

# 深層学習を用いた可視光画像からの瞳孔検出と注視点推定への応用

北角 一哲<sup>1,a)</sup> 中澤 篤志<sup>1</sup> 西田 豊明<sup>1</sup>

**概要:** 瞳孔の検出は視線検出や個人認証で重要だが、赤外光下の画像を入力として用いなければならない、赤外光源・カメラが必要であるという問題があった。本論文はこれに対し、可視光画像から瞳孔を検出する手法を提案し、眼球の三次元姿勢追跡及び視線検出を行う手法を提案する。

瞳孔検出には畳み込みニューラルネットワークによる画像セグメンテーション手法を利用する。得られた瞳孔輪郭と可視光画像からの瞳輪郭を組み合わせ、三次元の眼球モデルを用いて粒子フィルタでパラメータ検出を行うことで、頑健な姿勢推定を可能にした。実験では瞳孔位置による注視点検出方法および角膜イメージング法による注視点検出方法を評価する実験を行い、手法の有用性を確認した。

**キーワード:** 瞳孔検出, 可視光画像, 畳み込みニューラルネットワーク

## Pupil Detection from Visible Light Corneal Image Using Deep Learning and Application to Gaze Estimation

KAZUNORI KITAZUMI<sup>1,a)</sup> ATSUSHI NAKAZAWA<sup>1</sup> TOYOAKI NISHIDA<sup>1</sup>

**Abstract:** Detection of pupils is important for gaze detection and personal authentication, but there is a problem that an infrared light image must be used as an input, and an infrared light source and camera is necessary. In this paper, we propose a method to detect the pupil from the visible light corneal image and propose a method to track the 3D attitude of the eye and detect gaze point.

For the pupil detection, we use the image segmentation method by the convolutional neural network. we can estimate 3D attitude robustly by combining the obtained pupil contour with the pupil contour from the visible light corneal image and performing parameter detection with a particle filter using a 3D eye model. In our experiments, we evaluate method of point of gaze detection and method of tracking 3D attitude of the eye, and we confirm the usefulness of our method.

**Keywords:** pupil detection, visible light image, convolutional neural network

### 1. 序論

近年、様々な視線計測装置・デバイスが存在する。こうした視線計測装置は、コンビニや自動販売機における商品の配列の改善や、ホームページのボタン・画像の配置といったユーザーインタフェースの改善などに応用されており、視線計測の重要性が高まっている。

視線検出には様々な手法があるが、現在よく使われているのは瞳孔位置による手法 (Pupil Center Corneal Reflection: PCCR) である。この手法では、眼を計測するアイカメラから瞳孔を検出し、その中心位置などとシーン画像位置をマッピングすることで視線を検出する。

従来、瞳孔を検出するには、赤外光下の画像を入力として用いなければならない、赤外光源・カメラが必要である。しかし、赤外光源・カメラを用いるとコストがかかることや、太陽光下では瞳孔検出が困難であるという問題点が挙げられる。本論文はこれに対し、可視光画像から瞳孔を検

<sup>1</sup> 京都大学  
Kyoto University  
<sup>a)</sup> kitazumi@ii.ist.i.kyoto-u.ac.jp

出する手法を提案する。

本研究では画像セグメンテーションによる瞳孔の検出を考える。近年、畳み込みニューラルネットワークをベースにした画像セグメンテーション手法に Fully Convolutional Networks(FCN) が提案されている [1]。FCN は Fully Connected 層を持たないネットワーク構造をしており、画像の形で出力することにより画像セグメンテーションが可能となる。さらに FCN の発展として、deconvolution 層を導入した構造 [2]、領域ベースの R-fcn[3]、encoder-decoder 間にパスを導入した Segnet[4]、U-Net[5] などが提案されている。本研究では U-Net を適用した瞳孔検出手法を提案する。可視光の角膜画像を入力とし、瞳孔領域を出力とするような学習を行うことで瞳孔の検出が可能であると考えられる。

本研究では深層学習により検出した瞳孔の情報を利用した、角膜輪郭追跡や視線検出手法への応用についても述べる。

## 2. 関連研究

従来の瞳孔検出に関する研究には Morimoto らの研究 [6] や Qiang らの研究 [7] などが挙げられる。Morimoto らは現在一般的な瞳孔検出手法となっている明瞳孔・暗瞳孔法を提案している。この手法では二つの異なる波長を持つ赤外光源を交互に照射することで、瞳孔が明るく映る画像と暗く映る画像が得られ、その二つの画像の差分から瞳孔を検出できる。しかし、この手法では、大型な赤外光源と赤外カメラが必要となるという問題点がある。Qiang らは明瞳孔・暗瞳孔法を利用して、ドライビング中の被験者の瞳孔を検出し、ドライバーの警戒レベル状態を推定する手法を提案している。

