

角膜表面反射の画像特徴を用いた視線停留点検出

北角 一哲^{1,a)} 中澤 篤志¹ 西田 豊明¹

概要: 本研究では、角膜表面反射画像の局所特徴量を利用して、眼球映像中の瞬目と視線停留状態を検出することを目標とする。局所特徴量を用いたアルゴリズムの一つに Bag of Features がある。Bag of Features は一枚の画像を画像中の局所特徴量の出現頻度ヒストグラムで表現する手法であり、本研究ではこれを用いる。眼球映像の全フレームをヒストグラムで表現し、フレーム間のヒストグラムの類似度を求める。類似度を示す指標に Histogram Intersection を用いる。本手法は、瞬目では類似度が小さくなり、停留状態では大きくなることを利用して、類似度の値について閾値を設定し検出を行う。また視線停留点を検出するために、角膜イメージング法を用いて注視点推定を行う。実環境中での眼球映像を取得し、上記の手法で検出を行い、Precision, Recall, F-measure による精度評価を行ったところ、瞬目に関しては検出精度が高く本手法が非常に有効であることが示された。視線停留状態の検出に関しては、本手法が有効であることが言えるが、より精度を向上させることが今後の課題となる。

キーワード: 角膜表面反射, 瞬目, 停留状態, Bag of Features

Detecting fixation points of gaze by using image feature of corneal reflection

KAZUNORI KITAZUMI^{1,a)} ATSUSHI NAKAZAWA¹ TOYOAKI NISHIDA¹

Abstract: In this paper, we aim at detecting blinking and fixation in eye movies, using local features of corneal image. Bag of Features is one of the algorithms which uses local features. Bag of Features is a method that transforms an image into histogram of local feature frequency. We use Bag of Features. We transform all frames in eye movie into histograms and calculate degree of similarities between frames. We use Histogram Intersection as an index of similarities. To detect blinking and fixation, we determine the threshold value of Histogram Intersection. On the other hand, we estimate point of gaze by corneal imaging technique to detect fixation points of gaze. We record eye movies in actual environment and evaluate accuracy of detection. As a result, it is shown that this method is very useful for blinking detection. For fixation detection, this method is also useful but next issue is to improve accuracy of detecting fixation.

Keywords: corneal surface reflection, blinking, fixation, Bag of Features

1. 序論

人間の生活・行い・体験を映像・音声・位置情報などのデジタルデータとして記録する技術をライフログという。現在、ライフログに関するアプリケーションやサービスが多数存在している。例えばどこで何を食べたかを記録する食

ベログ、いつどのような本を読んだかを記録する読書ログなどのサービスが存在する。そして人々が利用したお店や読んだ本の情報が、インターネット上で共有できる。中村らの研究 [13] では、ライフログが備えるべきデータを分析し、ライフログのための標準的なデータモデルの構築を行っている。

ライフログの用途として、人の記憶の補助や体験の共有などが主に挙げられる。日常生活において人がいつ・何を見たかを時系列的に記録することは記憶の補助だけでなく、

¹ 京都大学
Kyoto University

^{a)} kitazumi@ii.ist.i.kyoto-u.ac.jp

マーケティングにも有用である。例えばコンビニに入ってから出るまでに見たものを記録し分析することで商品の配置に利用できるなど、応用の幅は広い。

従来の視線に関するライフログの生成にはウェアラブルカメラ等を用いて撮影された一人称視点映像が用いられることが多い。一人称視点映像とは頭部などに装着されたカメラから撮影されるその人が見ている風景の映像である。人が実際に見るシーンとほぼ同じシーンが映像として得られることから、日常生活における行動の記録や分析に有用である。実際に一人称視点映像が用いられた研究 [4],[5],[6] が様々行われている。一人称視点映像を用いる大きな問題点としてプライバシーの問題が挙げられる。そこでこれに対して頭部に装着したカメラから眼球の動きを撮影した映像を用いることで、見ている風景が角膜に映る範囲に制限されることによりプライバシーの問題を大きく削減できる。さらに一人称視点映像では正確な注視点を推定することができないという問題も、角膜表面反射映像を用いることで解決できる。nakazawa らの研究 [9] では角膜イメージングカメラを用いて、キャリブレーションが不要でリアルタイムに注視点を推定する手法を提案している。

視線計測に関する研究に、Bulling らの研究 [2] がある。Bulling らは、Electrooculography(EOG) と呼ばれるデバイスを用いて、眼球の動きを計測している。EOG は目の前面の角膜と後面の網膜の間の電位差を計測し、この電位差が眼球の回転角と比例するのを利用して眼球の動きを計測する。しかし、目を動かさずに顔や体のみを動かして注視する対象を変える場合もあり、こういった動作を EOG では認識することができない。角膜表面反射映像を用いた場合には顔や体のみを動かして注視対象を変えたとき、角膜に映るシーンも変わるため識別することが可能である。

