

キャリブレーションを必要としない視線領域推定アルゴリズムの構築と教育現場での活用

代表研究者	柴 里 弘 毅	熊本高等専門学校	電子情報システム工学系	教授
共同研究者	永 田 正 伸	熊本高等専門学校	拠点化プロジェクト系	教授
共同研究者	加 藤 達 也	熊本高等専門学校	電子情報システム工学系	助教

1 はじめに

今日、視線の推定に関しては数多くの手法[1], [2]が考案されており、その成果は様々な分野に利用されている。一例として、発話が困難な脳血管障害者や筋萎縮性側索硬化症患者の意思表示を目的とした視線入力装置がある[3]。視線で画面上のアイコンを選択したり、文字を入力したりすることで文章を書いて読み上げるなどの機能を備えており、意思や気持ちの伝達が可能になるため、生活の質が改善されることが期待されている。

特別支援学校は、学習上または生活上の困難を克服し自立を図ることを目的とした学校であり、病院から通学する重度の重複障害のある児童生徒を対象とする課程も存在する[4]。病気や障害の状態および教育的ニーズに応じた指導が行われることで、一人一人の能力を伸ばし、主体的・自立的に生活する力を育てている。著者らはこれまでも特別支援学校と連携して、学習教材や支援機器の開発を行ってきた[5], [6]。知的な遅れがあり身体を思うように動かすことが難しい場合、児童生徒の持つ力に応じて表情や目の動きによる意思表示が行われることもある。保護者や教諭らの、対面に限らず視線による意思表示を可能にすることで自発的な活動を広げたいという想いに触れた。

一般に、非接触型の視線入力装置では、視線推定の精度を高めるため、使用開始前に複数回指定された点を注視する操作（キャリブレーション）を必要としている。しかし、身体機能の障害だけでなく知的障害もある児童生徒には、キャリブレーションの操作指示に従い視線を頻繁に移動させることは大変難しく、現行の視線入力装置を使いこなして意思表示を行うことは極めて困難である。また、EDINBURGH Univ. が作成した Eye Gaze Software Curve[7]では、目の動きを利用したソフトウェアを操作性や対象レベルに合わせて7段階に分類している。この Eye Gaze Software Curve においてマウスオーバーやシングル左クリックなどは基礎的な機能と位置づけられているが、重度重複障害者や介助者にとって、視線を利用して手軽に意思の伝達が可能になることは実用面で意味がある。

本研究では、重度重複障害者の利用を想定して、視線の分解能は粗くとも特別な視線入力装置を必要とせずに、キャリブレーションの負担もなく利用可能なリアルタイム視線領域推定システムの開発を行った。はじめに、画面領域の分割、畳み込みニューラルネットワーク（Convolutional Neural Network; CNN）モデルの選択に関する予備実験を行い、提案システムに最適な組み合わせを選択した。次に、学習済みモデルを用いて、カメラ映像からリアルタイムに視線領域推定を行うシステムを構築した後、頭部の距離や位置、および姿勢の条件を変化させて実験を行った。提案システムにおいて、基本姿勢における実用性を示し、頭部姿勢の変化が視線領域推定に及ぼす影響について考察する。

2 視線領域推定手法

2-1 学習データセットの作成

システムの利用状況として、Fig. 1 に示されるように横になっている状態の対象者に対し、教諭や保護者がタブレット端末の画面を見せる、もしくは対象者がアーム等で固定されたタブレット端末の画面のアイコンを注視して意思表示をするといった場面を想定する。この場面では、「どの座標を注視しているのか」といった高精度な視線推定ではなく、「どの領域を見ているのか」といった視線領域推定ができれば意思表示が行える。これを特殊なハードウェアを用いずに実現することができれば、利便性は著しく向上する。

学習データセットには、MPIIGaze[8]を用いた。このデータセットは、ノートPCを注視している15人の画像とそのパソコン上の注視座標点など37667組で構成されている。本研究では、このデータセットを以下の手順で再構成する。



Fig. 1 提案システムの利用概観

- 1) MPIIGaze に含まれる注視座標点から注視領域のラベリングを行う.
- 2) Haar-Like 特徴量に基づき, 元画像から顔領域のみを抽出する.
- 3) 2) において Haar-Like 特徴量による顔領域検出ができない画像をスクリーニングし, 訓練用, 検証用, テスト用データに分割する.

今回は, 訓練用データを 18000 枚, 検証用データを 2860 枚, テスト用データを 2008 枚使用した. データセットを検証用とテスト用に分割するのは, 過学習を防ぐためである. 学習係数などのハイパーパラメータが適切に設定されていない場合, 訓練用データによる認識精度と検証用データによる認識精度に乖離が生じることからパラメータを再調整する. 4 分割した視線領域ラベリングの例を Fig. 2 に示す.

area 1	area 2
area 3	area 4

Fig. 2 視線領域のラベリング

2-2 CNN モデルの特徴

CNN には, 畳み込み層やプーリング層などの層が存在し, その数や組み合わせ方により, 多くのモデルが存在する. 本研究では大きく 3 種類の CNN モデルを用いて, 視線領域推定の学習を行った.

1 つ目のモデルには, 2014 年に開催された画像認識の競技会 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) で好成績を収めた VGG [9] を使用した. 中間層の違いにより VGG16 や VGG19 などがある.

2 つ目は, GoogLeNet (Inception) である [10]. 幾つかの version が存在するが, 本研究では Inception-v3 を使用した. 複数のフィルタを用いることで, 細部の表現だけでなくより抽象的な表現の獲得も一つのモジュール内で行うことができる.

3 つ目のモデルには, チャンネル方向と空間方向の畳み込みを分離した Xception モデル [11] を使用した. 畳み込みを分離することにより計算量を削減でき, 層をより深くすることができる. Batch Normalization やスキップ機構などの技術も組み込まれている.

