの光源環境を復元し、顔認識において高いスコアを残すことに成功した。また、Nishino ら [8] は角膜表面反射を利用して今見ているシーンを復元する、角膜イメージング法を提案した。そして、Nakazawa ら [4] は、角膜表面反射から眼球のモデルを構築し、画像の位置合わせ技術を組み合わせることで、角膜表面反射画像を用いて、シーン画像に対して注視点の推定を行うことができることを示した。Nitschke ら [9] は、幾何学的眼球モデルを構築することで、角膜表面反射からのシーン環境、および観測者とカメラ、シーンとの関係の復元方法について再検証し、角膜表面反射分野の包括的な説明と、その応用可能性について触れている。

[SURF 特徴量] 一般的に、2つの画像のマッチングを行う際には、2つの画像の対応点を探す必要がある。

SURF は、Baya らによって考案された特徴量である [1]。SURF は、Lowe によって考案された SIFT [3] の高速化版であり、画像認識において多く用いられている特徴量である。SURF は画像中のハイコントラストな点を検出し、検出した局所領域の勾配情報に対して主成分分析を適用し、特徴量ベクトルを形成する。SURF は特徴点について、オリエンテーションベクトルを持つため、向きと大きさで正規化でき、回転、スケール変化に対して不変である。本研究では、2画像間の位置合わせ精度を高めるため、SURF を利用することで、2画像間の変換行列を求めるために必要な対応点の数を少なくすると同時に、後述の RANSAC を適用する対象となる点を抽出している。

[画像の位置合わせ技術] 本研究では、画像の位置合わせ技術に、Fischler らによって発表された RANSAC [2] を用いている。RANSAC はロバストなパラメータ推定であり、画像のマッチングをはじめとして、幅広い分野に用いられている。RANSAC は、ランダムで取り出したデータからパラメータを求め、それを他のデータにも適用して誤差、あるいは一致度を求めることで、最も適するパラメータを探す手法である。本研究は、角膜表面反射を用いて眼球姿勢を推定するとともに、角膜表面反射とシーン画像の位置合わせを行う際に、RANSAC を SURF 特徴量の検出点に対して適用し、画像の位置合わせを試みる。

3. 注視点マップ作成システム

本研究の目的は、周囲 360 度を対象に注視点を推定することである。そこで、注視点を推定する対象となる、複数フレームの角膜表面反射画像、および注視点マップを生成する媒体となる、1枚の全天球画像を用いる。図2は角膜表面反射画像のうちの1フレームと、それを撮影する角膜

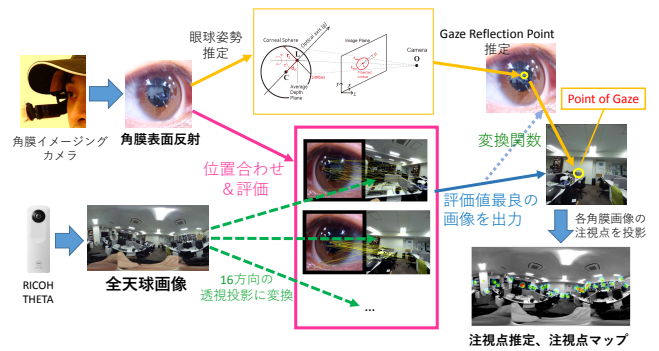


図3 注視点マップ生成の流れ

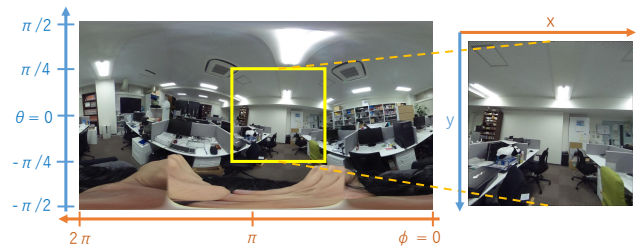


図4 全天球画像（左）と、ある方向に対する透視投影（右）

イメージングカメラであり、図1は全天球画像の一例である。全天球画像は、リコー社の製品、RICOH THETA を用いて撮影する。この画像は、上下 90 度、周囲 360 度の画像を1枚の画像上に収める全天球カメラである。本研究は、これに対して注視点推定を行うことを目標とする。注視点推定は図3の手順で行う。これを以下の節で説明する。