人が見たものを時系列的に記録するライフログを生成する際に、視線がある対象に停留している箇所を検出する必要がある。ここで人の眼球運動として視線停留状態、瞬目、サッケード(跳躍運動)がある。視線とは、目がある対象や点を見ている際の眼球と視対象物間を結ぶ線を指す。サッケード(跳躍運動)とは、人が見る対象を替える際に発生する眼球運動である。瞬目中やサッケード中はほとんど物体を認識していないとされている。そのため、瞬目とサッケードを検出することはライフログ生成には重要である。特に瞬目の検出はライフログ生成だけでなく、運転中のドライバの状態の検知など様々な分野に応用できる。保坂らの研究 [14] では、瞬目の頻度と眠気状態が対応することを明らかにしている。

画像認識とは画像をパターンに基づいて認識・分類する技術である。特別な視線計測デバイスを用いず、角膜表面反射映像から眼球運動を識別するにあたり、画像認識の技術に着目した。画像認識の技術を用いることで角膜表面反射映像を瞬目や視線停留状態を認識し、分類できると考えた。

そこで、本研究では視線ライフログを生成する際に重要な瞬目と視線停留点を、眼球映像から得られる角膜表面反射映像から画像認識の技術を用いて検出することを目標とする。

2. 関連研究

本章では視線検出や視線解析、画像認識に関する関連研究を挙げる。

2.1 視線検出

視線検出に関する研究には、西田らの研究 [11] が挙げられる。西田らは、人の眼球運動から強い興味や関心などが反映された視点の検出を行うため、一定の停留時間条件を満たす停留点の発生頻度に着目した視点検出の方法を提案した。視線計測では、被験者の眼球運動を、赤外 LED を用いた工学的な手法により測定している。竹上らの研究 [12] では、視線方向の変化に伴って角膜における光源の反射像の位置が虹彩領域内で変化することに着目し、虹彩領域内の角膜反射像の相対位置と視線方向との関係式に基づき、画像処理により視線方向を推定している。

2.2 視線解析

視線解析に関する研究では、視線の動きを用いた行動分析に関するものが多い。Shiga ら [10] は、日常生活における人の行動の認識を、視線の動きと画像特徴を組み合わせることで、視線の動きのみまたは画像特徴のみを用いたときよりも認識の精度が向上することを示した。視線計測では Eye Tracking Glass(ETG) と呼ばれる視線計測デバイスを用いている。また、Bulling ら [2] は、Electrooculography(EOG) と呼ばれる眼電位図を記録するゴーグル型のデバイスを用いて、日常生活における人の行動の認識を行った。EOG は目の前面の角膜と後面の網膜の間の電位差を計測し、この電位差が眼球の回転角と比例するのを利用して眼球運動を計測する。

現在では [10],[2] の研究のように視線計測デバイスにより計測された視線の動きを用いた行動分析に関する内容の研究が多く、視線計測自体に焦点を当てた研究は少ない。また、EOG は眼球運動を計測するのに非常に有用なデバイスであるが、1 章でも述べたように、目を動かさずに体や顔を動かして注視対象を変える場合には、EOG はそのことを認識できない。こうした問題は角膜表面反射映像を用いることで解決できると考える。

2.3 画像認識

現在、画像認識の分野の研究が盛んに行われている。その中でも一般物体認識の技術に関する研究が急速に進んでいる。一般物体認識とは制約のない実世界シーンの画像に対して、コンピュータが画像中に含まれる物体を一般的な名

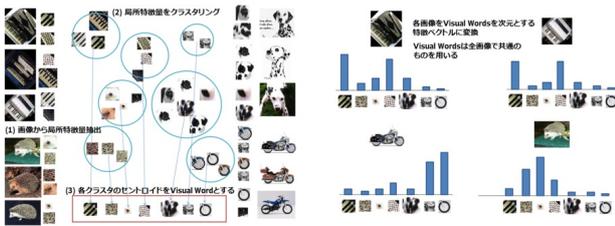


図 1 Bag of Features のアルゴリズム
Fig. 1 Algorithm of Bag of Features

称で認識する技術である。その他にも顔認識や人検出などの研究が行われており、現実に数多くのシステムやアプリケーションが存在する。例えば、夜間の運転中に見えにくい前方の歩行者や障害物を検知し、ディスプレイに表示して知らせるトヨタのナイトビューシステムや、携帯で撮影した写真などから被写体やテキストを解析して検索するツールである Google Goggles などがある。本章では、画像認識を行う際によく用いられる画像の局所特微量と関連研究、この局所特微量を応用した画像特徴表現である Bag of Features について述べる。

2.3.1 局所特微量

上でも述べたように画像認識には局所特微量がよく用いられる。局所特微量とは、特徴点検出法により画像中の濃淡の変化が大きい特徴点を検出し、その特徴点周りの領域を画素値や微分値により特徴ベクトルにしたものである。局所特微量としては SIFT 特徴 (Scale Invariant Feature Transform), SURF 特徴 (Speeded-Up Robust Features), HOG 特徴 (Histograms of Oriented Gradients) などがある。

