

高齢者向け視線推定技術開発のための データセット構築と基礎実験

中川 博貴^{1,a)} 青木 工太^{1,b)} 榎原 靖^{1,c)} 八木 康史^{1,d)}

概要：近年の新型コロナウイルス感染症の拡大や高齢化社会などに伴い、ICTを用いたオンライン高齢者支援システムが注目されている。リモートケアシステムに顔画像解析技術を応用することで、視線推定による集中度の測定や、表情認識によるうつ病などの健康の変化への気づきなどのさらなる発展が期待できる。本研究では、高齢者に特化した視線データセットを構築し、さらにその有効性を検証するため、深層学習手法を用いた視線推定に関する基礎実験を行った。視線データ収集実験では計128人のデータを収集し、視線情報の真値をつけられた顔画像が計235,107枚得られた。そして、VGG-16をファインチューニングして視線方向のクラス分類を用いた簡易的な視線推定を行った結果、誤差は4.90cmであった。また、推定結果から、アビアランスに基づく視線推定において瞳孔の見え方の重要性が確認でき、加齢による瞼や皮膚のたるみといった問題を抱える高齢者の視線推定が従来手法で十分な精度が得られるかどうかを検証する必要性が確認できた。

キーワード：視線推定, 視線データセット構築, オンライン高齢者支援

1. はじめに

近年の新型コロナウイルス感染症の拡大や高齢化社会に伴い、タブレット端末等を用いたICTによるオンライン高齢者支援が注目されている。例えば、フィンランドのヘルシンキ市では慢性的な看護職員不足への対策として、ビデオ会議サービスによる在宅看護を取り入れている[1]。同市のリモートケアでは、在宅医療を受ける高齢者のために、看護師がビデオ会議サービスを通じて健康観察や服薬管理を行うサービスを提供している。

このようなサービスに顔画像解析技術を適用すれば、さらなる発展が期待できる。例えば、視線推定による服薬指導などの説明中における集中度の測定や、表情認識によるうつ病などの健康の変化の気づきといった機能である。そのほかにも、オンラインによる同意取得においても視線推定技術は有用である。電子機器の操作に不慣れである高齢者が端末を使用して同意を行う際、説明中の集中度の測定に加え、端末操作中の視線推定によって同意の信憑性を担保することができる。

しかし、以上のような機能を既存技術で実現するには問題がある。現在公開されている視線データセットには20代から40代の若年層の被験者のデータが多く、高齢者のデータを多く含むものは存在しない。特に、視線推定において白目・黒目の見え方が重要だと考えられるが、加齢による瞼や皮膚のたるみといった高齢者特有の課題に従来手法で正しく推定することは難しいと考えられる。本研究では、高齢者向け視線推定技術開発のために加え、高齢者の視線推定が従来手法で十分に可能かどうかを検証するためにも、高齢者に特化した視線データセットを構築することを目的とし、さらにその有効性を検証するため、既存の深層学習ネットワークをファインチューニングし、視線方向のクラス分類による基礎実験を行う。

2. 関連研究

2.1 視線推定のアルゴリズム

視線推定の手法はモデルに基づく手法とアビアランスに基づく手法に分けられる[2]。各手法については2.1.1節と2.1.2節で説明する。さらに、人物依存と人物非依存の2つに分類できる。人物依存は、特定の人物の視線を推定するためにその人物に合わせて最適化する手法である。人物非依存は、一度学習すれば再学習等を行わなくても誰に対しても同じように推定できることを目指す手法である。

¹ 大阪大学
Osaka University

a) nakagawa_h@am.sanken.osaka-u.ac.jp

b) aoki.k@am.sanken.osaka-u.ac.jp

c) makihara@am.sanken.osaka-u.ac.jp

d) yagi@am.sanken.osaka-u.ac.jp

2.1.1 モデルに基づく手法

モデルに基づく手法は目を幾何学的にモデル化し、視線を推定するアプローチである。この手法では、さらに、目の特徴を検出するために外部光源を必要とするかどうかによって、角膜反射法と形状に基づく手法に分けることができる。角膜反射法は赤外光を眼球に照射し、その反射像(ブルキニエ像)から角膜球の中心、半径を推定する。その後、観測される瞳孔の中心点と角膜球の中心とを結んだ直線を視線とする。一方、形状に基づく手法は瞳孔の中心や虹彩の輪郭、目尻などの観測された目の形状から直接視線方向を推測するものである。これらの手法は赤外照明装置を必要とすることや低画質、照明条件の変化に頑健でない。

2.1.2 アピアランスに基づく手法

アピアランスに基づく手法は直接目の画像を入力とし、視線を推定する。そのため、低解像度の眼球画像でも動作するが、モデルに基づく手法に比べて大規模な学習データが必要となる。しかし、最近の研究 [3] では、アピアランスに基づく手法はモデルに基づく手法よりも優れた視線推定精度が得られることが示されている。これは、大規模なデータセットと最新の深層学習技術の利用により可能となったものである。