3 予備実験

3-1 画面分割に関する予備実験

はじめに、画面を 2, 4, 6, 9 分割とした際の視線領域推定に関する予備実験を行った。領域推定には VGG16 モデルを使用した (Fig. 3)。まず、データセットから切り出された顔領域を入力とし、分割された画面領域番号を出力とする学習を行った。ここで、学習係数を 0.001, 学習回数を 100 epoch とした結果を Tab. 1 に示す。分割数が多くなるほど訓練用データとテスト用データの精度に差が広がり過学習の傾向が強くなることが分かる。2 分割でシステムを構成すると信頼性は高いが選択肢は少なくなる。一方、9 分割では選択肢は増えるものの、信頼性が低くなるという問題が残る。

そこで、今回は想定している利用者の知的・身体的特性、および、選択肢の数と信頼性を踏まえ、テスト用データで 76.54 % の認識精度を示しバランスのとれた画面領域 4 分割について実験を進めることにした。

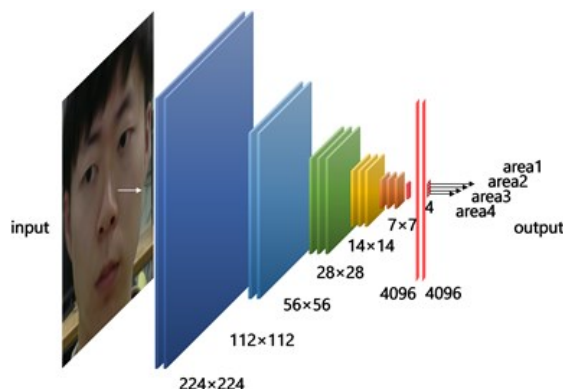


Fig. 3 VGG16 の構造

Tab. 1 視線領域の推定精度 (VGG16)

精度 (%) \ 画面領域数	2	4	6	9
訓練用データ	98.07	93.54	89.78	88.03
テスト用データ	91.38	76.54	67.48	55.63

3-2 CNN モデルの選択

画面領域推定に使用する CNN モデルの選択に関する実験を行った。3-1 節の結果を踏まえ画面分割は 4 分割とした。また、モデルには VGG (16, 19), Inception v-3, Xception の 3 つを用いた。3-1 節では、本研究に適した画面分割数を VGG16 モデルを用いて仮探索した。本節の実験では画面分割数を 4 に固定したことから、改めてハイパーパラメータのチューニングを行った。Tab. 2 に各 CNN モデルの認識精度を示す。

まず、VGG 同士を比較すると、VGG16 に比べ VGG19 の精度が少し劣ることが分かる。これは、層の深さが精度に直結するとは限らないことを示唆している。また、Inception v-3 と Xception については VGG より数 % 精度が向上しており、僅かな差ではあるが Xception モデルの認識精度が最も優れているという結果となった。

次に、Xception による画面 4 分割推定結果について考察を行う。Fig. 4 は、画面左上領域が正解ラベルのときの推定結果である。精度は 85.71 % で、対角領域と誤認識されるケースは 0.56 % と小さく、全く別領域と誤認識されることは少ないと判断できる。また、上下領域での誤認識 (10.90 %) は、左右領域での誤認識 (2.82 %) と比べて約 4 倍となった。また、他の 3 つの領域についても左右方向より上下方向での認識精度が低くなるという同様の傾向が見られた。これは、頭部や目が移動した際の特徴量の変化が、左右方向に現れやすく、上下方向では抽出が難しいことを表しているものと考えられる。

Tab. 2 各 CNN モデルの推定精度

model	VGG		Inception-v3	Xception
	16	19		
accuracy(%)	80.6	79.6	82.3	82.9

area 1 85.71%	area 2 2.82%
area 3 10.90%	area 4 0.56%

Fig. 4 左上領域注視時の視線領域推定精度

3-3 誤認識に関する考察

画面領域左右2分割において誤認識された画像の一部をFig. 5に示す. 分割数の少ない2分割においても, 人が直感的に判断することが難しい画像が含まれていることが分かる. このように領域推定を誤った画像の多くは, 画面中央の境界付近を注視しているものであった. テスト用データにはこのような画像も含まれているため, 精度が90%程度に留まったものと推測される. 実用的には, 境界付近に不感帯を設けるなどの改善策を講ずることに対応できると考えられる.

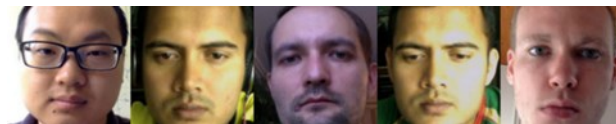


Fig. 5 誤認識された顔画像の例[8]

3-4 予備実験のまとめ

VGG16 において, 訓練用データでは9分割において88.03%の精度が得られるものの, テスト用データでは55.63%に留まる結果となった. 推論結果にフィルタ処理を加えることを考慮しても, テスト用データにおいて76.54%の精度が得られた4領域までが実用的な範疇に収まると考えられる. 3つのCNNモデルを比較した結果については, 中間層が多い方が必ずしも良い結果が得られるとは限らず, 僅かな差ではあるがXceptionモデルの認識精度が最も優れているという結果が得られた. 画面4分割においては, 左右方向に比べ上下方向に誤差が生じやすいことも明らかとなった.

4 基本姿勢に関する評価実験

4-1 リアルタイム視線領域推定手法

予備実験を踏まえ, 認識精度に優れていたXceptionモデルを用いてリアルタイム視線領域推定システムを構築した. 次の手順でリアルタイム視線領域推定を行う. なお, 今回の実験では, デスクトップPCで視線領域推定システムを構築した. Logicool社のwebカメラc270(画素数1280×720)では, フレーム取得から推定結果の表示まで25fps以上で動作した.

1. カメラから画像フレームを取得する.
2. Haar-Like 特徴分類器を用いて顔領域を切りだし, 画像として保存する.
3. 保存された顔画像をCNNの入力とする.
4. 画面注視領域の推定結果を比較し, 最も確率の高いものを推定領域として画像フレームに赤枠で表示する.
5. 上記1~4の処理を逐次繰り返す.