近年、Morita ら [8] によって可視光画像を利用した瞳孔検出手法が提案されている。この手法では角膜にシーン反射がある状態で、角膜と瞳孔の輝度差が小さい場合にも瞳孔輪郭が検出できることが確認されている。しかし、Morita らの研究における実験環境ではシーン反射による影響は小さい画像が用いられているが、実環境においてはシーン反射により瞳孔の大部分が隠されている場合も存在するため、本研究では、シーン反射の大きな角膜画像においても検出可能な手法の実現が目標となる。

次に、可視光の角膜反射画像を利用した視線計測手法として従来、眼球モデルを利用するもの [9]、回帰モデルを用いるもの [10] が挙げられる。Nakazawa らは三次元の眼球モデルを用いて粒子フィルタでのパラメータ検出を行い、頑健な眼球の三次元姿勢推定と注視点推定を可能にしている。しかし、この手法の問題点として角膜がまぶたで隠されてしまうと、角膜輪郭追跡が失敗してしまう場合がある。この問題を解決する手法を本研究では提案する。

また、Mori らは主成分回帰により、注視点と角膜画像

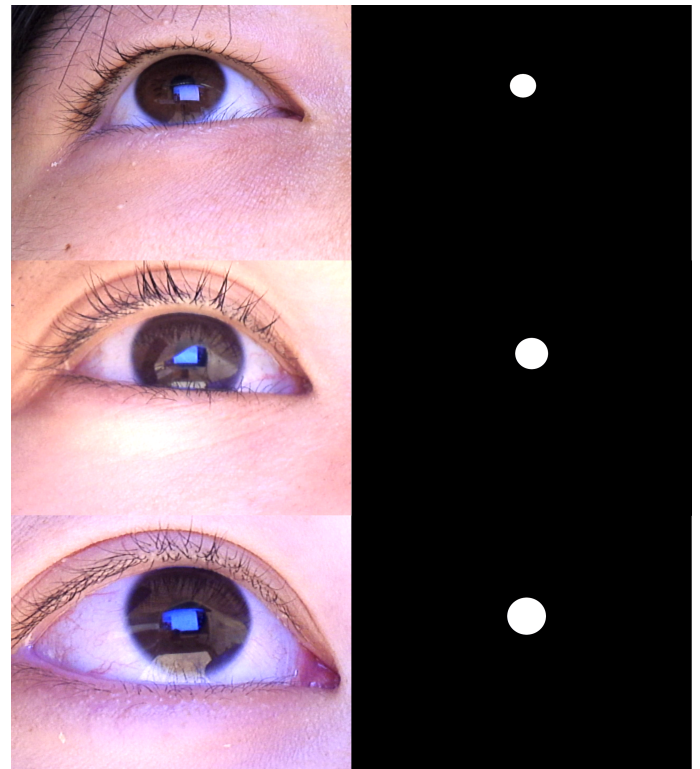


図 1 左：入力画像（可視光）、右：手動で検出した瞳孔領域

の関係を回帰モデルを用いて注視点を推定する手法を提案している。双曲面ハーフミラーを用いることで、カメラ画像に角膜とミラーに映りこんだシーンを同時に撮影し、視線推定を可能にしている。この技術により視線推定誤差が  $1^{\circ} \sim 2.5^{\circ}$  程度の結果が得られている。

## 3. 可視光画像からの瞳孔検出

### 3.1 U-Net を用いた瞳孔検出

提案手法では瞳孔を検出するための手法として U-Net[5] を用いる。U-Net とは Ronneberger らによって提案された、畳み込みニューラルネットワークによる画像セグメンテーション手法である。本研究では図 2 の構造をしたネットワークを用いる。前半の Convolution・Pooling 層により局所特徴の抽出を行い、後半の Convolution・upSampling 層により特徴を保持したまま画像を大きく復元する。これにより入力画像と同じサイズの画像が出力として得られる。

本研究では入力を可視光の角膜画像、出力を瞳孔領域のセグメンテーション画像となる。図 1 に示すような教師データを作成する。左が入力の可視光画像、右が手動で検出した瞳孔領域の正解画像である。こうして作成したデータセットを図 2 のネットワークにより学習を行う。

損失関数にはダイス係数を用いる。損失関数、ダイス係数は集合  $X$ 、集合  $Y$  に対して以下のように計算される。

$$Dice = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$
$$Loss = 1 - Dice$$