本研究では、高齢者が使用するタブレット端末のフロントカメラで撮影された画像から視線推定を行うことを想定しているため、アピアランスに基づく手法を用いる。

2.2 データセット

利用可能な視線データセットはすでに多数存在する [4], [5], [6], [7], [8], [9], [10]。主なデータセットの特徴を表 1 にまとめる。

アピアランスに基づく視線推定は日常的で自然な環境において良好な性能を発揮すると考えられている。すなわち、実験室環境下における過度に明るい照明や、正確な頭部の姿勢を前提として収集されたデータセットでは実用性に欠ける。

これらのデータセットは高齢者を多く含むわけではなく、加齢による眼や皮膚のたるみといった、視線推定における高齢者特有の課題に対して不十分であると考えられる。したがって、高齢者を対象とした顔画像解析技術の開発のために加え、既存技術が高齢者特有の課題に対応できるかどうかを検証するためにも、新たに高齢者に特化したデータセットを構築する必要がある。2022年1月28日の時点では被験者は128人であるが、データ収集は少なくとも今後1年間は継続され、最終的には500人分のデータが得られる予定である。

3. 視線データ収集

本章では、タブレットの画面に表示される注視点をタッチする実験による視線データ収集手法と、アノテーション

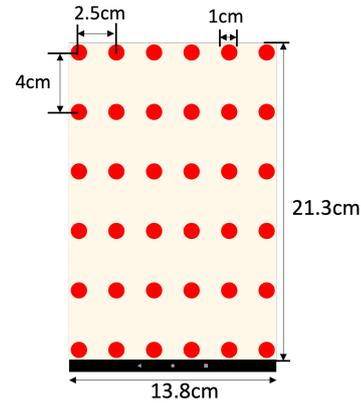


図 1 視線データ収集アプリの画面

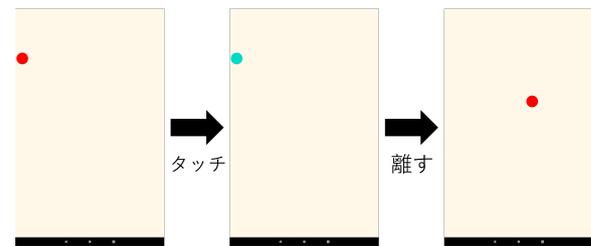


図 2 注視点表示の画面遷移

ン手法、構築したデータセットの特徴について述べる。なお、本データ収集は、大阪大学産業科学研究所研究倫理委員会の承認を得ている (R3-07「デュアルタスク歩行解析やヒューマンセンシングによる年齢関連指標および認知能力・身体能力の推定」)。

3.1 データ収集手法

3.1.1 データ収集実験に用いた Android アプリの仕様

本研究では、Android タブレットの NEC LAVIE Tab E 10FHD2 モデルを用いて視線データの収集実験を行った。タブレット本体のサイズは縦 24.42cm × 横 15.53cm、スクリーンのサイズは縦 21.3cm × 横 13.8cm である。

図 1 に視線データ収集のために開発した Android アプリの画面を示す。タブレットのスクリーンに 6 行 6 列、水平方向 2.5cm、垂直方向 4cm の間隔で、直径 1cm の注視点を合計 36 個、均等に配置する。データ収集実験の際には、この中の注視点のうち 1 つがスクリーンに表示され、ペン先 4.5mm のタッチペンを用いて注視点をタッチする。図 2 に注視点表示の画面遷移を示す。データ収集における注視点の挙動は以下のようにした。

- (1) 注視点は押下した後に離されるまで、同じ場所に、同じ大きさのまま表示され続ける。
- (2) 押下されている間、注視点の色が赤色から水色に変化する。
- (3) 離されると注視点が消え、同時に別の注視点が表示される。

被験者が先に次の注視点を予測することを防ぐため、次に

表 1 既存データセットの特徴

| データセット | 被験者数 | 頭部姿勢 | 注視点 | 照明条件 | データ数 | 解像度 |
|------------------|-------|------|---------|------|-----------|-------------|
| MPIIGaze[4] | 15 | 連続 | 20 | 日常生活 | 213,659 | 1280 × 1024 |
| GazeCapture[5] | 1,474 | 連続 | 13 + 連続 | 日常生活 | 2,445,504 | 640 × 480 |
| Columbia[6] | 56 | 5 | 21 | 1 | 5,880 | 5184 × 3456 |
| RT-GENE[7] | 15 | 連続 | 連続 | 日常生活 | 122,531 | 1920 × 1080 |
| UT Multi-view[8] | 50 | 8 | 160 | 1 | 64,000 | 1280 × 1024 |
| ETH-XGaze[9] | 110 | 18 | 525 | 16 | 1,083,492 | 6000 × 4000 |
| TabletGaze[10] | 51 | 4 | 35 | 日常生活 | ビデオ | 1280 × 720 |
| 本研究 | 128 | 1 | 36 | 4 | 213,435 | 1080 × 1920 |