4-2 実験手法

提案した視線領域推定システムについて次のような評価実験を行った. Fig. 6に示されるようにカメラと頭部の位置を一定に保つため顎台で頭部を固定し, ディスプレイ上に配置した36点(1領域あたり9点)を注視した(Fig. 7). ここで, No. 1~9を領域1, No. 10~18を領域2, No. 19~27を領域3, No. 28~36

を領域4と呼ぶこととする。ディスプレイ上に表示されている点をNo. 順に注視し、1点につき150フレームのデータを取得した。なお、被験者には瞬目をしないように指示している。瞬目が発生した場合は、目を休める時間を数秒設け、その点から再計測を行った。

次に、意思表示に使用される場面を想定して、カメラと顔までの距離や、顔の位置、姿勢を変化させ評価を行った。今回の実験では、4つの領域のうちCNNの出力確率が50%を上回った領域をシステムが推定した視線領域として扱うものとする。4節に示す基本姿勢に関する実験、および、基本姿勢において距離、位置を変更した実験では、複数名の20代～50代の健常男性被験者で実験を行った。また、5節の姿勢を変化させた実験については、20代被験者1名の結果を示し、定性的な関係を考察する。

データの取得にあたっては、本研究の目的やデータの取り扱い等について被験者に説明し、適切な倫理的配慮を行ったうえで実施した。また、熊本高等専門学校「ヒトを対象とする研究に関する倫理審査委員会」による承認を得た（承認番号2020-2）。

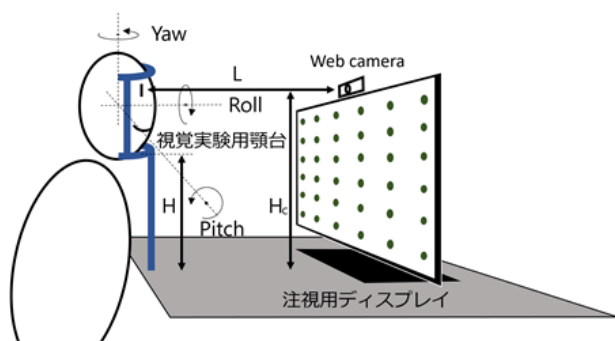


Fig. 6 実験環境

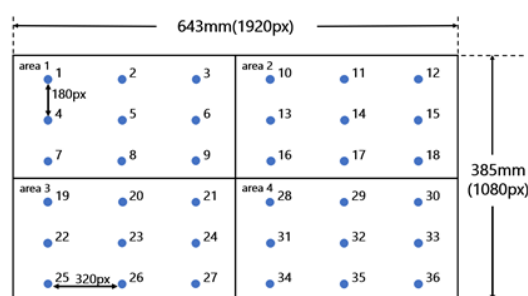


Fig. 7 注視点の配置パターン

4-3 基本姿勢における視線領域推定

Fig. 1に示される使用場面を想定し、視覚実験用顎台の高さを調整しカメラと正対させ、 $L = 700 \text{ mm}$ とした状態を基本姿勢とした。なお、 $H_c = 450 \text{ mm}$ である。この基本姿勢における被験者1名 ($H = 350 \text{ mm}$)の視線領域推定の結果をFig. 8に示す。横軸は注視点のNo.を表しており、縦軸は各注視点の推定精度を示している。

Fig. 7に示されるように領域1はNo. 1～9で構成されており、左上のNo. 1, 2, 4, 5では高い推定精度であることが確認される。同様に、画面四端付近の注視点においても高い精度で領域推定が行なわれている。また、注視領域が1から2に切り変わるNo. 9, No. 10において推定領域が1から2へと正しく切り替わっていることも分かる。同様の結果が領域2から3に切り替わるNo. 18, No. 19, 領域3から4に切り替わるNo. 27, No. 28でも観察され、4つの視線領域推定が適切に行われていることが分かる。しかし、領域1と領域3の境界や領域2と領域4の境界など上下の領域境界付近の点であるNo. 7とNo. 19, 20およびNo. 16, 17, 18においては精度が90%程度になる様子が確認された。一方、左右方向の境界となるNo. 24とNo. 31, 34に関しては推定精度が僅かに低下した。以上は、上下方向よりも左右方向の分解能に優れているというFig. 4の結果と一致する。

次に、同様の実験を5名の被験者に対して行い、基本姿勢における平均推定精度を求めた結果をFig. 9に示す。

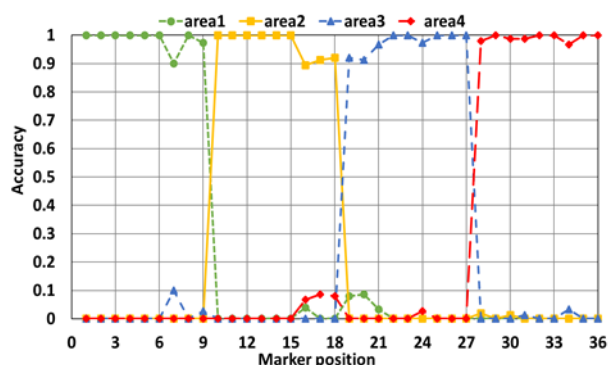


Fig. 8 基本姿勢における推定精度

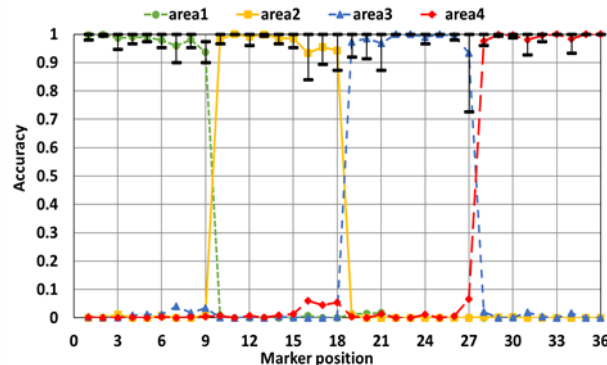


Fig. 9 基本姿勢における平均推定精度 (N=6)

これより Fig. 8 と同様の傾向が見られることが分かる。被験者 6 人の平均推定精度は 98.2 % で、個人毎では最高 99.2 %、最低 97.0 % であった。標本標準偏差の範囲は 0.00 % ~ 11.3 %、中央値は 1.35 %、平均値は 2.01 % であることから、ばらつきは一定の範囲内に収まっており、再現性があることが確認される。