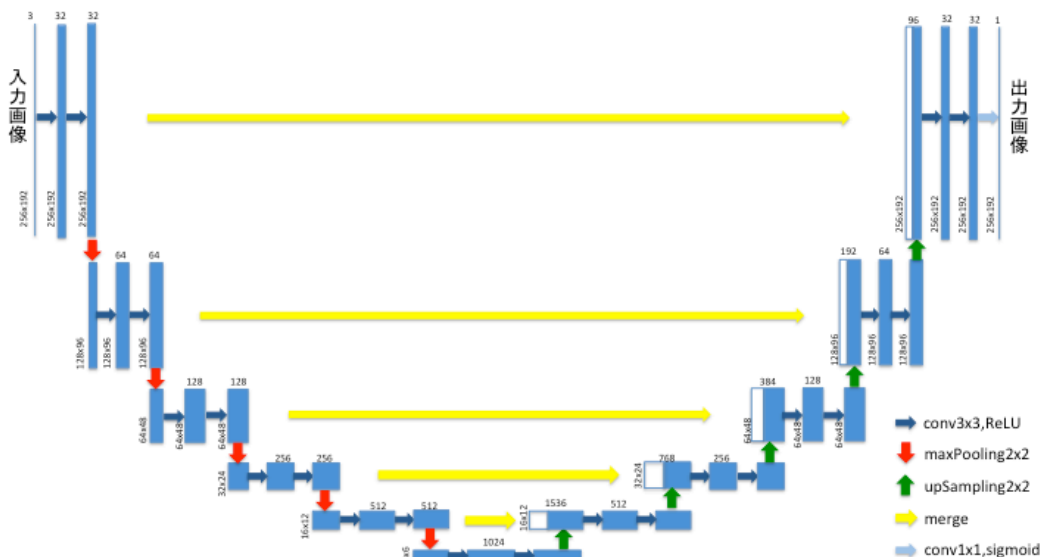


図 2 学習に用いるネットワーク構造 (U-Net)



図 3 二つ以上の領域が出力される例

ダイス係数は 0 ~ 1 の値を取り, 集合  $X, Y$  が一致していれば 1 となる. したがって損失関数を  $1 - Dice$  とすることで, 損失が 0 すなわちダイス係数が 1 に近づくように学習が進むため, ダイス係数を用いた学習が可能になる.

### 3.2 検出ノイズの除去

U-Net における瞳孔検出では図 3 のように二つ以上の領域がノイズとして出力されてしまうことがある. 現在観測されたものについては, 複数の領域の中に瞳孔領域も含まれているため, こうした場合には三次元眼球モデルと組み合わせることで正しい瞳孔領域を選択することを試みた. 各フレーム毎に眼球モデルのパラメータの一つとして, 画像中の眼球中心  $(c_x, c_y)$  を粒子フィルタを用いて推定する. ここであるフレームの瞳孔検出結果が二つ以上の領域  $S = (s_1, \dots, s_i)$  ( $i$  は領域数) を含んでいた時, 正しい瞳孔を選択するために, 各領域の中心と前フレームで推定された眼球中心との距離  $D = (d_1, \dots, d_i)$  を計算し, 最小となる領域を選択する. つまり, 選択される瞳孔領域  $S_{pupil}$  は以下で求められる.

$$k = \operatorname{argmin}(D)$$

$$S_{pupil} = s_k$$

## 4. 検出された瞳孔の応用

本章では, 深層学習により検出された瞳孔を利用した視線検出手法および角膜姿勢追跡手法について述べる.

### 4.1 瞳孔中心を用いた視線検出

本節では, 検出された瞳孔を利用した視線検出手法について述べる.

あらかじめ, 角膜映像から得られる瞳孔中心と見ているシーンの対応点を利用して幾何学的変換を作成する, すなわちキャリブレーションを行う. 具体的に本研究ではディスプレイに 12 個のマーカーを順に表示し, マーカーを注視する操作を行うことで対応するマーカー位置と瞳孔中心の組が 12 組得られる. 本研究の環境では, 一つのマーカーに対し角膜映像がそれぞれ 60 枚ずつ得られるが, 視線移動の遅延を考慮し, 中央の 20 枚の画像を利用する. また瞬目により瞳孔推定不可能なフレームが存在することも考慮し, 20 枚の画像から得られる瞳孔中心  $p = (p_1, p_2, \dots, p_{20})$  位置の中央値  $p_{med} = \operatorname{median}(p)$  を対応点として利用する. こうして得られたマーカー位置  $M = (M_1, M_2, \dots, M_{12})$  と瞳孔中心  $p_{med} = (p_{med1}, p_{med2}, \dots, p_{med12})$  の 12 組の対応点から二次多項式変換  $tform$  を計算する.