3.1 全天球画像の透視投影への変換

全天球画像は、水平方向が角度 ϕ 、鉛直方向が角度 θ に対応している。全周型注視点推定を行うにあたって、水平方向全て、および鉛直方向の $-\pi/4$ から $\pi/4$ を使用する。1枚の透視投影は、水平方向、鉛直方向ともに画角 $\pi/2$ で、大きさ $N \times N$ ピクセルであるとする。全天球画像と透視投影の関係を図4に示す。いま、ある透視投影 I_n の中心を全天球画像における $\theta = 0, \phi' = 0$ とし、 I_n の左上端を $I_n(0, 0)$ とすれば、

$$I_n(x, y) = (N(1 - \tan \phi')/2, N(1 - \tan \theta)/2) \quad (1)$$

$$(-\pi/4 \leq \phi', \theta \leq \pi/4)$$

が成り立つ。全天球画像の鉛直方向の中央、つまり $\theta = 0$ の部分に等間隔で置かれた 16 点それぞれを $I_n (1 \leq n \leq 16)$ の中心点とし、各 I_n について、全天球画像の各画素に対してこの式を適用し、16 方向の透視投影を生成する。

3.2 眼球姿勢の推定

まず、角膜表面反射画像から得られた眼球領域の画像から、カメラに対する眼球の 3 次元位置を定める。本節は従来法 [8],[4] と同様の手法を用いている。

角膜を球面とみなし、瞳の画像上での中心位置を i_r 、長

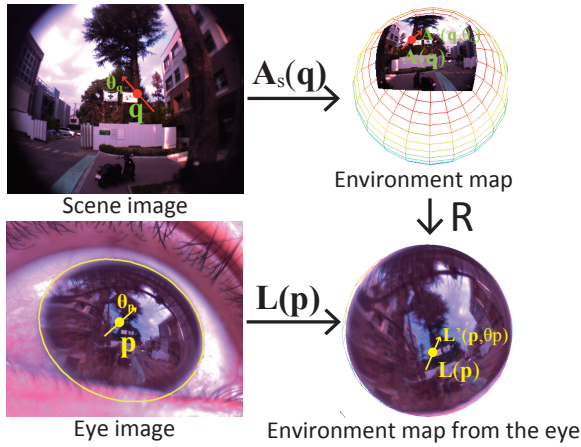


図 5 球面座標への変換 [5]

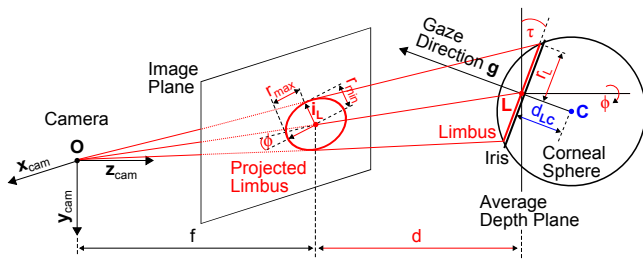


図 6 構築された眼球モデル [4]

軸，短軸を r_{\max}, r_{\min} ，回転角 ϕ とすると，角膜球面の 3 次元位置は，

$$d = r_L \cdot f / r_{\max}$$

である。ただし， $\mathbf{L} = d\mathbf{K}_e^{-1}\mathbf{i}_L$ であり， \mathbf{K}_e はアイカメラの 3×3 内部行列， f は焦点距離を示す。(図 6) また，Gaze Direction (角膜中心の表面と垂直な方向) \mathbf{g} は，

$$\mathbf{g} = [\sin \tau \sin \phi \quad -\sin \tau \cos \phi \quad -\cos \tau]^T$$

である。ただし， $\tau = \pm \arccos(r_{\min}/r_{\max})$ のうち，角膜輪部の傾きに一致する方である。

3.3 SURF による角膜表面反射画像と透視投影の対応点検出

本節は従来法 [4] と同様の手法を用いている。全天球画像の撮影座標と，眼球の位置が十分に近いことを仮定すると，角膜表面反射画像と透視投影は球面座標への変換によって関係づけることができるようになる (図 5)

角膜表面反射画像の座標 \mathbf{p} における，角膜の表面に入射する光の式 $\mathbf{L}(\mathbf{p})$ を求めたい。すなわち， $\mathbf{L}(\mathbf{p})$ は角膜表面の 3 次元座標 \mathbf{P} からシーン画像へのベクトルであるとする。まず，アイカメラ中の点 \mathbf{p} に入射する，正規化された逆投影ベクトル $\mathbf{A}_e(\mathbf{p})$ は，

$$\mathbf{A}_e(\mathbf{p}) = \mathbf{K}_e^{-1}\mathbf{p} / \|\mathbf{K}_e^{-1}\mathbf{p}\|$$

である。 \mathbf{K}_e はアイカメラの 3×3 内部行列である。ここで， \mathbf{P} とアイカメラの距離を t_1 とすれば， $\mathbf{P} = t_1\mathbf{A}_e(\mathbf{p})$ の