図 3 タブレットと被験者の位置関係



図 4 カメラと照明の配置

表示される注視点はランダムに表示される。制限時間を 1 分として注視点をタッチし続ける作業を 1 回の計測とし、1 回の計測の間に注視点が表示される回数に偏りはなく、36 個の注視点が満遍なく表示される。

被験者がデータ収集実験を行っている際、バックグラウンドでタブレットのフロントカメラで撮影を行い、被験者がスクリーンをタッチするごとに以下の項目を記録する。

- スクリーンをタッチした時刻
- 計 36 個の注視点の番号
- 表示されていた注視点の中心座標
- 被験者がタッチした座標
- 注視点をタッチできたかどうか

3.1.2 実験環境

図 3、図 4 に示すように、机の上にタブレットスタンドを設置し、床からタブレットの上端までが 114cm の位置にタブレットを設置する。被験者はタブレットの正面に着席し、カメラから約 50cm 離れた状態でタッチペンを用いてタブレットを操作する。この実験を行っている被験者の様子を、タブレットのフロントカメラとその上部に設置したウェブカメラ (logicool C920n PRO HD WEBCAM) で撮影し、静止画列を保存する。タブレットのフロントカメラの解像度は 1920 × 2560、フレームレートは 2fps である。一方、ウェブカメラの解像度は 1080 × 1920、フレームレ

率は 20fps である。タブレットのフロントカメラのみではなくウェブカメラも用いた理由は、タブレットのカメラで 1 フレームをキャプチャし、ストレージに書き込むまでの処理に約 400 ミリ秒要する実装となっているため、フレームレートが 2fps と低く、タブレットのフロントカメラのみでは十分なデータ数が取得できないためである。

また、本研究における照明条件は、照度が強い場合 (約 300 ルクス) と弱い場合 (約 100 ルクス)、また、それぞれの場合における近赤外線照明の有無を組み合わせた計 4 通りである。照度の調整は実験室の蛍光灯の調光で行い、照度の計測は机上に照度計を置き、実験日の午前と午後の 2 回に分けて計測する。アピアランスに基づく手法にもかかわらず、被験者の正面に近赤外線照明を設置した理由は、実験室の環境下で撮影された画像では、被験者によっては目の領域が影となり、黒目と白目の区別がつかない場合があったためである。特に高齢者の場合、加齢による皮膚のたるみにより、上記のような現象は多くの被験者に見られた。

3.1.3 実験の流れ

実験の流れは、以下のとおりである。

- (1) 電子機器の操作に不慣れである高齢な被験者に対して、Android アプリの操作方法とタッチペンの使用方法、実験内容の説明を行う。

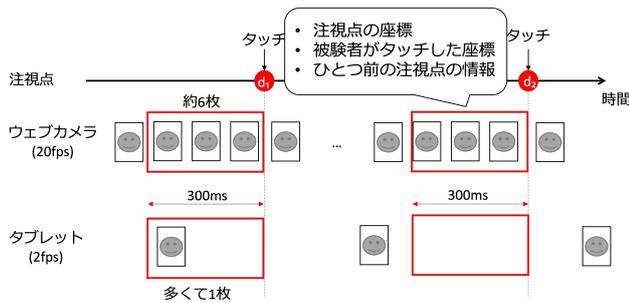


図 5 視線情報のアノテーション方法

- (2) 30 秒間の練習モードで実験を行う。この際も、データ収集を行う。
 - (3) 休憩を挟んだ後、1 分間の計測モードで実験を行う。
 - (4) 照明条件を変更し、3. を残り 3 回繰り返す。
- ひとりの被験者に対して、30 秒間の練習を 1 回と、1 分間の計測を計 4 回を行い、合計 4 分 30 秒間データを収集する。

3.2 アノテーション

3.2.1 視線情報のアノテーション方法

収集した画像データに対する視線情報のアノテーションの方法について説明する。注視点をタッチしようとするとき視線はその注視点に向いているという仮定から、図 5 に示すように、被験者が注視点をタッチした時刻より 300 ミリ秒前からキャプチャされた画像に対してアノテーションを行う。ウェブカメラは 20fps でキャプチャするため、被験者が 1 度スクリーンをタッチすることによって得られる視線情報につき約 6 枚の画像データが抽出でき、タブレットでは 2fps でキャプチャするため、多くて 1 枚の画像が抽出できる。抽出された画像に対し、3.1.1 節で示したデータ収集中に記録した項目に加えて、以下に示す項目を紐づける。