キャリブレーション後, 得られた二次多項式変換を利用して, 角膜映像中の瞳孔中心位置  $p$  をシーン画像上の点  $u_p$  にプロットする. これは以下の式で計算される.

$$u_p = tform(p)$$

このシーン画像上の点  $u_p$  が視点となる.

この手法の特徴として, キャリブレーションを行うこと

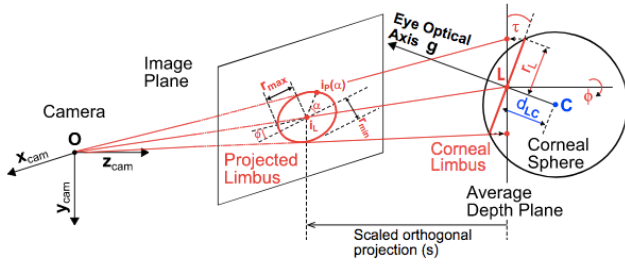


図4 三次元眼球モデル [9]

で、カメラを頭に固定した場合、体の動かしたとしても頑健な視線検出を行うことができる。

#### 4.2 角膜姿勢追跡への応用

本節では、検出された瞳孔を利用した角膜姿勢追跡手法について述べる。従来法には Nakazawa らによって提案された粒子フィルタによる眼球姿勢追跡手法がある [9]。本研究では、この手法と検出された瞳孔の中心を組み合わせることで、頑健な角膜輪郭追跡手法を提案する。

三次元眼球モデルの詳細を図4に示す。画像に投影される角膜は5つのパラメータ  $[c_x, c_y, \phi, \tau, s^*]$  によって表される。 $c_x, c_y$  は画像中における眼球中心座標、 $\phi, \tau$  はカメラに対する眼球角度、 $s^*$  はスケールを表している。このパラメータを粒子フィルタを用いて各フレーム毎に推定する。

追跡手法は以下に示すステップで行われる。

##### (1) 角膜輪郭選択および眼球モデル推定。

まず角膜輪郭を手動で選択し、それにより眼球モデルパラメータを計算する。パラメータは  $[c_x, c_y, \phi, \tau, s^*]$  の5つからなる。

##### (2) 粒子生成。

前フレームの眼球パラメータの近傍から粒子を  $K$  個ランダムに決定する。本研究では  $K = 50$  としている。

##### (3) 角膜輪郭点生成。

各粒子に対して、角膜輪郭点  $N$  点  $e = (e_1, \dots, e_N)$  を以下の数式により計算する。

$$[a_L, a_S] = [r_L s^*, r_L s^* \cos(\tau)]$$

$$e_i = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} a_L \cos(2\pi i/N) \\ a_S \sin(2\pi i/N) \end{bmatrix}$$

ここで  $a_L, a_S$  は角膜画像における角膜輪郭の長径・短径を示し、 $r_L$  は実際の角膜の半径であり  $r_L = 6.0\text{mm}$  としている。本研究では  $N = 100$  としている。

##### (4) 角膜輪郭の尤度計算。

各粒子に対する角膜輪郭の尤度を計算する。尤度は以下の数式により計算する。

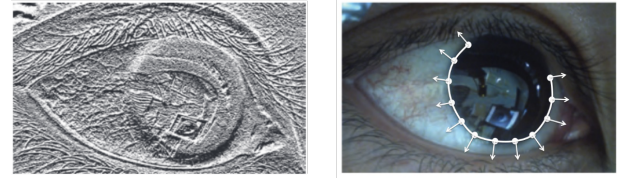


図5 左：勾配画像、右：角膜輪郭の法線

$$ang = |\arctan(\frac{e_i(y) - c_y}{e_i(x) - c_x})|$$

$$w_i = \exp(-(\frac{ang}{2})^2)$$

$$d = \sqrt{(c_x - p_x)^2 + (c_y - p_y)^2}$$

$$L = \left\{ \sum_{i=1}^N (w_i \hat{\mathbf{i}}(e_i) \cdot \mathbf{n}_i) \right\} \times \frac{1}{\cosh(d/100)}$$

ここで  $e_i(x), e_i(y)$  は角膜輪郭点  $e_i$  の  $x, y$  座標、 $\hat{\mathbf{i}}$  は角膜画像上の点  $e_i$  における勾配方向、 $\mathbf{n}_i$  は角膜輪郭点  $e_i$  における法線方向、 $p_x, p_y$  は瞳孔中心の  $x, y$  座標を示している。ここで評価値  $L$  が最大のものを結果として出力し、2.へ戻る。