次に、誤認識が最も起こりやすい点の 1 つと考えられる中央部の No. 16 について、被験者 1 名の 150 フレームの詳細データを Fig. 10 (a) に示す。正答である領域 2 との推定結果が概ね支配的で、対角方向の領域 3 と誤認識されたフレームは検出されなかった。しかし、80 フレーム付近までは領域 1、それ以降は領域 4 と誤認識される瞬間が散見され、誤りに偏りが見られる。瞬間的に誤認識される現象は他の被験者でも確認されたが、その大部分は領域 4 であった。Fig. 10 (a) において観察される現象はアルゴリズムに由来するものとも考えられるが、無意識に近接点の方にちらちらと視線の揺らぎが生じていた可能性も考えられる。今後、実験時の目の動きについても観察を行うなどして、注視に関する特性を調査する必要がある。

さらに、このデータに 10 フレームの移動平均処理を施したものを Fig. 10 (b) に示す。処理前は 89.3 % だった推定結果が処理後は 100 % になることが分かる。なお、他の 5 名の被験者についても同様に、10 フレーム移動平均で 100 % となることが確認された。このことから、実際にアプリケーションに使用する際には、移動平均処理を施すことで瞬間的な誤認識を抑えることができ、精度が安定すると考えられる。

なお、以降の実験データについては特に断りがない限り、移動平均を取らずに生データの実験結果を示す。

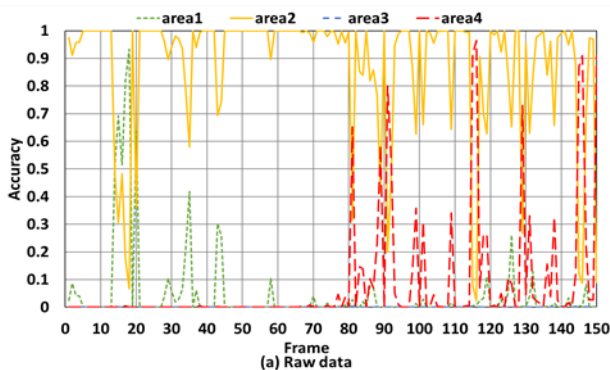
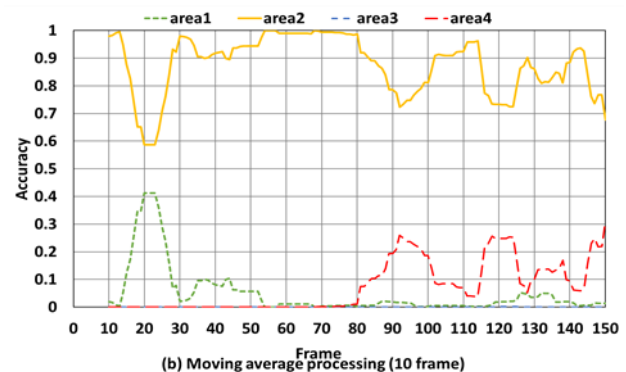


Fig. 10(a) 推定精度の時間変化(Raw)



(b) 推定精度の時間変化(移動平均)

4-4 カメラと顔の距離が推定精度に与える影響

基本姿勢を維持しつつ、カメラから顔までの距離 L を 1200 mm, 1700 mm に変更した実験を行った。まず、被験者 1 名の実験結果を Fig. 11 に示す。距離が 1200 mm の場合は精度の低下はほとんど見られないが、1700 mm においては上下左右の領域境界付近の点において、精度が大きく低下することが確認される。距離が離れることで抽出される顔領域の画素数が少なくなることから、分解能の低下に起因するものと考えられる。今回は一般性を考慮して 1280×720 の解像度でシステムを構築したが、近年はタブレット内蔵カメラの高解像度化も進んでいるため、カメラの解像度を高くすることができれば推定精度の低下を抑制できるものと推察される。

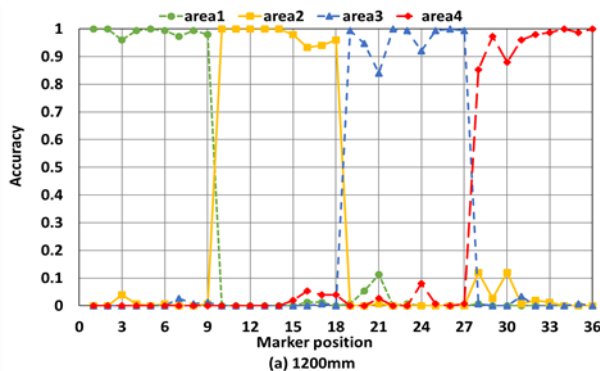
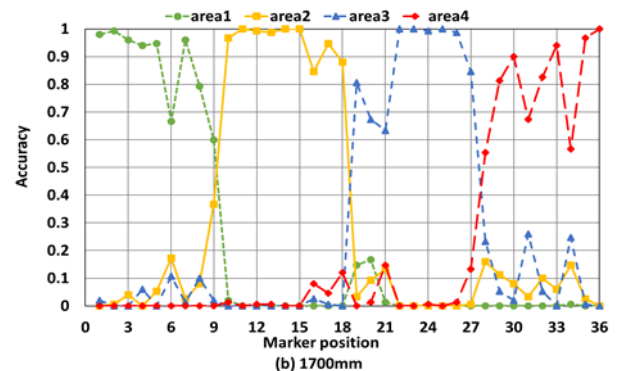


Fig. 11(a) 距離を変更した際の推定精度(L=1200mm)



(b) 距離を変更した際の推定精度(L=1700mm)

次に、距離 1200 mm における被験者 4 名の結果を Fig. 12 に示す。平均推定精度は 97.4 %，個人毎の平均推定精度は最高 98.4 %，最低 95.9 %であった。また、標本標準偏差の範囲は 0.333 %～10.4 %，中央値は 1.04 %，平均値は 2.46 %となった。領域境界付近ではばらつきがやや大きくなる傾向が見られるものの、距離 700 mm と同程度の平均推定精度が維持されており、再現性があることが示された。