関係が成り立つ。角膜球の半径を r_C とすれば，角膜球について，式 $\|\mathbf{P} - \mathbf{C}\|^2 = r_C^2$ が成り立つ。これを用いると，

$$t_1^2 \|\mathbf{A}_e(\mathbf{p})\|^2 - 2t_1(\mathbf{A}_e(\mathbf{p}) \cdot \mathbf{C}) + \|\mathbf{C}\|^2 - r_C^2 = 0$$

ただし， $\|\mathbf{A}_e(\mathbf{p})\|^2 = 1$ である。これらより t_1 ， \mathbf{P} を求めると， $\mathbf{L}(\mathbf{p})$ は，

$$\mathbf{L}(\mathbf{p}) = 2(-\mathbf{A}_e(\mathbf{p}) \cdot \mathbf{N}(\mathbf{p}))\mathbf{N}(\mathbf{p}) + \mathbf{A}_e(\mathbf{p})$$

$$\mathbf{N}(\mathbf{p}) = (\mathbf{P} - \mathbf{C}) / \|\mathbf{P} - \mathbf{C}\|$$

となる。ただし $\mathbf{N}(\mathbf{p})$ は \mathbf{P} における角膜球の法線ベクトルである。

このとき，2 画像間の位置合わせ問題は，両画像から得られた Environment map どちらの変換を表す 3×3 回転行列 \mathbf{R} を求める問題として定式化できる。

$$\mathbf{L}(\mathbf{p}) = \mathbf{R}\mathbf{A}_s(\mathbf{q}), \quad \mathbf{A}_s(\mathbf{q}) = \frac{\mathbf{K}_s^{-1}[\mathbf{q}^T \quad 1]^T}{\|\mathbf{K}_s^{-1}[\mathbf{q}^T \quad 1]^T\|}$$

ただし， \mathbf{K}_s は全天球カメラとシーン画像の 3×3 対応行列である。球面座標の変換 \mathbf{R} を求めるには，2 組の対応点が必要である。しかし，角膜表面反射画像はノイズが多く，2 組の正しい候補点を発見するのは困難である。そこで，今回は回転，スケールの変化に対して不変な画像特徴量 SURF[1] を用いて，2 画像の位置合わせを行う。SURF は回転に対して不変かつ，特徴量の方向を持っているため，これを用いることで，1 組の対応点のみで \mathbf{R} を求めることができる。SURF 候補点の位置 \mathbf{x} ，特徴ベクトル $\mathbf{F}(\mathbf{x})$ ，方向 θ^x とする。SURF の初期対応点検出において得られた M ペアの特徴点と特徴量を， $\{(\mathbf{p}_i, \mathbf{F}(\mathbf{p}_i), \theta_i^{\mathbf{p}}) | i = 1, \dots, M\}, \{(\mathbf{q}_i, \mathbf{F}(\mathbf{q}_i), \theta_i^{\mathbf{q}}) | i = 1, \dots, M\}$ とすると，各ペア i について，推定変換行列 \mathbf{R}_i は，以下の式により表現できる。

$$\mathbf{R}_i = \begin{bmatrix} \hat{\mathbf{L}}_x & \hat{\mathbf{L}}_y & \hat{\mathbf{L}}_z \end{bmatrix} \begin{bmatrix} \hat{\mathbf{A}}_x & \hat{\mathbf{A}}_y & \hat{\mathbf{A}}_z \end{bmatrix}^{-1},$$

$$\mathbf{L}_x = \mathbf{L}(\mathbf{p}_i)$$

$$\mathbf{L}_y = \mathbf{L}(\mathbf{p}_i) \times (\mathbf{L}(\mathbf{p}_i) \times \mathbf{L}'(\mathbf{p}_i, \theta_i^{\mathbf{p}}))$$

$$\mathbf{L}_z = \mathbf{L}(\mathbf{p}_i) \times \mathbf{L}'(\mathbf{p}_i, \theta_i^{\mathbf{p}})$$

$$\mathbf{L}'(\mathbf{p}, \theta^{\mathbf{p}}) = \mathbf{L}(\mathbf{p} + \mathbf{u}(\theta^{\mathbf{p}})) - \mathbf{L}(\mathbf{p})$$

$$\mathbf{A}_x = \mathbf{A}_s(\mathbf{q}_i)$$

$$\mathbf{A}_y = \mathbf{A}_s(\mathbf{q}_i) \times (\mathbf{A}_s(\mathbf{q}_i) \times \mathbf{A}'_s(\mathbf{q}_i, \theta_i^{\mathbf{q}}))$$

$$\mathbf{A}_z = \mathbf{A}_s(\mathbf{q}_i) \times \mathbf{A}'_s(\mathbf{q}_i, \theta_i^{\mathbf{q}})$$

$$\mathbf{A}'_s(\mathbf{q}, \theta^{\mathbf{q}}) = \mathbf{A}_s(\mathbf{q} + \mathbf{u}(\theta^{\mathbf{q}})) - \mathbf{A}_s(\mathbf{q})$$

$$\mathbf{u}(\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \end{bmatrix}^T$$