従来法では尤度計算において図5に示すような角膜画像の勾配と法線方向のみを用いていた。それに対して本手法では検出された瞳孔中心を利用して、各粒子の尤度  $L$  を計算する。これにより  $d$  が小さい、すなわち瞳孔と眼球中心が近いほど評価値が大きくなるため、従来法では、尤度が局所解に落ちてしまい追跡が失敗していた問題を解決することが可能となる。

## 5. 実験

5.2章と5.3章では以下で述べる実験環境で取得したデータを利用する。

### 5.1 実験詳細

評価実験では7名を対象にデータを取得し、各被験者に対し瞳孔検出と視線推定を行った。図8に示したように被験者は図6のアイカメラを頭部に装着し、ディスプレイに表示されるマーカーを注視している際の角膜映像を撮影する。被験者の前方140cmにディスプレイを設置し、ディスプレイには図7に示す12個のマーカーをひとつずつ順に表示する。マーカーの表示間隔は2秒で、約60フレームである。この映像を被験者が注視するという作業を2回連続で行う。5.3章の視線推定実験では、一回目の映像を瞳孔位置とマーカー位置のキャリブレーションに用い、二回目の映像を視線推定と誤差評価に用いる。今回の実験では前方シーンカメラが設置されていないため、瞳孔位置とマーカー位置の対応を保持するため頭の位置を固定している。目の位置から見たとき、マーカー同士が約  $8.2^\circ$  の間隔に位置しており目と最下段のマーカーの高さがおおよそ水

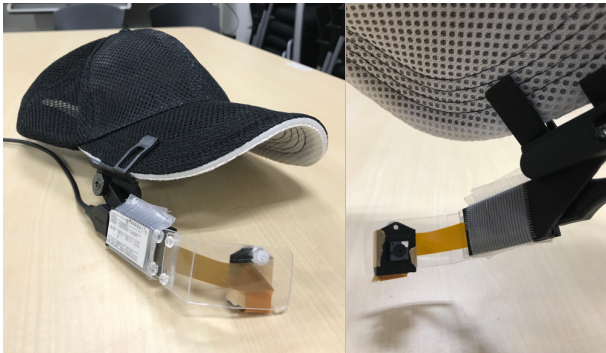


図 6 装着型アイカメラ

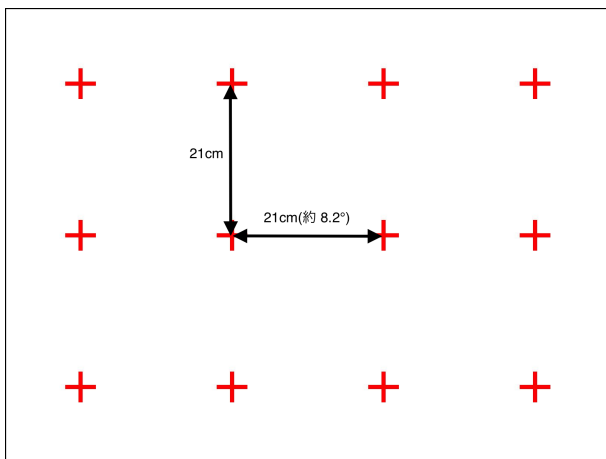


図 7 ディスプレイに表示するマーカー画像

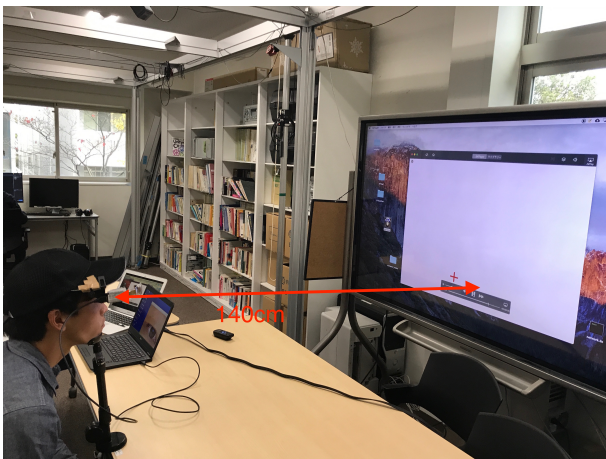


図 8 実験環境

平となっている。

## 5.2 瞳孔検出

本節では、U-Net を利用した可視光画像からの瞳孔検出の精度評価を行う。

### 5.2.1 実験詳細

で述べた実験環境で取得した角膜映像から図 1 のように手で瞳孔領域を検出し教師データを作成する。各被験者のデータから約 12 枚取り出し、合計 83 組のデータセットを