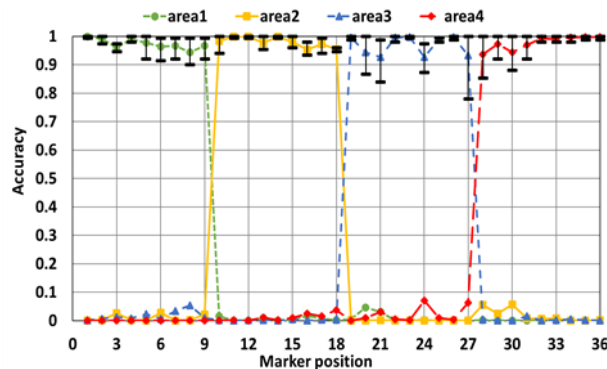


Fig. 12 距離を変更した際の平均推定精度 (N=4)

4-5 顔の位置が推定精度に与える影響

カメラと顔の位置が変化した場合を想定して、顎台の高さを 50 mm，100 mm，150 mm 下げて実験を行った。まず、被験者 1 名における 100 mm，150 mm の結果を Fig. 13 に示す。100 mm 低くした場合は、上下の領域境界付近の点で精度が 80 %程度まで低下するものの平均推定精度は 98.1 %となり、領域推定に支障をきたさないことが確認された。また、50 mm でも同様の傾向が見られた。しかし、150 mm 低くした場合は、領域 1，2 の下部の点 (No. 7, 8, 9, 16, 17, 18) については領域 3，4 の点と誤認識され、約 10 %～30 %の精度まで低下した。

次に、顎台を 100 mm 下げた実験について、被験者 4 名の結果を Fig. 14 に示す。個人毎の平均推定精度は 97.4 %，最高は 98.4 %，最低は 96.6 %であった。標本標準偏差の範囲は 0.00 %～8.60 %，中央値は 0.835 %，平均値は 2.33 %であった。Fig. 13(a) と同様に領域 1，2 の下部の点を領域 3，4 と誤認識する傾向が見られる。以上より、相対的に顔の位置を低くすると、上下の領域境界付近で推定精度の低下が起こり、下部領域と誤認識される傾向が顕著になることが示された。学習に用いた MPIIGaze データセットの顔画像は、ノート PC の内蔵カメラで取得されたものである。顔の位置がカメラより高い画角の画像であるため、実験のシチュエーションとは大きく異なる。このことが推定精度の低下に影響したものと考えられる。

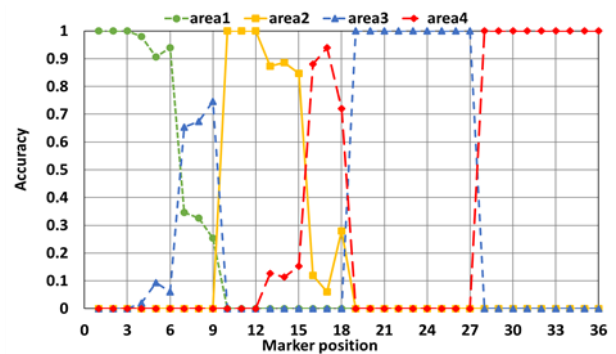
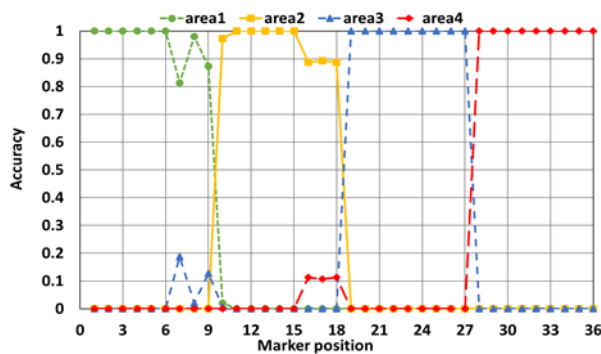


Fig. 13 (a) (b) 顔位置を下に変更した際の推定精度
(左(a) 100mm，右(b) 150mm)

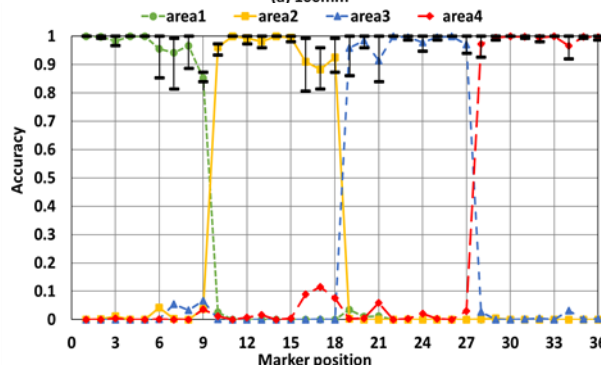


Fig. 14 顔位置を下に変更した際の平均推定精度
(N=4)

4-6 基本姿勢に関するまとめ

以上の結果より、基本姿勢 (L=700 mm) においては平均推定精度が 98.2 % となり、再現性があることが示された。10 フレームの移動平均処理を行うことで、推定精度が 100 % に向上することも確認された。また、基本姿勢を保ちつつ L=1200 mm とした場合において、基本姿勢と同程度の平均推定精度が維持されることが示された。顔の位置を 100 mm 下方にずらした場合は、基本姿勢と同程度の平均推定精度が維持される一方、領域 1、2 下方の領域境界付近で下部領域と誤認識される傾向が顕著になることが明らかとなった。

5 姿勢変化に関する評価実験

5-1 ヨー角が推定精度に与える影響

姿勢の変化については、被験者 1 名の実験結果をもとに定性的な関係について考察する。頭部のヨー角を右に 10 度、20 度とした場合の結果を Fig. 15 に示す。今回の被験者では、ヨー角 10 度における No. 17 において 20 % 程度まで精度が低下するものの、ほとんどの点について基本姿勢と同等の精度が得られることが確認された。顔はやや右側に向いているため、領域 1、3 を注視する場合は頭部と瞳の向きが異なる不自然な姿勢になる。それに関わらず、推定精度が維持されていることが分かる。