2.3.2 Bag of Features

Bag of Features は画像を局所特徴の出現頻度ヒストグラムで表現する手法であり, Bag of Visual Words や Bag of Keypoints と表記されることも多い。2.3 章の最初で触れたように一般物体認識においてよく用いられる。Bag of Features はもともと自然言語処理でよく使われる Bag of Words を画像へ応用した手法である。Bag of Words とは文書を単語の集合として扱い、既知の文章から特徴語を抽出し、各特徴語の頻度を文書ベクトルの値として文書を表現する方法である。ここで画像において文書の単語に対応するものは前章で取り上げた局所特微量である。Bag of Features は Visual Words と呼ばれる代表的な局所特微量を抽出し、画像を Visual Words の出現頻度ベクトルで表現する方法である。図 1 は Bag of Features のアルゴリズムを簡潔に示したものである。Bag of Features では Visual Words を構築する画像集合を準備し、それぞれの画像から多数の局所特微量を抽出する。図 1 では局所特微量を小さな画像により表現しているが実際には SIFT 特徴や SURF 特徴の 128 次元ベクトルである。次にすべての特徴ベクトルををクラスタリングする。クラスタリングには非階層型クラスタリングのアルゴリズムである K-means 法がよく用いられ

る。K-means 法などにより K 個のクラスタに分割した後に各クラスタのセントロイド (重心) を求める。このセントロイド一つ一つを以降 Visual Word として扱う。Visual Word も SIFT 特徴や SURF 特徴同様 128 次元の特徴ベクトルである。Visual Word の数は K-means 法でクラスタリングしたクラスタ数 K である。そして各画像を Visual Words を次元とするヒストグラムに変換する。これは画像中の各局所特微量について一番近い Visual Word を求め、その Visual Word に投票することで行う。一番近い Visual Word を求める際にはユークリッド距離やマンハッタン距離が用いられることが多い。

Bag of Features に関する研究には Jegou らの研究 [7] がある。Jegou らは, Bag of Features のアルゴリズムを改良した画像検索手法を提案し、検索精度を向上させている。一般物体認識などで Bag of Features が用いられる場合には、上で述べたアルゴリズムによって作成されたヒストグラム同士を比較して一般物体認識を行う。ヒストグラムを比較する際の尺度としてよく用いられるものに Histogram Intersection がある。Histogram Intersection は 2 つの正規化されたヒストグラムが与えられたときの類似度 D が以下の式で求められる。

$$D = \sum_{i=1}^n \min\{a_i, b_i\} \left(\sum_{i=1}^n a_i = 1, \sum_{i=1}^n b_i = 1 \right)$$

ここで, D は 2 つのヒストグラムの類似度, a_i, b_i はそれぞれヒストグラム A, B の i 番目のピンの値, n はピンの個数を表す。2 つのヒストグラム A, B が類似しているほど, D の値は大きくなり, A, B のヒストグラムが一致しているとき D の値は 1 となる。このときヒストグラム A, B は括弧内の条件を満たす必要がある。

画像特徴を用いることによって画像の局所的な類似性を求め画像認識を行っている。そのため、瞬目やサッケードなどによる、目の開き具合・角膜に映り込む風景・瞳の位置などの変化を画像特徴を用いることで認識できると考えられる。これまで視線の計測に画像認識の技術を用いた例は見つからない。そこで本研究では、角膜表面反射画像に画像認識の技術を適用し、瞬目や視線の停留状態を検出することを試みる。

3. 角膜表面反射の画像特徴を用いた視線停留点検出

3.1 システムの概要

まず、数分間の眼球映像を目を撮影するウェアラブルカメラを用いて取得する。この際眼球映像はフレームごとに 1 枚の角膜表面反射画像として保存される。そして、取得した画像に以下の処理を行う。

- (1) 瞳の輪郭追跡と注視点推定
- (2) 瞬目の検出

(3) 視線の停留状態の検出

推定された注視点と検出された停留状態を用いることで視線停留点の検出を行うことができる。

瞳の輪郭追跡と注視点推定に関しては、画像特徴のみで実装するのは難しいため、nakazawa らの研究 [9] で提案された手法を用いる。瞬目の検出と視線の停留状態の検出には 2.3 章で述べた Bag of Features を用いる。

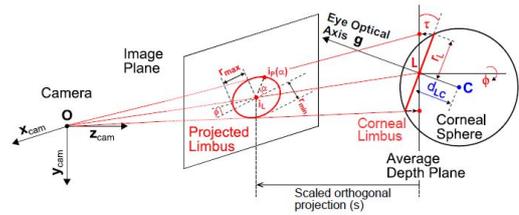


図 2 投影された角膜からの 3 次元姿勢推定

Fig. 2 Pose estimation from projected cornea

3.2 瞳の輪郭追跡と注視点検出

瞳の輪郭追跡と注視点の推定には、nakazawa らの研究 [9] で提案された手法を用いる。この手法では角膜表面反射画像列のみを用いてリアルタイムでの注視点の推定を行っている。また目の画像のみを扱うため、キャリブレーションが不要である。瞳の輪郭追跡と注視点推定の手法を述べる。