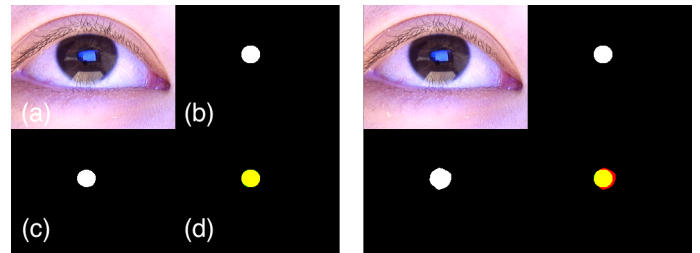


図 9 被験者 A についての瞳孔検出結果。左：caseA，右：caseB。各画像について、(a):入力画像、(b):正解画像、(c):推定画像、(d):評価画像

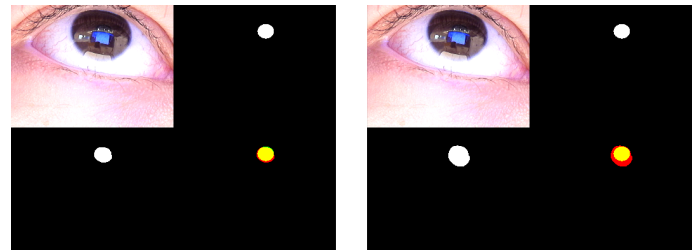


図 10 被験者 B についての瞳孔検出結果

作成した。本実験ではデータ拡張は行っていない。入力画像サイズは  $256 \times 192 \times 3$ ，出力画像サイズは  $256 \times 192 \times 1$  としている。活性化関数として各層の出力に ReLU，最後の出力層に sigmoid 関数を用いた。最適化関数に SGD，損失関数に Dice 係数を採用した。学習は 1epoch あたり約 8 秒の実行時間で 300epoch で学習器を作成する。

精度評価では 2 通りのパターンで学習を行った。

case A 被験者のデータを学習データセットの含めて学習を行う。このケースでは各被験者データから約 8 組をトレーニング用，3 組を評価用として用いる。

case B 被験者のデータを学習データセットに含めずに学習を行う。このケースでは検出対象の被験者以外の 6 人のデータを用いて学習を行い、対象の被験者のデータ約 10 枚で瞳孔検出を行い評価する。

実験では GPU に NVIDIA の GeForce GTX 970 を搭載した計算機を使用した。

### 5.2.2 精度評価

得られた学習器を用いてテストデータセットに対して瞳孔検出を行った結果を図 9，図 10 に示す。

各図の左側が caseA，右側が caseB の結果である。(a) が入力，(b) が正解，(c) が推定，(d) が評価画像となっている。評価画像は緑チャンネルに正解，赤チャンネルに推定結果を入れているため，正解と推定が重なっている部分は黄色で表示される。caseA では大部分が黄色で表示されているため，高い精度で検出できていることが確認できるが，caseB の結果については赤や緑の部分が大きくなっている。

図 9 に示されているように，角膜に見ているシーンが写り込んでいる場合やまぶた・睫毛による影がある場合に対

しても頑健に瞳孔検出が出来ていることが確認できる。

また, Precision, Recall, F 尺度により pixel 単位での精度評価を行った。Precision, Recall, F 尺度は以下の式で計算する。

$$Precision = \frac{tp}{tp + fp}$$

$$Recall = \frac{tp}{tp + fn}$$

$$F_{measure} = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$tp$  は検出されかつ正解であるピクセル数 (図 9 の黄色),  $fp$  は検出されたが瞳孔領域ではないピクセル数 (図 9 の赤色),  $fn$  は検出されなかった瞳孔領域のピクセル数 (図 9 の緑色) を表している。

評価結果を表 1 に示す。この結果より, caseA では 90% 以上の精度で, caseB においてもおよそ 80% 以上の精度で検出が出来ていることが確認できる。被験者 E の数値が大きく下がっているが, このデータは他の被験者のデータに比べ, 虹彩と瞳孔の輝度差が非常に小さく, 人による瞳孔の確認すら難しく検出が困難なものとなっていた。しかし学習データに加えることで 80% 以上の結果が出ていることから, データセットを十分に準備し汎化性能を持たせることができれば, 虹彩と瞳孔の輝度差が非常に小さい場合においても検出は可能であると言える。caseB の精度を caseA に近い精度を実現するために, どれだけの人々のデータが必要であるか, どういったデータ拡張が有効かを確かめる必要がある。