ただし， $\hat{\mathbf{x}}$ は \mathbf{x} を正規化したものであり， $\mathbf{L}'(\mathbf{y}_i, \theta)$ および $\mathbf{A}'_s(\mathbf{y}_i, \theta)$ は，画像中の点 \mathbf{y} における回転角 θ を投影する関数である。

まとめると，角膜表面反射画像中の点から透視投影中の点に変換する関数 \mathbf{W} は次式で与えられる。

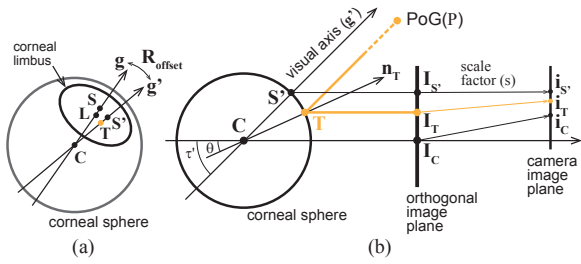


図7 視線の方向, およびカメラと眼球の関係 [6]

$$\mathbf{W}_i(\mathbf{p}) \equiv \frac{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{K}_s \mathbf{R}_i^{-1} \mathbf{L}(\mathbf{p})}{\mathbf{K}_s \mathbf{R}_i^{-1} \mathbf{L}(\mathbf{p})} \quad (i = 1, \dots, M)$$

3.4 RANSAC を用いた画像の位置合わせ

各行列 \mathbf{W}_i の妥当性を評価するため, RANSAC (RANDOM SAMPLE CONSENSUS)[2] を, 先ほどの各 SURF 初期対応点 M ペアに対して適用する. これを以後, **1point-RANSAC** と呼ぶことにする. これは, ある候補点で作った変換関数が他の候補点をどれだけ正しく移動させるかによって評価するものである.

角膜表面反射画像 \mathbf{I}_{eye} と, 透視投影 \mathbf{I}_n ($n = 1, \dots, 16$) における SURF 初期対応点のペア i ($i = 1, \dots, M_n$) それぞれについて, まず $\mathbf{p}_i, \mathbf{q}_i$ の組から, \mathbf{W}_i を得る. ここで, 全てのペア ($\mathbf{p}_j, \mathbf{q}_j$) ($j = 1, \dots, M_n$) について, $\|\mathbf{W}_i(\mathbf{p}_j) - \mathbf{q}_j\| < \text{threshold}$ を満たすペア j の数を $N_{\text{valid}}(i)$ とする. このとき \mathbf{W}_i の評価値 $\mathbf{E}(\mathbf{W}_i)$ を

$$\mathbf{E}(\mathbf{W}_i) = \frac{N_{\text{valid}}(i)}{M_n} \quad (i = 1, \dots, M_n)$$

と定義し, \mathbf{I}_{eye} と, \mathbf{I}_n の対応に適するものとして

$$\mathbf{W}_{\mathbf{I}_n} = \arg \max_{\mathbf{W}_i} \mathbf{E}(\mathbf{W}_i) \quad (i = 1, \dots, M_n)$$

を満たす行列 $\mathbf{W}_{\mathbf{I}_n}$ を選択する.

これを各透視投影 \mathbf{I}_n ($n = 1, \dots, 16$) に対して行い, 最も高い評価値を与えた $\mathbf{W}_{\mathbf{I}_n}$ を結果とする. つまり, 角膜表面反射画像に最も近いと推測される画像 \mathbf{I}_n^* および, \mathbf{I}_{eye} から \mathbf{I}_n^* への座標変換関数 \mathbf{W}^* は次のように表される.

$$\mathbf{I}_n^* = \arg \max_{\mathbf{I}_n} \mathbf{E}(\mathbf{W}_{\mathbf{I}_n}) \quad (n = 1, \dots, 16)$$

$$\mathbf{W}^* = \mathbf{W}_{\mathbf{I}_n^*}$$

3.5 角膜表面反射画像における注視点の推定

本節は [6] と同様の手法を用いている.

本節の目的は, 注視点 Point of Gaze(PoG) の角膜表面における位置, Gaze Reflection Point(GRP) \mathbf{T} の角膜表面反射画像上の座標 \mathbf{i}_T を得ることである.