- ひとつ前に表示されていた注視点の番号
- ひとつ前に表示されていた注視点の中心座標
- ひとつ前に被験者がタッチした座標
- 注視点が表示されてから経過した時間

3.2.2 PC とタブレットの正確な時刻の同期方法

被験者が注視点をタッチしたタイムスタンプはタブレットのシステムから取得した時刻であり、ウェブカメラの画像に記録されているタイムスタンプは PC のシステムから取得した時刻である。注視点をタッチした時刻より 300 ミリ秒前からキャプチャされていた画像に対してアノテーションを行うため、PC とタブレットは正確に時刻が同期されていなければならない。PC が参照する NTP サーバを Android タブレットが参照している pool.ntp.org に統一したが、300 ミリ秒以上の時刻のずれが生じることがあった。そこで、時刻をミリ秒まで表示するウェブサイトを作成し、図 6 のように PC とタブレットで時刻を表示させ、

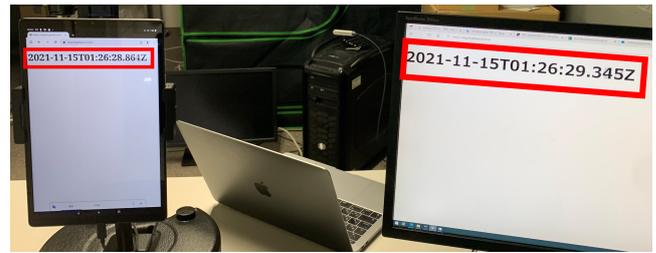


図 6 PC とタブレットの正確な時刻の同期方法

実験の開始時と終了時に時刻のずれを記録しておく。そして、時刻のずれを修正した上でアノテーションを行う。

3.3 データセットの特徴

2021 年 11 月 22 日から 2022 年 1 月 24 日までの実験で、合計 128 人のデータを収集した。男性の合計が 78 人、女性の合計が 50 人であり、最年少が 65 歳、最年長が 86 歳、平均 73.5 歳の大規模な高齢者のみの視線データセットが構築できた。また、3.2.1 節で示したアノテーションを行い、視線の真値をつけられた画像の枚数は、ウェブカメラで撮影された画像が 213,435 枚、タブレットのフロントカメラで撮影された画像が 21,654 枚である。

4. 視線推定ネットワーク

構築したデータセットの有用性を評価するための基礎実験として、顔画像 (300 × 300) のみから表示されていた注視点の番号を推定する 36 クラス分類と、図 8 に示すように注視点の粒度を落として 9 クラス分類を行う。

4.1 VGG-16 のファインチューニング

図 7 に本研究で使用したネットワークを示す。本研究で使用したネットワークとオリジナルの VGG-16[11] のネットワーク構造との差分は、入力層と全結合層である。また、図 7 における最終全結合層の FC-n の n の値は、5.2 節と 5.3 節で行うクラス分類のクラス数、すなわち 36 または 9 である。VGG-16 は、畳み込み層が 13 層と全結合層が 3 層の計 16 層からなる畳み込みニューラルネットワークであり、画像認識に優れたネットワークである。前半の畳み込み層の 13 層は ImageNet[12] と呼ばれる大規模画像データセットにより事前学習を行った。そこで、図 7 の Conv_block1 から Conv_block4 までの重みを固定し、収集した視線データセットを用いてファインチューニングを行う。

5. 実験

5.1 実験設定

視線推定モデルの汎化性能を評価するため、2021 年 11 月 22 日から 2022 年 1 月 24 日までのデータ収集実験で収集した 128 人のデータを 3 つのサブセットに分割し、3 分