表 1 瞳孔検出精度評価。

被験者	Precision		Recall		F-measure	
	caseA	caseB	caseA	caseB	caseA	caseB
A	0.951	0.756	0.911	0.958	0.931	0.845
B	0.849	0.619	0.959	0.924	0.900	0.741
C	0.939	0.99	0.936	0.772	0.937	0.83
D	0.960	0.958	0.891	0.765	0.924	0.851
E	0.774	0.136	0.912	0.412	0.837	0.205
F	0.988	0.937	0.878	0.792	0.930	0.859
G	0.961	0.99	0.912	0.73	0.936	0.84

### 5.3 視線検出

本節では, 瞳孔を利用した視線検出の評価実験について述べる。実験詳細は 5.2.1 で述べた実験環境でデータを取得し, 各被験者に対して視線推定を行い, その誤差を評価した。被験者 G について, 実験設定の不備が生じていたため視線検出実験からは除外した。

被験者 A に対する視線推定結果を図 11 に示す。これは caseB における結果である。この結果より, 誤差  $1^\circ$  程度の範囲で視線推定ができていることが確認出来る。これは安定した瞳孔検出ができていけると言える。また各被験者につ

表 2 視線推定の平均角度誤差

被験者	A	B	C	D	E	F
caseA	0.80°	0.51°	0.87°	1.18°	1.40°	0.60°
caseB	0.91°	0.64°	0.7°	1.15°	2.49°	0.62°

表 3 Mori らによる視線推定の平均角度誤差 [10]

被験者	A	B	C	D
平均角度誤差	1.09°	1.40°	2.12°	2.61°

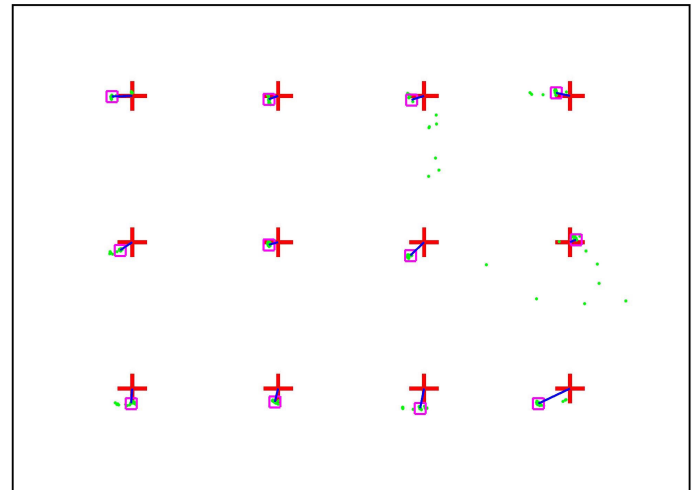


図 11 視線推定結果

表 4 角膜姿勢追跡における従来法と提案手法の accuracy の比較

	正解フレーム数	全フレーム数	accuracy
従来法	204 フレーム	2353 フレーム	204/2353 (8.67%)
提案手法	2031 フレーム	2353 フレーム	2031/2353(86.3%)

いて caseA と caseB で比較した結果を表 2 に示す。参考に Mori らの手法における視線推定誤差を表 3 に示す。被験者, 実験環境ともに異なっているため, 単純な比較にはならないが誤差として提案手法の方がほとんどの場合, 小さな値が出ていることが確認出来る。また, caseB ときの被験者 E については瞳孔検出の精度が低かったために視線推定が失敗してしまったと考えられる。

### 5.4 角膜姿勢追跡

大学キャンパス内を移動している時の角膜映像を取得し, その映像に対し瞳孔検出および角膜輪郭追跡を行う。得られた映像は約 30fps, フレーム数 2354, 画像サイズは  $1280 \times 960 \text{ pixel}$  となっている。瞳孔検出に関しては, 得られた映像の内 13 フレームを 5.2 章で利用した学習データセットに追加することで実現した。

従来の角膜イメージング法による角膜姿勢追跡と, 瞳孔を利用した角膜姿勢追跡の比較を図 12 に示す。(a) が従来法の結果, (b) が提案手法の結果を 10 フレームごとに表示したものである。また, 姿勢追跡のフレーム単位での精度を従来法と提案手法で比較したものを表 4 に示す。正解は