図6はカメラが映す映像と眼球姿勢との関係, そして図7は別視点のもの, さらに角膜中心の表面と垂直な方向 (Eye

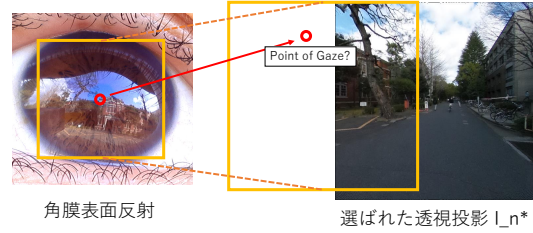


図8 GRP の投影先が透視投影に存在しない例

Optical Axis) \mathbf{g} と, 実際の視線の方向 (Visual Axis) \mathbf{g}' の違い, およびその2つの関係について説明したものである. この視線の方向の違いには個人差があるが, その関係は個人差パラメータ $\mathbf{R}_{\text{offset}}$ を用いて, $\mathbf{g}' = \mathbf{R}_{\text{offset}} \mathbf{g}$ で表すことができる. 眼球は球であることを仮定しているが, 角膜表面反射画像は2次元画像である. そこで, 眼球の表面を平面に仮想的に投影した Orthogonal image plane (図7-(2)(b)) を考えると,

$$\begin{aligned} \mathbf{C} \cdot \mathbf{n}_T &= \begin{bmatrix} \cos \tau' & \sin \tau' \end{bmatrix} \cdot \mathbf{n}_T, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \end{bmatrix}, \\ \mathbf{n}_T &= \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix}, \quad \tau' = \arccos(g'_z), \\ \mathbf{g} &= \begin{bmatrix} g_x & g_y & g_z \end{bmatrix}^T, \quad \mathbf{g}' = \begin{bmatrix} g'_x & g'_y & g'_z \end{bmatrix}^T. \end{aligned}$$

また, 角膜の角度 θ は τ' を使って表すことができ,

$$\theta = \arctan((1 - \cos \tau') / \sin \tau'). \quad (2)$$

である. ここで, $|\mathbf{I}_T - \mathbf{I}_C| = r_C \sin \theta$ であり, Orthogonal image plane とカメラのイメージはスケール変化の関係にあることを考えると, GRP の座標 \mathbf{i}_T は

$$\mathbf{i}_T = \mathbf{i}_L + s \left(-d_{LC} \begin{bmatrix} g_x & g_y \end{bmatrix}^T + r_C \sin \theta \begin{bmatrix} g'_x & g'_y \end{bmatrix}^T \right). \quad (3)$$

である. ただし, s は角膜の中心 \mathbf{L} を通る Orthogonal image plane と, 角膜表面反射画像間のスケール倍率である.

3.6 注視点の投影と画像方向修正

ここまでの節で,

- 3.5 節で決定した角膜表面反射画像の GRP
- 3.4 節で決定した座標変換関数 $\mathbf{W}^* = \mathbf{W}_{\mathbf{I}_n^*}$

が得られるので, 次は GRP を透視投影 \mathbf{I}_n^* に投影し, 注視点の位置 Point of Gaze を得たい. しかし 3.4 節の結果によっては, 透視投影の方向が微妙に違うために, GRP の投影先が \mathbf{I}_n^* に存在しないことが時折起こる (図8). これは, 投影先の x 座標を見ることで判別することができる. 透視投影は $N \times N$ ピクセルなので, 投影先の点の x 座標が 0 未満ならば, 透視投影よりも左にあり, N より大きいならば, 右にある. そこで, それぞれのケースに合わせて, 用いる透視投影を変更して再度投影することで, この問題を解決する. 透視投影の画角は 90° である. ここで, 投影先の点が \mathbf{I}_n^* に存在しないとき, 透視投影の中心と比較して