3.2.1 まぶたの検出と追跡

眼球の 3 次元姿勢は幾何学的な眼球モデルと粒子フィルタを用いて角膜画像列から追跡する。2 次元空間の角膜検出の技術は既にいくつか存在するが、この手法では注視点推定の精度を向上させるため、3 次元眼球モデルを用いている。まず、目の画像に対し、目尻と目頭の位置を手動で指定する。次にまぶたの境界と、瞳輪郭部を分け、眼球姿勢推定の精度を上げるために、まぶたの輪郭を追跡する。上まぶたと下まぶたの位置を $i_j^{UE}, i_j^{LE} (j = 1, \dots, N^E)$ とすると、評価関数は次の式 (1) で表される。

$$Eval^E = - \sum_{j=1}^{N^E} \dot{I}(i_j^{UE}) \cdot \frac{\ddot{i}_j^{UE}}{|\ddot{i}_j^{UE}|} - \sum_{j=1}^{N^E} \dot{I}(i_j^{LE}) \cdot \frac{\ddot{i}_j^{LE}}{|\ddot{i}_j^{LE}|} \quad (1)$$

ここで、 $\dot{I}(x)$ は位置 x での入力画像の勾配で、 $\ddot{i}_j^{UE}, \ddot{i}_j^{LE}$ はそれぞれ、上まぶたと下まぶたの境界線の二次導関数である。まぶたの形によって、 $\ddot{i}_j^{UE}, \ddot{i}_j^{LE}$ は目の内側の方法を示す。この評価関数が最大となる点群を粒子フィルタを用いて探し、まぶたの輪郭を推定する。

3.2.2 リアルタイムでの 3 次元眼球姿勢の追跡

眼球の 3 次元姿勢推定と光の反射を計算するために固定したサイズの眼球モデルを用いる。図 2 は角膜表面をカメラの画像平面に投影したものである。眼球モデルの姿勢は円形の角膜の姿勢によって決まる。角膜は中心点 L と法線ベクトル g で表される。角膜は楕円の形をしており、楕円を表現するパラメータは、中心点 i_L 、長径と短径 r_{max}, r_{min} 、回転角 ϕ の 5 つである。光軸 g は $g = [-\sin \tau \sin \phi \quad \sin \tau \cos \phi \quad -\cos \tau]^T$ の式で得られ、角度 τ は $\tau = \pm \arcsin(r_{min}/r_{max})$ の式で得られる。スケール s は、 $s = r_{min}/r_L$ で計算され、投影された角膜の中心 i_C と角膜輪郭 i_P は以下の式 (2), (3) で表される。

$$i_C = i_L + s d_{LC} \begin{bmatrix} \sin \tau \sin \phi \\ -\sin \tau \cos \phi \end{bmatrix} \quad (2)$$

$$i_P(\alpha) = i_L + s r_L \begin{bmatrix} \cos \phi \cos \alpha - \sin \phi \cos \tau \sin \alpha \\ \sin \phi \cos \alpha + \cos \phi \cos \tau \sin \alpha \end{bmatrix} \quad (3)$$

ここで α は $\{\alpha | 0 \leq \alpha \leq 2\pi\}$ を満たす画像平面での楕円の角度、 r_L と r_C はそれぞれ $6.0\text{mm}, 7.7\text{mm}$ で、 $d_{LC} = \sqrt{r_C^2 - r_L^2}$ である。これにより、瞳輪郭を求めることができる。

3.2.3 注視点推定

得られた眼球の 3 次元姿勢を用いて、見ている場所が角膜反射画像中で反射している点を検出する。図 3(a) は目の光軸 g と視線方向 g' とそれぞれが角膜表面と交わる点 S, S' を示している。目の光軸と実際の視線方向はわずかにずれており、この誤差は人それぞれ異なる。また、 g と g' のずれは回転角によって表現できるため回転行列 R_{offset} を用いて視線方向は $g' = R_{offset} g$ の式で求められる。

図 3(b) は角膜表面での光の反射を示している。注視点は視線方向からの光が反射している眼球表面上の点に位置する。視線方向 $g' = [g'_x \quad g'_y \quad g'_z]^T$ としたとき、図 3(b) 中の τ' は $\tau' = \arccos(g'_z)$ で求められる。これより注視点を決める眼球の角度 θ は τ' を用いて次の式 (4) で求められる。

$$\theta = \arctan((1 - \cos \tau') / \sin \tau') \quad (4)$$

結果として直交画像平面中の眼球の中心 (I_T) と注視点 (I_C) の距離が得られ、 $|I_T - I_C| = r_C \sin \theta$ となる。角膜輪郭の中心 L が位置する平面を考えると、カメラの画像平面における注視点 i_T は以下の式 (5) で求められる。

$$i_T = i_L + s(-d_{LC} \begin{bmatrix} g_x & g_y \end{bmatrix}^T + r_c \sin \theta \begin{bmatrix} g'_x & g'_y \end{bmatrix}^T) \quad (5)$$