また、ヨー角 20 度の場合は、領域 2、4 において推定精度の低下が顕著に現れた。逆に、顔を左に 20 度向けると領域 1、3 の誤認識が増えるという対称的な傾向も確認された。ヨー角が大きくなると画面を注視する際の瞳の特徴量が基本姿勢と逆向きになると考えられ、このことが推定精度の低下を招くものと推察される。

以上より、学習に用いられたデータセットから視線領域推定するための顔の特徴量が適切に学習されており、今回の被験者では 10 度程度のヨー角のずれまでは許容されることが確認された。しかし、ヨー角が大きくなり顔と瞳の向きが異なるという不自然な注視をしている場合には、正しい領域推定は困難となることも明らかとなった。

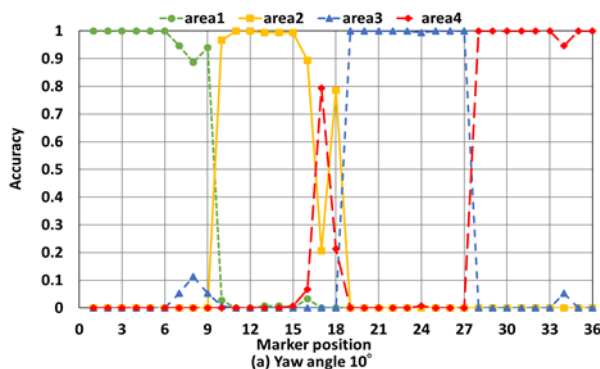
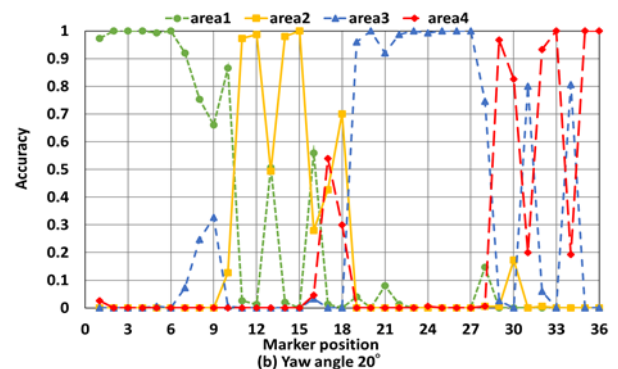


Fig. 15(a) ヨー角を変更した際の推定精度(10 度)



(b) ヨー角を変更した際の推定精度(20 度)

5-2 ピッチ角が推定精度に与える影響

次に、頭部のピッチ角を上 15 度、30 度とした場合の結果を Fig. 16 に示す。この被験者の場合、ピッチ角が 15 度の場合は全ての点において、94.0 % 以上の精度であり、基本姿勢の場合と同程度の精度であるこ

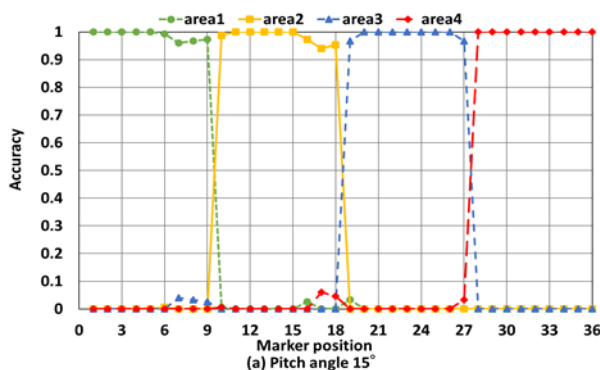
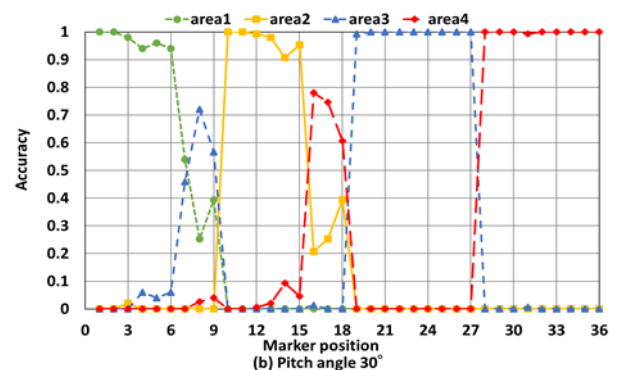


Fig. 16(a) ピッチ角を変更した際の推定精度(15 度)



(b) ピッチ角を変更した際の推定精度(30 度)

とが確認された。ピッチ角が 15 度においても精度が低下しなかった理由として、学習に用いたデータセットの撮影環境とピッチ角 15 度の実験環境が比較的類似しているためと推察される。しかし、ピッチ角 30 度になると上領域の境界部分の点 (No. 7, 8, 9 や No. 16, 17, 18) を下領域と誤認識する傾向が表れた。ピッチ角が大きくなり過ぎると学習に用いたデータセット画像の環境との差異が大きくなり誤認識が増加するものと考えられる。

5-3 ロール角が推定精度に与える影響

次に、ロール角を右に 10 度、15 度と変化させて実験を行った。頭部を 10 度傾けた場合は、36 点の平均精度が 98.5 %であった。Fig. 17 に 15 度傾けた実験の結果を示す。36 点の平均精度は 97.7 %であり、基本姿勢と同程度であることが確認される。ただし、領域 1, 3 の領域境界付近 (No. 3, 6, 9 や No. 21, 24, 27) の精度は下がり、右領域 (領域 2, 4) との誤認識が増加する傾向がうかがえる。また、左側に傾けた場合、検出精度は右側と同程度となり、誤認識については対称の傾向が見られることが確認された。

本システムでは Haar-Like 特徴量に基づき顔検出を行っており、ロール角約 20 度付近からは顔検出精度が著しく低下する。今回の被験者では、顔領域が問題なく抽出可能なロール角 15 度までは、精度的にもシステム的にも許容範囲内にあるといえる。