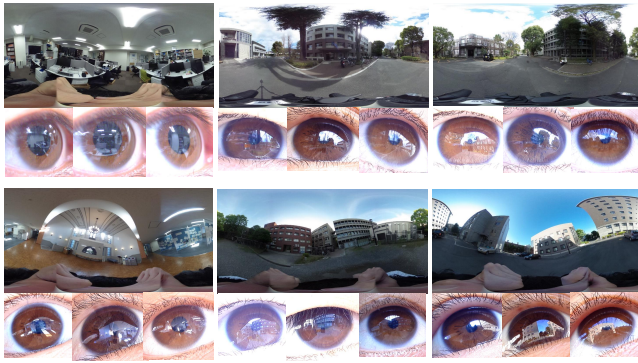


図 9 実験に用いた全地球画像と角膜表面反射の一例

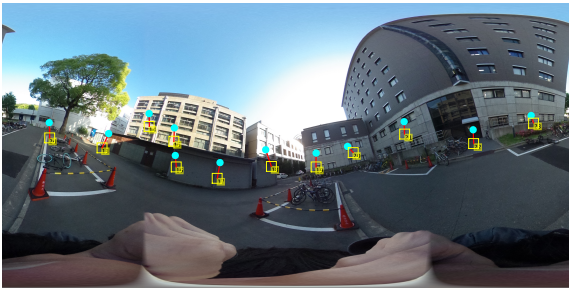


図 10 個人誤差を考慮しない、屋外 5 における推定注視点 (黄色枠), 水色丸は実際の注視点, 赤線は誤差

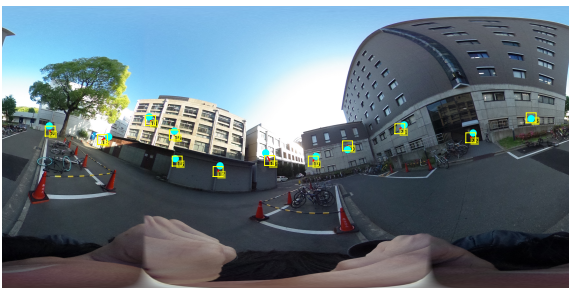


図 11 個人誤差を考慮した、屋外 5 における推定注視点 (黄色枠), 水色丸は実際の注視点, 赤線は誤差

45 度以上異なる場所に注視点が存在する。本研究では、全地球画像を 16 方向の透視投影に変換しているため、2 つ離れた透視投影を選択して再投影すれば、投影先が透視投影の外に存在することはなくなる。まとめると、以下の条件のとき、用いる透視投影と座標変換関数を変更する。

- GRP を I_n^* に投影した際、 x 座標が 0 未満だったとき

$$I_n^* \rightarrow I_{(n+2)}^*, \quad W^* \rightarrow W_{I_{(n+2)}^*} \quad (n = 1, \dots, 16)$$

- x 座標が N より大きかったとき

$$I_n^* \rightarrow I_{(n-2)}^*, \quad W^* \rightarrow W_{I_{(n-2)}^*} \quad (n = 1, \dots, 16)$$

に変更して再度注視点の推定を行う。ただし、 n は 1 から 16 で循環するものとする。

4. 実験

4.1 実験内容と結果

まとめると、本研究の注視点推定は、大きく次の 2 つの段階で構成されており、これらを検証する必要がある。

表 1 個人誤差を考慮した実験結果

Scene	User #	画像数	位置合わせ 成功率	鉛直方向	
				水平方向 平均誤差 (度)	鉛直方向 平均誤差 (度)
屋内 1	1	28	28(100.00%)	5.056($\sigma = 2.795$)	3.931($\sigma = 4.110$)
屋内 2	1	10	5(50.00%)	2.565($\sigma = 2.682$)	2.322($\sigma = 1.745$)
屋内 2	2	10	7(70.00%)	3.313($\sigma = 2.817$)	1.927($\sigma = 1.911$)
屋内 2	3	10	6(60.00%)	1.770($\sigma = 1.324$)	1.580($\sigma = 1.127$)
屋外 1	1	10	10(100.00%)	2.406($\sigma = 1.501$)	3.929($\sigma = 2.860$)
屋外 1	2	10	8(80.00%)	1.579($\sigma = 1.173$)	2.426($\sigma = 2.701$)
屋外 1	3	10	10(100.00%)	1.861($\sigma = 0.921$)	5.258($\sigma = 3.986$)
屋外 2	1	10	8(80.00%)	2.364($\sigma = 1.340$)	2.279($\sigma = 1.901$)
屋外 2	2	10	8(80.00%)	2.973($\sigma = 3.019$)	3.197($\sigma = 2.146$)
屋外 2	3	10	8(80.00%)	2.034($\sigma = 1.525$)	2.264($\sigma = 1.984$)
屋外 3	1	12	11(91.67%)	1.247($\sigma = 0.741$)	2.490($\sigma = 2.287$)
屋外 3	2	12	11(91.67%)	1.798($\sigma = 1.522$)	4.866($\sigma = 3.859$)
屋外 3	3	12	8(66.67%)	1.116($\sigma = 0.780$)	3.692($\sigma = 2.980$)
屋外 4	1	10	7(70.00%)	3.361($\sigma = 2.621$)	3.639($\sigma = 4.047$)
屋外 4	2	10	10(100.00%)	2.808($\sigma = 1.641$)	1.876($\sigma = 1.716$)
屋外 4	3	10	10(100.00%)	1.748($\sigma = 1.862$)	1.964($\sigma = 1.676$)
屋外 5	1	12	10(83.33%)	1.785($\sigma = 0.921$)	1.340($\sigma = 1.088$)
屋外 5	2	12	12(100.00%)	1.529($\sigma = 1.373$)	2.657($\sigma = 1.235$)
屋外 5	3	12	9(75.00%)	3.050($\sigma = 2.845$)	4.023($\sigma = 2.780$)