これより目の画像中の注視点の座標を眼球の 3 次元姿勢の追跡により得られる。

図 4 と表 1 は以上の瞳の輪郭追跡と注視点推定を行った結果の例である。図の青の実線がまぶたの輪郭、黄の点線で描かれた楕円が瞳の輪郭、黄の四角が注視点 i_T を示している。表 1 は注視点の座標などの眼球情報である。これらの眼球情報の詳細を以下に記す。

- r_{inside} : 図 4 中の黄色の丸で示した目頭の画像中の座標。左の値が x 座標、右の値が y 座標。
- $r_{outside}$: 図 4 中のピンクの丸で示した目尻の画像中の座標。左の値が x 座標、右の値が y 座標。
- c_{center} : 図 4 中の灰色の丸で示した眼球の中心位置の

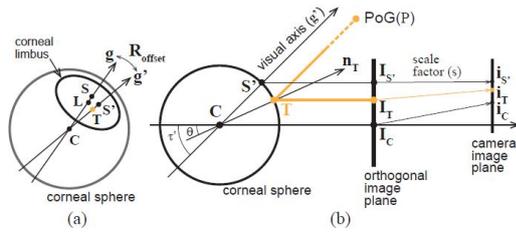


図 3 (a) 目の光軸 g と視線方向 g' の違い.(b) 注視点 T と目の画像中の位置 i_T .

Fig. 3 (a)Difference between optical axis(g) and eye direction(g'). (b)Position of Gaze Reflection Point(T) and eye in the picture(i_T).

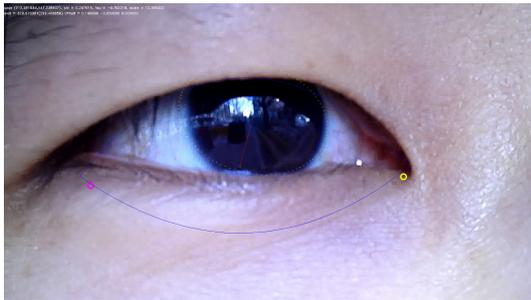


図 4 瞳輪郭追跡と注視点推定の結果.

Fig. 4 The result of corneal tracking and GRP estimation

表 1 眼球情報

Table 1 Information of eye

r_{inside}	$r_{outside}$	c_{center}	ϕ, τ
1440.0, 628.0	312.0, 660.0	850.3, 558.9	0.267, -0.762
scale	l_{center}	long-axis	short-axis
12.18	897.31, 417.32	273.04	197.49
rL	rLC	rC	grp
5.6	5.284	7.7	920.71, 394.7

座標. 左の値が x 座標, 右の値が y 座標.

- ϕ : 眼球の主に水平方向の角度を示す.
- τ : 眼球の主に垂直方向の角度を示す.
- scale: 画像中での瞳の輪郭の大きさを示す. $rL \times scale \times 4 = \text{long-axis}$ となる.
- l_{center} : 図 4 中の灰色の丸と反対の赤線の端の画像中の座標. 瞳の輪郭の楕円の中心の座標を表している. 左の値が x 座標, 右の値が y 座標.
- long-axis: 瞳の輪郭の長径を表すパラメータ.
- short-axis: 瞳の輪郭の短径を表すパラメータ.
- rL: 実際の眼球モデルでの瞳の半径. 単位は mm .
- rLC: $rLC = \sqrt{rC^2 - rL^2}$. 実際の眼球モデルでの眼球の中心から瞳までの距離.
- rC: 実際の眼球モデルでの眼球の半径. 単位は mm .



図 5 トリミングした目の画像

Fig. 5 An eye image

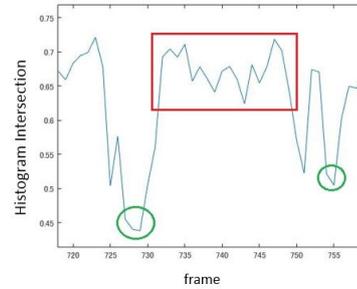


図 6 Histogram Intersection の時間変動.

Fig. 6 Time series of Histogram Intersection

3.3 瞬目の検出

瞬目は目全体が大きく動く運動である. したがって, 瞬目のフレームはその前後のフレームと, Bag of Features で作成されたヒストグラムが大きく異なる. つまり, Histogram Intersection によって求められた値が小さくなると考えられる. このことを利用して瞬目を検出する.

得られた眼球映像の, 全てのフレームに対し 3.2 章で求めた瞳の輪郭の情報と目頭と目尻の位置情報を用いて画像を図 5 のようにトリミングする. 幅は目尻の位置から目頭の位置まで, 高さは瞳の中心から上下に瞳の輪郭を表す楕円の長径の長さである. 図 5 の中央付近の小さな赤い円は注視点を表している. そして, トリミングされた画像に対して Bag of Features を適用し, ヒストグラムに変換する. つぎにフレーム間の Histogram Intersection を計算する. Histogram Intersection の値で閾値を設定し, 閾値を下回った箇所を瞬目と判定する. Histogram Intersection は 2 つの画像のヒストグラムから求められるものであるため, 閾値を下回ったときの 2 つの画像を「瞬目である」と判定してしまうと, 前後の瞬目でないフレームも含まれてしまう. そのため閾値を上回ったときの 2 つの画像を「瞬目でない」と判定するよう実装した. つまり, 2 回以上連続で閾値を下回ったときに「瞬目である」と判定される.