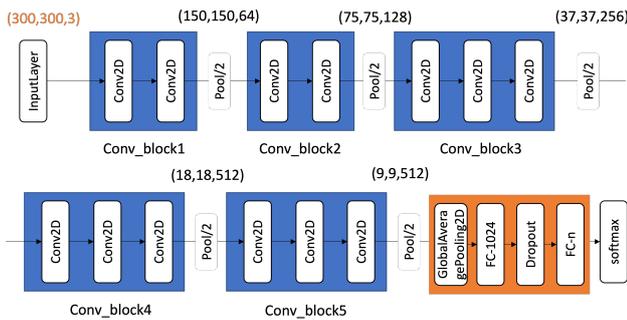


図 7 本研究で使したネットワーク

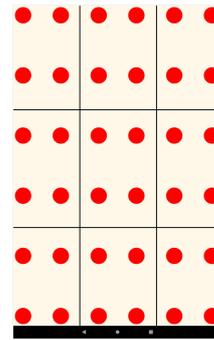


図 8 9 クラスの定義

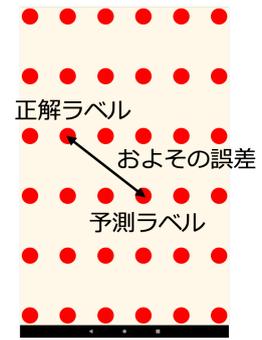


図 9 およその誤差の定義

割交差検証を行う。3つに分割したサブセットのうち、2つのサブセットを用いて学習を行い、もうひとつのサブセットを用いて評価を行う。学習データセットとテストデータセットの組み合わせを変えて、計3回の学習と評価を繰り返す。バリデーションデータは、学習データセットに含まれるすべての被験者のデータから均等に抽出し、訓練データ:バリデーションデータ=9:1になるように設定する。3つのサブセットに含まれる被験者数とデータ数を以下の表2に、3通りの訓練データ、バリデーションデータ、テストデータ数を表3に示す。ただし、本研究では、タブレットで撮影された画像とウェブカメラで撮影された画像の正規化を行っていないため、学習に用いるデータはウェブカメラで撮影された計213,435枚のデータのみを用いる。

表 2 3-分割交差検証のためのデータセットの分割

| サブセット | 人数 | データ数 |
|-------|----|--------|
| A | 43 | 62,363 |
| B | 43 | 78,639 |
| C | 42 | 72,433 |

表 3 3通りの訓練、バリデーション、テストデータ数

| 学習/評価データセット | train | val | test |
|-------------|---------|--------|--------|
| AB/C | 126,901 | 14,101 | 72,433 |
| AC/B | 121,901 | 13,480 | 78,639 |
| BC/A | 135,964 | 15,108 | 62,363 |

36クラス分類、9クラス分類における精度の評価基準として、

$$\text{再現率 (\%)} = \frac{\text{正検出数}}{\text{正検出数} + \text{検出漏れ数}} \times 100$$

を用いる。また、36クラス分類の場合のみにおいて、図9に示すような、およその平均誤差 (cm) = 正解ラベルが示す注視点の中心座標と予測ラベルが示す注視点の中心座標のユークリッド距離を用いて評価を行う。

表 4 36クラス分類の3-分割交差検証の結果

| 学習/評価データセット | 学習時の再現率 (%) | 評価時の再現率 (%) |
|-------------|-------------|-------------|
| AB/C | 76.5 | 15.9 |
| AC/B | 71.4 | 16.4 |
| BC/A | 77.5 | 17.8 |

5.2 36クラス分類による視線推定結果

表3に示した3通りのデータセットを用いた学習、評価結果を以下の表4に示す。

学習データをサブセットBとCとして学習し、テストデータをサブセットAとして評価したモデルの精度が17.8%で最も高かった。このモデルを用いて、テストデータであるサブセットAに含まれる62,363枚に対して行った視線推定の結果を以下の図10に示す。図10は、図1の36個の注視点の配置に基づいて、正解ラベルごとに予測ラベルの割合を示したものである。黒枠で囲っているマス目は正解ラベルが示す注視点の位置である。例えば、図10における左上のヒートマップは、正解ラベルがタブレットのスクリーン左上の注視点であるテストデータに対する予測ラベルの割合であり、正しく推定できている割合が約60%で、右隣の注視点に約10%、ひとつ下の注視点に約20%の割合で誤った推定を行っていることがわかる。その他のヒートマップを見ると、正解ラベルがタブレットのスクリーンの外枠の注視点である場合は予測が正解位置周辺に集中している。しかし、正解ラベルがスクリーンの中央付近の注視点である場合は予測がばらついていることがわかる。

また、このときのおよその平均誤差は4.90cmであった。本研究と同様にモバイルデバイスを用いて収集されたGazeCapture[5]とTabletGaze[10]と精度を比較する。GazeCaptureはデータ収集用のアプリを公開してクラウドソーシングを行い、iPhoneやiPadの様々な種類の端末でデータ収集を行ったデータセットである。一方、TabletGazeは本研究と同様に1台のタブレット端末を用いてデータ収集を行っている。この2つの先行研究は画面上の注視点の位置を推定しているのに対して、本研究では注視点の範囲を推定しているという違いはあるが、それぞれの誤差と注視点の最近傍点との距離で正規化した誤差を以下の表5