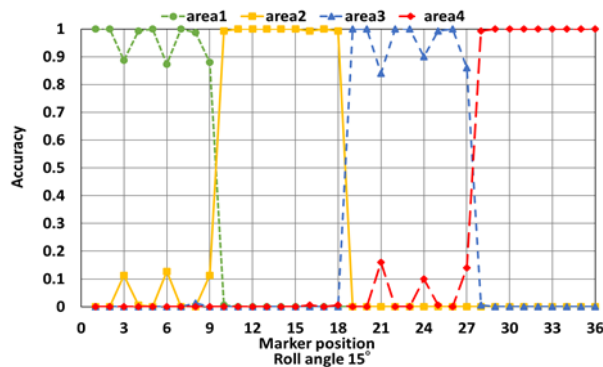


Fig. 17 ロール角 15 度における推定精度

5-4 複合的な姿勢角変化が推定精度に与える影響

実際に視線による意思表示システムを使用する場面では、ベッド上の使用者の横で介助者がタブレットを向ける状況も想定される。アームで画面を固定する場合と比べ、基本姿勢からのずれが生じやすくなると考えられる。そこで、頭部のヨー角、ピッチ角、ロール角が複合的に変化した場合の実験を行った。Fig. 18 に示されるように、ヨー角右 10 度、ピッチ角上 15 度、ロール角右 15 度においては、最悪ケースでも 67.3 %の精度で領域推定され、36 点の平均では 92.5 %の精度で領域推定が行えることを確認した。

さらに、これに移動平均処理を施した結果を Fig. 19 に示す。推定精度が、10 フレーム移動平均で 96.8 %、

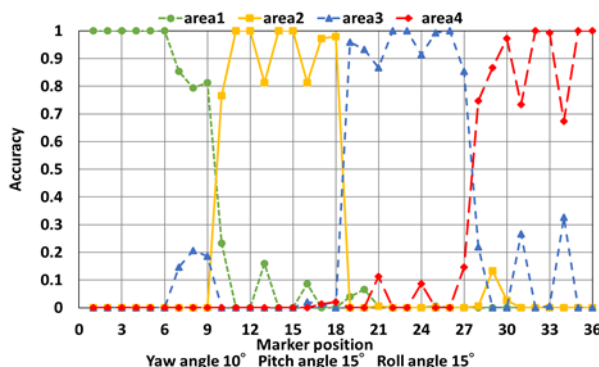


Fig. 18 ヨー角 10 度、ピッチ角 15 度、ロール角 15 度における推定精度

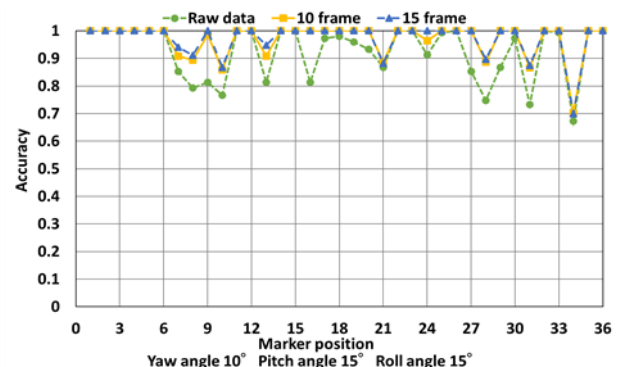


Fig. 19 移動平均処理を施した際の推定精度の推移

15 フレーム移動平均で 97.3 %に改善されることが分かる。15 フレームは、約 0.5 秒に相当する。したがって、ある領域を一定時間注視することで意思表示を行う場合、移動平均処理を施すことで、タイムラグの影響をほとんど受けることなく、推定精度を向上させることができる。今回の被験者では、ヨー角 10 度、ピッチ角 15 度、ロール角 15 度までは視線領域推定による 4 択での意思表示が可能であるといえる。ただし、確証的に論じるためには、今後、複数の被験者で実験を重ねる必要がある。また、境界部においては、精度低下が観察される点に注意が必要である。汎用的なアプリケーションに用いる際には、境界付近に不感帯を設けるなどの対応策も併用することが有効であると考えられる。

6 意思表示支援アプリケーションの開発

6-1 開発環境

意思表示アプリケーションの開発環境を Tab. 3 に示す。今回は、視線領域推定を Python 側で行い、アイコンと組み合わせた GUI 画面をブラウザで表示するシステム構成とした。Python には Tkinter や wxPython などの GUI フレームワークや GUI モジュールが存在する。しかし、これらのツールではレイアウトの調整に限界があり、アニメーションや複雑な GUI の作成は困難であるため、直感的に操作を理解できるようにしユーザビリティを向上させる必要がある。

そこで、Eel を用いて Python と JavaScript 間でのデータの受け渡しを行えるようにした。これにより、視線領域推定を Python で行い、それに連動した画面描画をブラウザで行うことができる。

Tab. 3 開発環境

環境	バージョン等
OS	Windows10 (Ver.1909)
GPU	Inter Iris Plus Graphics
開発言語 1	Python (3.8.6)
開発言語 2	JavaScript
使用ライブラリ	OpenCV (4.4.0)
	TensorFlow (2.3.1)
	Keras (2.4.3)
	Eel (0.14.0)

6-2 インターフェースデザインについて

視線入力を行うインターフェースにおいては、漫然と画面を見ているのか、選択するために注視しているのかを区別する必要がある。ユーザが求めるアイコンを探し出している最中に、意図していない領域が選択されるような Midas touch problem[12]を抑制することが求められる。そこで、Jacob らが提案している一定時間注視することで選択を決定する手法を採用した。また、今回作成したアプリケーションでは、注視している領域をユーザにフィードバックするために円形のメータを用いている。メータの増加により、限られた表示スペース内で時間の経過を視覚的に伝える効果が期待される。また、注視領域のアイコンの形を四角から丸に変更することで、システムが注視領域をどのように判断しているのかをユーザが直感的に認識できるようにした。