図 6 は Histogram Intersection の値を時系列的に表したグラフである. 縦軸が Histogram Intersection の値, 横軸が時間である. この場合, 例えば閾値を 0.55 とした場合緑で囲まれた箇所が「瞬目である」と判定される.

3.4 視線の停留状態の検出

視線が停留しているときには, 眼球の動きはほとんどなく角膜に反射するシーンもほとんど変化がない. したがっ

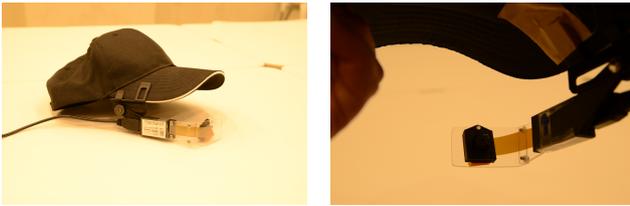


図 7 実験で使ったウェアラブルカメラ

Fig. 7 A Wearable eye camera



図 8 実験で使ったルート。

Fig. 8 The route used in this experiment.

て、停留状態にあるときのフレーム間のヒストグラムは類似し、Histogram Intersection で求められる値は大きな値をとると考えられる。このことを利用して視線の停留状態を検出する。ここで本稿では視線の停留状態を、0.6~0.7 秒以上同一の対象に視線が停留している状態とした。

視線停留状態を検出する際には、3.3 章で瞬目の検出によって「瞬目である」と判定された箇所以外のフレームを用いる。また、瞬目の検出の際にトリミングした画像と各画像のヒストグラムを用いる。そして、フレーム間の Histogram Intersection を計算する。ここで瞬目が検出された画像に対しては計算せず、その前後のフレーム間の Histogram Intersection を計算する。Histogram Intersection の値で閾値を設定し、一定数以上連続して閾値を上回った箇所を視線停留状態とする。例えば、図 6 で閾値を 0.6 とし、10 回以上連続して閾値を上回った箇所を停留状態とする場合、赤で囲った箇所が停留状態であると判定される。

4. 実験

4.1 実験概要

4 人の被験者に、まず眼球映像を撮影するための図 7 のウェアラブルカメラを装着した状態で、大学構内の指定した建物を順に探していく合計 10 分程度のタスクを課した。実験で指定した建物と、被験者が実際に通ったルートを下の図 8 に示す。図 8 で示した青い矢印が被験者が実際に通ったルートである。移動する際には、キャンパス内に配置された看板で建物の位置を確認しながら進むという条件を課した。これは視線が停留している状態を作りやすくする



図 9 角膜表面反射画像の例。

Fig. 9 Examples of corneal surface image.

ために課した条件である。撮影した眼球映像は指定した建物に到着した時点で区切っているため、2 分から 3 分程度の 4 つの眼球映像を被験者一人から得られる。

その際に得られる眼球映像はフレーム単位で画像で保存される。今回の実験では 2 人の被験者から得られた眼球映像を用いた。その際得られた眼球映像と一秒当たりのフレーム数は以下のとおりである。

- 画像サイズ: 640 × 480, 一秒あたりのフレーム数: 30 フレーム

2 人のデータを合わせて 31425 フレームを得た。その後撮影した眼球映像の中で視線が停留状態にある箇所をフレーム単位で解析するタスクを課した。15~20 フレーム続けて同じものを見ていると判断したときを停留状態として正解データを作成した。今回の実験では眼鏡をかけた人の眼球映像を撮影した場合、眼鏡に青や緑の光が反射し、映像の解析が難しいと考えられたため、裸眼またはコンタクトを装着した人を対象に実験を行った。

瞬目に関しても正解データを作っておく。本稿ではまぶたを 3 割以上閉じた状態を瞬目とし、フレーム単位で正解データを作成した。

図 9 は今回の実験で取得した角膜表面反射画像と、映り込んだ風景の通常のシーン画像である。この角膜表面反射画像を用いて視線停留点を検出することが本実験の目標である。瞬目の検出と停留状態の検出の評価に、Recall(再現率), Precision(適合率), F 尺度を用いた。

Bag of Features で用いる代表的な局所特徴量を表す Visual Words は画像データセットの caltech101 の画像を用いて作成した。Visual Word の個数は 300 個で、それぞれ 128 次元の SURF 特徴量で表現した。Visual Words を作成する際、100 万個程度の局所特徴量を K-means 法でクラスタリングするため、実行時間は非常に大きくなる。しかし一度 Visual Words を作成すれば、それ以降は作成する必要がなくなるため、Visual Words を作成する為にかかる実行時間は実験結果には考慮しない。

図 10 のように光がカメラに直接入り込んでしまっている

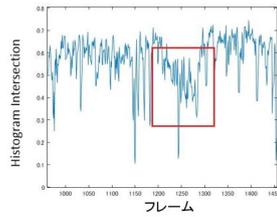
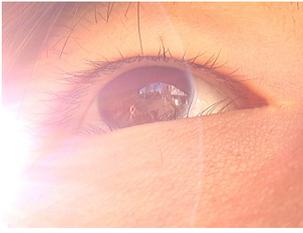


図 10 光が直接入り込んだ画像
図 11 Histogram Intersection
Fig. 10 Light comes in the camera directly
Fig. 11 Graph of Histogram Intersection

表 2 瞬目に関する Precision,Recall,F 尺度の結果.