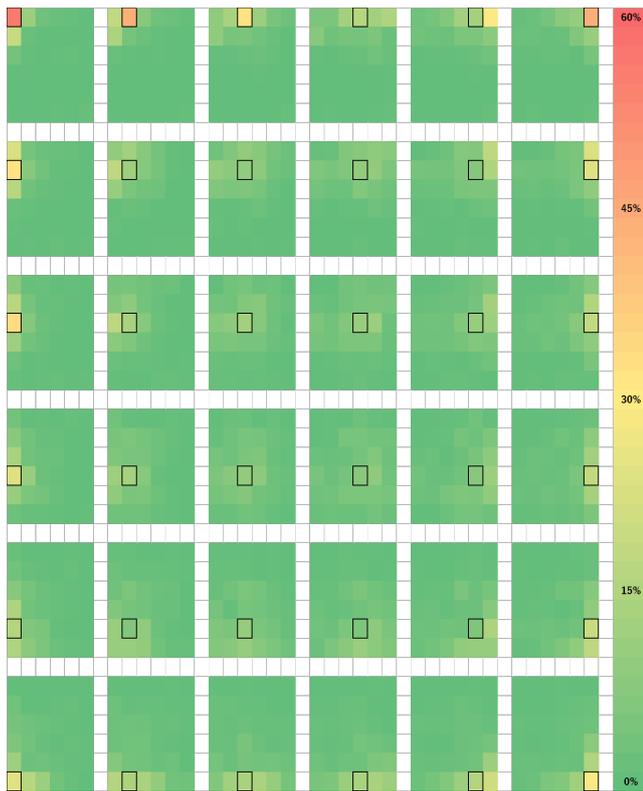


図 10 36 クラス分類を用いた、サブセット A に対する視線推定結果における正解ラベルごとの予測ラベルの割合

に示す。GazeCapture は iPad でも様々な機種が用いられているため正規化ができないが、TabletGaze では 3.4cm で、本研究は 2.5cm で正規化を行った。正規化した値が、TabletGaze は 1 に近いため、1 つ離れた注視点と間違ふことがあるのに対し、本研究は値が 2 に近いため、平均的に 2 つ離れた点と間違ふことがある。

表 5 先行研究との精度比較

| | 誤差 (cm) | 最近傍点との距離で正規化した誤差 |
|-------------|---------|------------------|
| GazeCapture | 2.53 | - |
| TabletGaze | 3.17 | 0.932 |
| 本研究 | 4.90 | 1.96 |

また、リモートケアシステム等の実際の利用環境を想定し、タブレットで撮影された画像に対する視線推定の性能を検証するため、同じモデルを使用してタブレットのフロントカメラで撮影された画像 21,654 枚をテストデータとして視線推定を行った。このときの再現率は 17.7%、およその平均誤差が 5.58cm、最近傍点との距離で正規化した誤差が 2.23 であった。図 10 と同様に、正解ラベルと予測ラベルの割合を図 11 に示す。

5.3 9 クラス分類による視線推定結果

表 3 に示した 3 通りのデータセットを用いた学習、評価結果を以下の表 6 に示す。

学習データをサブセット A と C として学習し、テス

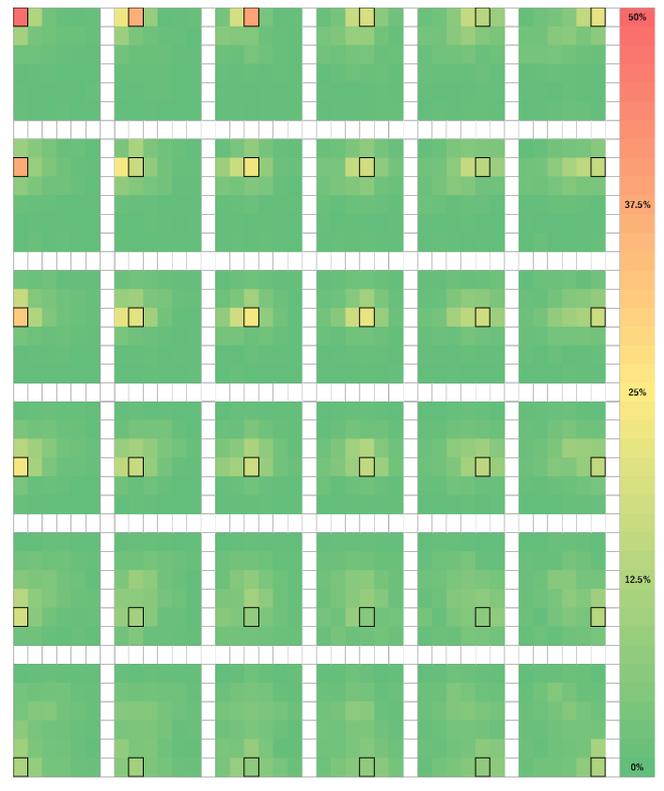


図 11 36 クラス分類を用いた、タブレットで撮影されたデータに対する視線推定結果における正解ラベルごとの予測ラベルの割合

表 6 9 クラス分類の 3-分割交差検証の結果

| 学習/評価データセット | 学習時の再現率 (%) | 評価時の再現率 (%) |
|-------------|-------------|-------------|
| AB/C | 90.8 | 41.3 |
| AC/B | 91.3 | 45.6 |
| BC/A | 92.2 | 43.1 |