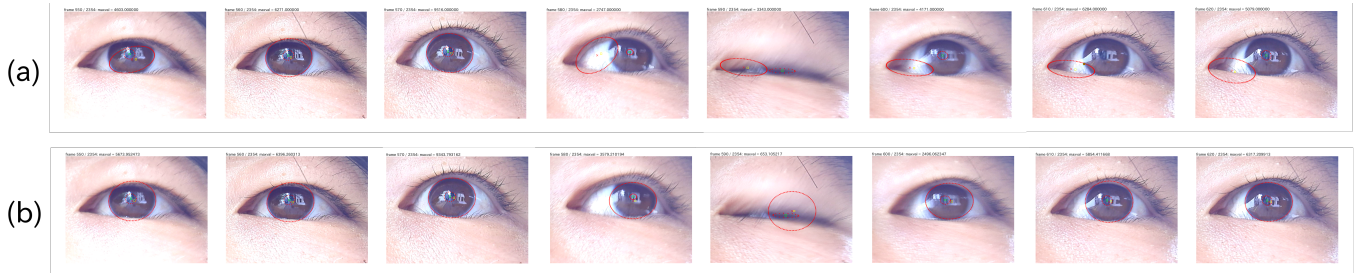


図 12 角膜追跡結果 . (a):従来法 , (b):提案手法

手動で判定した . 提案手法の精度が従来法を大きく上回ることが確認できた . 従来法では一度推定を失敗すると , その後正しい位置に修正することができなかつたのに対し , 提案手法では安定した追跡が行えていた . 図 12 を見ると従来法で角膜輪郭推定結果が目尻に寄ってしまっている箇所が , 提案手法では正しく追跡が出来ていることが確認できる . 従来法の結果について , 目尻に沿った箇所を誤って角膜輪郭と検出しており , 評価値も比較的大きな値が出ているため , 局所解を抜け出すことが困難であると考えられる . それに対して提案手法では , 瞳孔から離れた箇所は尤度が大きく下がるため , 局所解に入る可能性が小さくなりこうした結果が出ていると考えられる . 以上より瞳孔中心を利用することで従来法より頑健な追跡が可能であると言える .

## 6. 結論

本研究では , 深層学習を利用した可視光画像からの瞳孔検出手法を提案し , 獲得した瞳孔を利用した眼球の三次元姿勢推定手法の改善 , 視線推定手法を提案を行った . 実験の結果から , 角膜にシーンが映り込んでいる状態であっても , 高い精度で瞳孔領域を検出することが可能であることが示された . また , 角膜輪郭の追跡や視線推定手法についても従来手法以上の結果が得られている . 視線推定手法については ,  $1^\circ$  程度の誤差で推定可能であることが確認できた . 角膜画像と同時に一人称視点画像が獲得できれば , 人の動きが伴った場合においても視線推定が可能であると考えられる .

本研究の問題点として , 瞳孔検出モデルの汎化性能が不十分であることが挙げられる . 汎化性能を向上させるために必要な学習データ数や被験者人数を確認するとともに , 最適なデータ拡張手法を見つける必要がある . また , 本研究における視線推定方法ではキャリブレーションが必要であるが , キャリブレーションなしで GRP を推定する角膜イメージング法と組み合わせることで , キャリブレーション不要な視線推定手法を実現することが今後の課題となる .

謝辞 本研究は科研費 17H01779, 26249029, 15H02738, および JST, CREST, JPMJCR17A5 の支援を受けたものである .

## 参考文献

- [1] Long, J., Shelhamer, E. and Darrell, T.: Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015).
- [2] Noh, H., Hong, S. and Han, B.: Learning deconvolution network for semantic segmentation, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528 (2015).
- [3] Dai, J., Li, Y., He, K. and Sun, J.: R-fcn: Object detection via region-based fully convolutional networks, *Advances in neural information processing systems*, pp. 379–387 (2016).
- [4] Badrinarayanan, V., Kendall, A. and Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *arXiv preprint arXiv:1511.00561* (2015).
- [5] Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 234–241 (2015).
- [6] Morimoto, C. H., Koons, D., Amir, A. and Flickner, M.: Pupil detection and tracking using multiple light sources, *Image and vision computing*, Vol. 18, No. 4, pp. 331–335 (2000).
- [7] Ji, Q. and Yang, X.: Real-time eye, gaze, and face pose tracking for monitoring driver vigilance, *Real-Time Imaging*, Vol. 8, No. 5, pp. 357–377 (2002).
- [8] Morita, Y., Takano, H. and Nakamura, K.: Pupil diameter measurement in visible-light environment using separability filter, *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, IEEE, pp. 000934–000939 (2016).
- [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] Mori, H., Sumiya, E., Mashita, T., Kiyokawa, K. and Takemura, H.: A wide-view parallax-free eye-mark recorder with a hyperboloidal half-silvered mirror and appearance-based gaze estimation, *IEEE transactions on visualization and computer graphics*, Vol. 17, No. 7, pp. 900–912 (2011).