Table 2 The result of precision, recall and F-measure for blinking

Threshold	Precision	Recall	F 尺度
$avg \times 0.8$	0.7699	0.6116	0.6817
$avg \times 0.82$	0.7521	0.6583	0.7021
$avg \times 0.84$	0.7323	0.6993	0.7154
$avg \times 0.86$	0.7107	0.7397	0.7249
$avg \times 0.88$	0.6886	0.7714	0.7276
$avg \times 0.9$	0.6555	0.7948	0.7184

画像は今回省いて実験を行った。図 11 の赤い長方形で囲った部分は図 10 のように光がカメラに直接入り込んだ区間を示しており、この区間の Histogram Intersection の値が全体的に低くなっている。そのため今回の実験のように閾値を用いて瞬目や停留状態を検出する方法では、光が入り込んだ画像で検出を行うのは難しいと判断した。これにより、実際に実験に用いた画像の全フレーム数は 28440 フレームとなった。

4.2 瞬目の検出

瞬目の検出に関して、得られたデータから Precision,Recall,F 尺度を求めた。瞬目を判定する際、区切られた 2~3 分の眼球映像全体での Histogram Intersection の平均値を求め、平均値の何倍かした値を閾値と設定した。これは、画像サイズやフレーム数、また眼球映像を撮影する環境の明るさなどにより、全体的な Histogram Intersection の値が変わるため、具体的な数値を決めずに撮影したときの環境に応じた閾値を設定している。実験の結果を表 2 に示す。表 2 は Histogram Intersection の閾値を変えたときの Precision,Recall,F 尺度の値の推移を示している。表中の Threshold は閾値を表しており、その中の avg は一つの眼球映像全体での Histogram Intersection の値である avg の具体的な値は眼球映像によって異なる。閾値を Histogram Intersection の平均値の 0.88 倍としたとき、F 尺度が最も大きい数値を示した。表 3 は瞬目フレーム数、検出フレーム数の詳細を示している。表 3 は閾値を F 尺度が最も高い値を示した $avg \times 0.88$ としたときの結果を示している。

表 3 瞬目フレーム数、検出フレーム数の詳細。閾値は $avg \times 0.88$ としている。

Table 3 Detail of the number of frames of blinking and detection

	瞬目	非瞬目	合計
検出	3533	1598	5131
不検出	1047	22262	23309
合計	4580	23860	28440

表 4 停留状態に関する Precision,Recall,F 尺度の結果.

Table 4 The result of precision, recall and F-measure for fixation

Threshold	Precision	Recall	F 尺度
$avg \times 0.82$	0.5847	0.8343	0.6876
$avg \times 0.84$	0.5908	0.8295	0.6901
$avg \times 0.86$	0.5868	0.8107	0.6808
$avg \times 0.88$	0.5920	0.7824	0.6740
$avg \times 0.9$	0.6027	0.7469	0.6671

表 5 停留フレーム数、検出フレーム数の詳細.

Table 5 The details of the number of fixation and detection

	停留	非停留	合計
検出	11690	8058	19748
不検出	3251	5441	8692
合計	14941	13499	28440

4.3 停留状態の検出

停留状態を検出する際に、瞬目を検出するときと同様に、一つの眼球映像全体の Histogram Intersection の値を求め、その平均値にある値をかけた値を閾値と設定した。停留状態とする基準は、閾値を 15 フレーム連続で超えたときに停留状態と判定する。

実験の結果を表 4 に示す。表 4 は Histogram Intersection の閾値を変えたときの Precision,Recall,F 尺度の推移を示している。表 5 は停留フレーム数、検出フレーム数の詳細を示している。閾値は $avg \times 0.88$ としている。

4.4 実験結果と考察

図 12 に Intersection の時間変動と、瞬目と停留状態の正解フレームと検出フレームを示した。図の上段は Histogram Intersection の時間変動で、中央の直線は検出する際の閾値を示しており、閾値は Intersection の平均値の 0.88 倍としている。中段上が瞬目の検出結果、中段下が瞬目の正解データを示しており、値が 1 の箇所が検出、正解フレームを示している。下段上が停留状態の検出結果、下段下が停留状態の正解データを示している。

瞬目の検出に関して図 12 から非常に高い精度で検出できていることが確認できる。停留状態の検出に関しては網羅性を示す Recall の値は高く出ているが、正確性を示す