トデータをサブセット B として評価したモデルの精度が 45.6% で最も高かった。このモデルを用いて、テストデータであるサブセット B に含まれる 78,639 枚に対して行った視線推定の結果を以下の図 12 に示す。図 12 は、図 10、図 11 と同様に正解ラベルごとの予測ラベルの割合を示したものである。黒枠で囲っているマス目は正解ラベルが示す注視点の区画の位置である。

また、同じモデルを使用してタブレットのフロントカメラで撮影された画像 21,654 枚をテストデータとして上と同じ実験を行った。このときの再現率は 35.5% であった。図 12 と同様に、正解ラベルごとの予測ラベルの割合を図 13 に示す。

5.4 クラス分類による視線推定結果の考察

36 クラス分類において未知のデータに対する再現率は 17.8% であり、粒度を落とした 9 クラス分類においても 45.6% にとどまり、精度は低い結果となった。図 10 から図 13 より、正解ラベルがタブレットのスクリーンの外枠の注視点であるテストデータに対する推定精度は比較的高いことが確認される。しかし、36 クラス分類と 9 クラス分

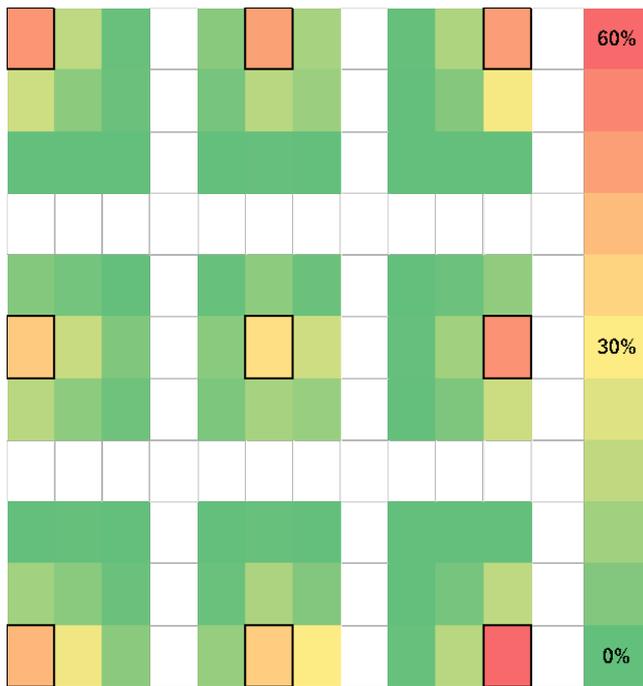


図 12 9 クラス分類を用いた, サブセット B に対する視線推定結果における正解ラベルごとの予測ラベルの割合

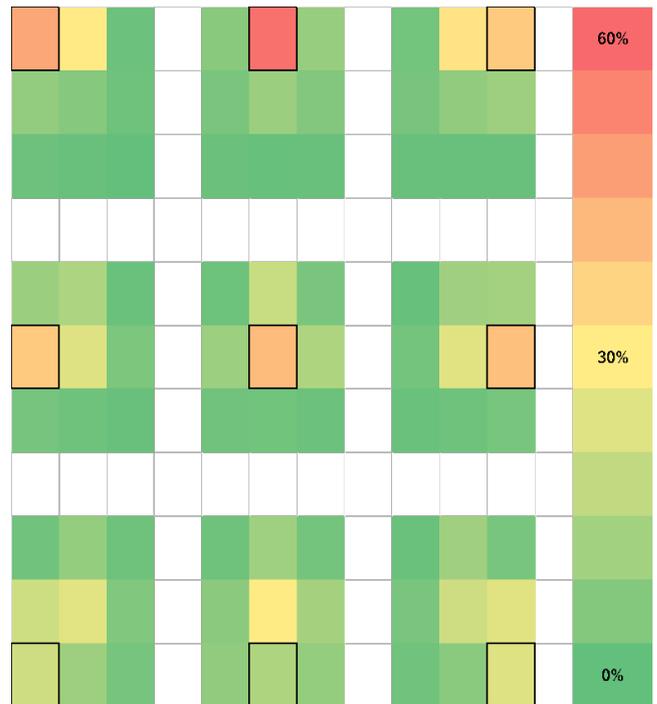


図 13 9 クラス分類を用いた, タブレットで撮影されたデータに対する視線推定結果における正解ラベルごとの予測ラベルの割合

類ともに中央付近への推定にはばらつきが確認される。これらの結果より、瞳の動きが大きいと予想されるスクリーン外枠にある注視点への視線精度は比較的高いが、瞳の動きの違いが小さいと予想されるスクリーンの中央付近にある注視点への推定精度は低いことがわかった。