表 2 個人誤差を考慮しない場合とした場合の比較

Scene	User #	個人誤差非考慮		個人誤差考慮	
		水平方向 平均誤差	鉛直方向 平均誤差	水平方向 平均誤差	鉛直方向 平均誤差
屋内 1	1	5.609°	6.282°	5.056°	3.931°
屋内 2	1	3.621°	4.693°	2.565°	2.322°
屋内 2	2	4.355°	7.604°	3.313°	1.927°
屋内 2	3	5.674°	6.045°	1.770°	1.580°
屋外 1	1	2.410°	7.798°	2.406°	3.929°
屋外 1	2	0.890°	9.065°	1.579°	2.426°
屋外 1	3	5.361°	7.523°	1.861°	5.258°
屋外 2	1	5.690°	7.978°	2.364°	2.279°
屋外 2	2	5.334°	10.533°	2.973°	3.197°
屋外 2	3	3.864°	10.114°	2.034°	2.264°
屋外 3	1	1.322°	4.114°	1.247°	2.490°
屋外 3	2	4.299°	5.647°	1.798°	4.866°
屋外 3	3	6.229°	4.919°	1.116°	3.692°
屋外 4	1	2.211°	4.041°	3.361°	3.639°
屋外 4	2	3.183°	5.432°	2.808°	1.876°
屋外 4	3	4.783°	5.146°	1.748°	1.964°
屋外 5	1	1.606°	4.078°	1.785°	1.340°
屋外 5	2	1.372°	8.032°	1.529°	2.657°
屋外 5	3	3.030°	8.526°	3.050°	4.023°

- (1) lpoint-RANSAC による評価で、角膜表面反射画像と 16 方向の透視投影のうちから、1 つの対応点を導き、位置合わせを行う。このとき、対応点は正しくなければならない。
- (2) 対応点から得られた座標変換関数と、眼球モデルから導いた Gaze Reflection Point(3.5 節)を用いて、全地球画像上の注視点を推定する。

屋内 2 シーン、屋外 5 シーンにおいて、角膜イメージングカメラ (図 2) を用いて、1~3 名の被験者に 10~28 点の指定した点を注視してもらい、その角膜表面反射画像を撮影し (図 9)、(1) の位置合わせの成功率を確認した。また、そのうち対応点導出が成功したものについて、(2) の注視点推定を行ったときの、全地球画像上の実際の注視点との誤差を確認した。

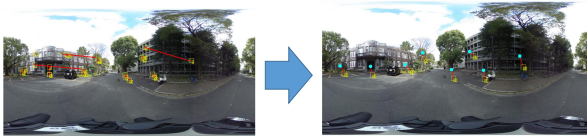


図 12 屋外 2 における、画像方向修正前 (左) と修正後 (右) の注視点推定結果

さらにこのとき、個人誤差に関するキャリブレーションを行い、全体の注視点推定誤差がどの程度縮まるかを検証した。

実験結果を表 1, 表 2 に示す。位置合わせの成功率は、屋外はおよそ 80~90%程度の精度であったが、屋内では幅があり、屋外よりも成功率が低くなるが多かった。また、位置合わせが成功したものについて注視点推定を行った結果は、個人誤差を考慮することにより、1~5 度程度であることが確認できた。

4.2 考察

透視投影と角膜表面反射画像の位置合わせが正しくなければ、正しい注視点を推定することができないことは既に述べた。つまり、注視点の推定で重要となる要素の 1 つに、1point-RANSAC の精度がある。位置合わせの結果の出力は、複数の候補の中から、3.4 節で扱った評価値が最大のものを選ぶ。今回の実験の際に、プログラム実行の様子を見ると、「正しい結果が出た位置合わせ」の中でも、他の候補点を大きく上回って出力した正解と、そうでないものに分かれていた。例えば、屋内 1 ではどの正解点の組も非常に良い評価値を出していたが、屋内 2 では、そうでない正解点の組が多く、「かろうじて」正解を出力することができたものがいくつか見られた。屋外の実験でも同様の現象が見られた。屋外 1 の全天球画像の中心にある建物は距離が近く、窓なども多くあったため、その方向を見ていた画像については注視点推定は非常に安定していたが、それ以外の方向では位置合わせが失敗することもあった。このことから、画像中の特徴物の多さが、1point-RANSAC の精度に深く関わっていることがわかる。したがって、1point-RANSAC は、特徴物がよく写っている角膜表面反射画像に対しては、位置合わせの成功が良く望める一方で、何も特徴がないもの (例えば、単色の壁、床、空など) が大半を占める角膜表面反射画像に対しては、位置合わせを成功させにくいと考えられる。