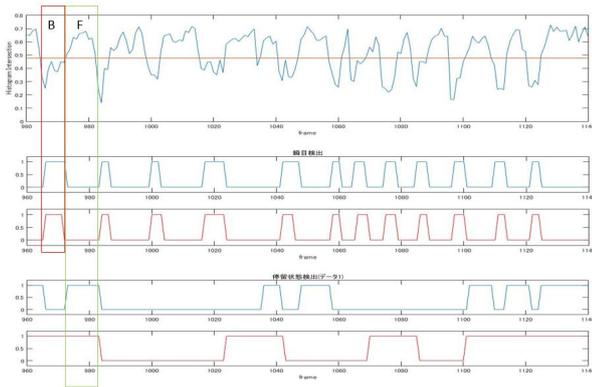


図 12 Histogram Intersection と瞬目、停留状態の関係

Fig. 12 Relationship between Histogram Intersection and blinking or fixation

Precision の値は低くなっている. 図 12 を見ると誤って検出している箇所が見られるが, 実際に視線が停留している区間を高い割合で検出できていることが確認できる. 今回の実験では大学のキャンパスという広い空間で行ったため, 見ている対象との距離があった. そのため停留状態の正解データを作成する際に, 角膜表面反射画像からは見えている対象がわかりづらく, 視線が停留しているかを判定するのが難しい. そのためコンビニやスーパーマーケットなどの限定された空間で同様の実験をして検証する必要があると考える.

5. 結論

本稿では視線ライフログを生成するために重要な, 瞬目と視線停留状態の検出を, 角膜表面反射画像のみから画像特徴を用いて行う手法を提案した. 本手法では Bag of Features によって眼球映像の各フレームでの局所特徴量の出現頻度ヒストグラムを生成し, フレーム間での眼球運動によるヒストグラムの変化に着目して, 瞬目と視線停留状態を検出することを試みた. その結果, 瞬目の検出に関して本手法が非常に有効であることが示された. 視線の停留状態の検出に関しても本手法が有効であると言えるが, 正解データの正確さを考慮すると, 実験環境を変更し改めて検証を行う必要があると考えられる.

停留状態の検出精度を向上させることで, 4.2 章での注視点推定の結果と合わせて, 視線停留点を検出できる. これによって注視対象を正確に推定することができるため, 注視対象を時系列的に解析することで, ライフログの生成にもつながる. また, 本手法が確立されれば, 瞬目の頻度や視線の停留時間などを解析することで, その人の集中度や眠気の検知など, 様々な研究分野に応用できる.

今後の課題としては, 停留状態の検出に関する実験を改めて行い, 精度を向上させる手法を検証するとともに, 視線停留点を用いた注視物体の識別などが挙げられる.

参考文献

- [1] Bay, H., Ess, 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] Bulling, A., Ward, J., Gellersen, H., Tröster, G. et al.: Eye movement analysis for activity recognition using electrooculography, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 33, No. 4, pp. 741–753 (2011).
- [3] Dalal, N. and Triggs, B.: Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, IEEE, pp. 886–893 (2005).
- [4] Doherty, A. R., Ó Conaire, C., Blighe, M., Smeaton, A. F. and O'Connor, N. E.: Combining image descriptors to effectively retrieve events from visual lifelogs, *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, ACM, pp. 10–17 (2008).
- [5] Doherty, A. R. and Smeaton, A. F.: Automatically segmenting lifelog data into events, *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*, IEEE, pp. 20–23 (2008).
- [6] Doherty, A. R., Smeaton, A. F., Lee, K. and Ellis, D. P.: Multimodal segmentation of lifelog data, *Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE, pp. 21–38 (2007).
- [7] Jégou, H., Douze, M. and Schmid, C.: Improving bag-of-features for large scale image search, *International Journal of Computer Vision*, Vol. 87, No. 3, pp. 316–336 (2010).
- [8] Lowe, D. G.: Object recognition from local scale-invariant features, *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Vol. 2, Ieee, pp. 1150–1157 (1999).
- [9] 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).
- [10] Shiga, Y., Dengel, A., Toyama, T., Kise, K. and Utsumi, Y.: Daily activity recognition combining gaze motion and visual features, *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, ACM, pp. 1103–1111 (2014).
- [11] 西田謙太郎, 吉高淳夫, 平嶋宗: 停留点の時間的な発生頻度に着目した注目点検出法の検討 (セッション 6: モニタリング+ 管理), 情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告, Vol. 2007, No. 11, pp. 121–126 (2007).
- [12] 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, Vol. 82, No. 10, pp. 1295–1303 (1999).
- [13] 中村匡秀, 下條彰, 井垣宏: 異なるライフログを集約するための標準データモデルの考察, 電子情報通信学会技術研究報告, Vol. 109, No. 272, pp. 35–40 (2009).
- [14] 保坂良資, 渡辺瞭: まばたき発生パターンを指標とした覚せい水準評価の一方法, 人間工学, Vol. 19, No. 3, pp. 161–167 (1983).