また、画面上部と下部に対する推定の分布の違いを確認できる。画面上部に対する推定は上部に集中し、下部への誤った推定は少ない。一方、画面下部に対する推定は上部への誤った推定も確認できる。この精度の差の原因として考えられるのは、図 14 に示すように瞼の開き方による瞳領域の大きさの差によって推定の難易度が異なることだと考えられる。この結果からも分かるとおり、アピランスに基づく視線推定手法において瞳領域の大きさは推定精度への影響が大きいことが分かる。そのため、加齢による瞼や皮膚のたるみといったアピランスにおける若年者との違いを有する高齢者の視線推定において、既存技術で正確な推定が可能かどうかを検証する必要がある。

さらに、図 10 と図 11, 図 12 と図 13 を比較すると、タブレットで撮影されたデータをテストデータとして視線推定を行った結果のほうが、画面下部に対する推定の精度が低いことがわかった。これは、図 3 に示したタブレットの配置において、タブレットの下部が固定されていないことによるものだと考えられる。データ収集実験の際に、タブレット下部を強くタッチすることでわずかにタブレットが揺れ、その際にキャプチャされていた画像のブレが確認できた。



(a) 画面上部を見ているとき (b) 画面下部を見ているとき

図 14 目の周辺領域を切り取った画像

6. おわりに

本研究では、高齢者に特化した視線データセットを構築し、さらに基礎実験としてクラス分類による視線推定を行った。顔画像に対して後から視線の真値をつけることは難しいため、注視点をタッチするゲームを行う自作の Android アプリを用いてデータ収集を行い、アノテーションを行った。アノテーションでは、注視点をタッチする際、視線はその注視点に向いているという前提のもと、注視点をタッチした時刻より 300 ミリ秒前からキャプチャされた画像に対して、表示されていた注視点の座標や実際にタッチした座標などの視線情報を紐づけた。2021 年 11 月 22 日から 2022 年 1 月 24 日までの視線データ収集実験で計 128 人のデータを収集できた。視線情報の真値をつけられた顔画像は、ウェブカメラで撮影された画像が 213,453 枚、タブレットのフロントカメラで撮影された画像が 21,654 枚であった。このうちウェブカメラで撮影された顔画像データのみを用いて、表示されていた注視点のクラス分類問題として、深層学習手法による視線推定の基礎実験を行い、誤差は 4.90cm であった。

収集したデータおよび実験結果から、アピランスに基

づく視線推定手法における瞳領域の大きさの重要性が確認できた。高齢者の視線推定においては、加齢による瞼や皮膚のたるみといったことが課題となる可能性がある。今後の研究において、本研究で構築した高齢者の視線データセット、および、今後収集されるデータを用いて、年齢による視線推定精度の差異を検証し、さらに、高齢者に最適化された視線推定技術の開発に取り組む。

謝辞 本研究活動は、文部科学省による Society 5.0 実現化研究拠点支援事業（グラント番号: JPMXP0518071489）によって行われたものである。

参考文献

- [1] 中塚富士雄: 高齢者へのオンライン形式の支援に関する考察, 高齢者のケアと行動科学, Vol. 26, pp. 26–40 (2021).
- [2] Hansen, D. W. and Ji, Q.: In the Eye of the Beholder: A Survey of Models for Eyes and Gaze, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 3, pp. 478–500 (online), DOI: 10.1109/TPAMI.2009.30 (2010).
- [3] Zhang, X., Sugano, Y. and Bulling, A.: Evaluation of Appearance-Based Methods and Implications for Gaze-Based Applications, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 416 (online), DOI: 10.1145/3290605.3300646 (2019).
- [4] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A.: Appearance-Based Gaze Estimation in the Wild, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015).
- [5] Krafska, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W. and Torralba, A.: Eye Tracking for Everyone, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016).
- [6] Smith, B., Yin, Q., Feiner, S. and Nayar, S.: Gaze Locking: Passive Eye Contact Detection for Human-Object Interaction, *ACM Symposium on User Interface Software and Technology (UIST)*, pp. 271–280 (2013).
- [7] Fischer, T., Chang, H. J. and Demiris, Y.: RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments, *Proceedings of the European Conference on Computer Vision (ECCV)* (2018).
- [8] Sugano, Y., Matsushita, Y. and Sato, Y.: Learning-by-Synthesis for Appearance-Based 3D Gaze Estimation, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1821–1828 (online), DOI: 10.1109/CVPR.2014.235 (2014).
- [9] Zhang, X., Park, S., Beeler, T., Bradley, D., Tang, S. and Hilliges, O.: ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head Pose and Gaze Variation, *European Conference on Computer Vision (ECCV)* (2020).
- [10] Huang, Q., Veeraraghavan, A. and Sabharwal, A.: TabletGaze: dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets, *Machine Vision and Applications*, Vol. 28, No. 5, pp. 445–461 (2017).
- [11] Simonyan, K. and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* (2014).
- [12] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: Imagenet: A large-scale hierarchical image

database, *2009 IEEE conference on computer vision and pattern recognition*, Ieee, pp. 248–255 (2009).