また、本研究では画像方向の修正を行ったが、これを行うことにより、注視点推定の精度が図 12 のように向上した。これは、透視投影の中に対応点があるという前提で変換関数を作っているため、それに適するように方向を修正したからであると考えられる。

見ている方向の明るさや、画像の明るさも位置合わせ精度に関わっていると考えられる。屋外 1 に比べて暗い場所

が多い屋外 2, 屋外 3 の実験では、屋外 1 に比べて位置合わせが失敗する点が多かった。これに加えて、角膜表面反射画像には多数のノイズが含まれている。角膜の光彩、まぶたが角膜表面反射画像に映りこんでいることや、眼球への光の入射量や光源の種類 (電灯、太陽など) は位置合わせの精度に関わっていると言える。

5. 結論

本稿では、全天球画像を用いて全周型注視点推定を行った。角膜表面反射画像のみを用いて眼球モデルの構築を行い、全天球画像の分割による、平面画像への変換、角膜表面反射画像と、全天球分割画像の位置合わせによる、2 画像の組み合わせの決定を通して、注視点を導いた。実験により、2 画像の位置合わせにおいて高い成功率を出すことができる可能性、およびそこから推定された注視点の妥当性が示唆された。しかし一方で、環境の違いや角膜表面反射画像のノイズの多さにより、位置合わせの成功率の低下を招き、その結果注視点の推定の成功率が低下してしまうケースも見られた。したがって、2 画像の位置合わせの側面では、鮮明な角膜表面反射部の安定した取得と、ノイズに対処する手法の検討が今後の課題として挙げられる。注視点推定の側面から見ると、眼球姿勢や、注視点の軸 (Visual Axis) には個人差により多少の誤差があることが示唆された。そこで、より精度の高い注視点推定のため、個人差の修正パラメータを導く手法の検討が今後の課題として挙げられる。

全周型の注視点推定は、従来の平面に対しての注視点推定から一歩広がり、3 次元的な概念が導入されている。注視点推定の範囲を拡大することができれば、将来的には角膜表面反射を通じて、3 次元空間における注視点推定、および注視点マップの生成が可能になることが期待される。これを実現するにあたって、ある複数の角膜表面反射画像から得られる差分から、注視点の距離を推定することなどをアプローチとして、3 次元的な位置を把握していきたい。

現実空間に即した注視点推定を行うことで、特に注視点推定や視線を扱う研究分野や、店内全体での注視点推定、マーケティングへの応用など、商業的分野においても、大きな利用価値が生まれることが期待されるから、角膜表面反射画像による注視点推定の応用可能性は十分に残されているといえる。

謝辞 本研究は、文武科学省科学研究費 26280058, 26249029, 15H02738 の援助で行われた。

参考文献

- [1] Baya, H., Essa, A., Tuytelaars, T. and Van Gool, L.: Speeded-up robust features (SURF), *Computer vision and image understanding*, Vol. 110, No. 3, pp. 346-359 (2008).
- [2] Fischler, M. A. and Bolles, R. C.: Random sample con-

- sensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395 (1981).
- [3] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, Vol. 60, No. 2, pp. 91–110 (2004).
 - [4] Nakazawa, A. and Nitschke, C.: Point of gaze estimation through corneal surface reflection in an active illumination environment, *Computer Vision–ECCV 2012*, Springer, pp. 159–172 (2012).
 - [5] Nakazawa, A., Nitschke, C. and Nishida, T.: Random re-sample consensus 法を用いた角膜表面反射とシーン画像の位置合わせ, 第 17 回画像の認識・理解シンポジウム (MIRU2014) (2014).
 - [6] Nakazawa, A., Nitschke, C. and Nishida, T.: Non-calibrated and real-time human view estimation using a mobile corneal imaging camera, *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on*, IEEE, pp. 1–6 (2015).
 - [7] Nishino, K., Belhumeur, P. N. and Nayar, S. K.: Using eye reflections for face recognition under varying illumination, *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, Vol. 1, IEEE, pp. 519–526 (2005).
 - [8] Nishino, K. and Nayar, S. K.: Corneal imaging system: Environment from eyes, *International Journal of Computer Vision*, Vol. 70, No. 1, pp. 23–40 (2006).
 - [9] Nitschke, C., Nakazawa, A. and Takemura, H.: Corneal imaging revisited: An overview of corneal reflection analysis and applications, *Information and Media Technologies*, Vol. 8, No. 2, pp. 389–406 (2013).