6-3 意思表示支援アプリ

アプリの動作画面を Fig. 20(a)～(e)に、離れた場所にいる介助者等に意思を伝えるための LINE 通知画面を Fig. 20(f)に示す。アプリ起動時に Fig. 20(a)が表示され、「開始」を選択することで意思表示システムが動作を開始する。Python 側の視線領域推定エンジンにより注視領域の推定が逐次行われ、推定された領域の JavaScript 関数が呼び出されることにより、アイコンがアニメーション動作する。4つの領域のうち、右上の領域はアイコン選択肢の切り替えに用いられ、意思表示の選択肢を拡張できるようにしている。それ以外の3領域に表示されたアイコンにより意思の表示を行う。この3領域については、所定の時間連続して注視することで Fig. 20(e)に示される確認画面に切り替わる。ここで左側の○領域を選択すると最終決定となり、音を鳴らすと同時に事前に登録されている LINE ユーザにメッセージを送信する。LINE に通知することで、



Fig. 20 操作画面(a)～(e)とLINE通知画面(f)

介助者が常に近くにいなくても意思を伝えることが可能になる。また、着信履歴が残るため見落としの軽減に繋がり、生活の振り返りにも用いることができる。選択に必要な注視時間やアニメーションの種類については、個人の特性や好みもあるため、変更できるようにする必要があると考えている。

7 おわりに

本研究では、CNN を用いてキャリブレーションを必要としない視線領域推定システムを構築した。予備実験の結果から、想定している利用者の知的・身体的特性やシステムの信頼性を考慮して画面領域を4分割とし、CNNモデルにはXceptionを採用した。

次に、MPIIGaze データセットを用いて学習を行った後、カメラ映像からリアルタイムに視線領域推定を行うシステムを構築した。画面上に36か所の注視点を設け、カメラと顔までの相対的な距離や位置、および顔の姿勢角を変化させて評価実験を行った。その結果、基本姿勢においてはカメラまでの距離 $L=1200\text{ mm}$ まで、顔位置については基本姿勢 ($L=700\text{ mm}$) を維持した状態で 100 mm 下方まで、キャリブレーションフリーで一定の精度が得られ実用的であることが示された。アプリケーション化の際には、移動平均を取るなどの対策により、領域境界付近の瞬間的な誤認識抑制が可能になることも示された。以上より、4択程度までは特殊

なハードウェアを用いることなく、キャリブレーションフリーで意思表示支援が可能であるといえる。

また、姿勢角の変化については、1名の被験者において、ヨー角については10度、ピッチ角については15度、ロール角については15度まで対応できることが確認された。これにより、厳密な基本姿勢でなくとも視線領域推定が行える可能性が示されたと考えられる。ただし、姿勢角の変化については、さらに複数の被験者で実験を重ね、考察する必要がある。

この視線推定システムを用い、アイコンを一定時間注視することで、離れた場所にいる介助者に意思を伝えられるアプリケーションを開発した。EelによってPythonとJavaScriptを連携動作させることでGUI画面を実現している。視線による入力、注視時間を長く設定すると利便性が損なわれユーザのストレスとなる。一方、短すぎるとMidas touch problemが生じやすくなるという問題がある。今後は、注視時間、アイコンやアニメーションの種類等を介助者で自由に設定できるように改良し、実用性を高める必要がある。その後、個人適応性にも留意しつつ、重度重複障害者の方に使用してもらおう社会実装実験へと展開したい。

謝辞

本研究にご助言いただいた熊本県立黒石原支援学校の皆様に深く感謝申し上げます。本研究は公益財団法人電気通信普及財団の助成を受けたものです。

【参考文献】

- [1] 伊藤和幸, 数藤康雄, 他. 重度肢体不自由者向けの視線入力式コミュニケーション装置. 電子情報通信学会論文誌. 2000, 83(5), p.495-503.
- [2] 新井康平, 上滝寛美. 利用者の動きを許容する黒目中心検出による視線推定に基づくコンピュータ入力システム. 電気学会論文誌. 2007, 127(7), p.1107-1114.
- [3] 後関利明, 星飛雄馬, 他. Tobii 社製眼球運動計測装置. 神経眼科. 2016, 33(1), p.47-52.
- [4] 文部科学省. 特別支援学校学習指導要領解説総則等編 (幼稚部, 小学部, 中等部). 開隆堂出版, 2018, 465p.
- [5] 柴里弘毅, 大塚弘文. 学外協力者とのニーズ発掘を起点とする社会実装プロジェクトの実践. 工学教育. 2015, 63(1), p.48-53.
- [6] 柴里弘毅, 大塚弘文. ニーズをとらえ解決策を生み出す社会実装型の学生教育—熊本県立黒石原支援学校における教材・支援機器開発—. 独立行政法人国立高等専門学校機構, 論文集高専教育, 2015, 38, p.135-140.
- [7] University of Edinburgh, “Eye gaze software curve”, <https://www.callscotland.org.uk/common-assets/cm-files/posters/eye-gaze-software-curve.pdf>, (参照 2020-9-2)
- [8] Zhang, X.; Sugano, Y.; et al. “Appearance-based gaze estimation in the wild”. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, 2015-07-07/12, p.4511-4520, doi:10.1109/CVPR.2015.7299081.
- [9] Liu, S.; Deng, W. “Very deep convolutional neural network based image classification using small training sample size”. Proceedings of 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). Kuala, Lumpur, 2015-11-03/06, p.730-734, doi: 10.1109/ACPR.2015.7486599.
- [10] Szegedy, C.; Liu, W.; et al. “Going deeper with convolutions”. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, 2015-07-07/12, p.1-9, doi: 10.1109/CVPR.2015.7298594.
- [11] Chollet, F. “Xception: deep learning with depthwise separable convolutions”. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, 2017-07-21/26, p.1800-1807, doi: 10.1109/CVPR.2017.195.
- [12] Jacob, R.J.K.; “What you look at is what you get: eye movement-based interaction techniques”, Proceedings CHI’90, pp.11-18, ACM Press, doi: 10.1145/97243.97246.

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
畳み込みニューラルネットワークを用いたリアルタイム視線領域推定システムの構築と評価	人 間 工 学 , Vol.56, No.5, pp.181-190, 2020	2020 年 10 月
視線による意思表示支援アプリケーションの開発	令和 2 年度電気学会九州支部高専研究発表会講演論文集, pp.55-56, 2021	2021 年 3 月