

顔画像の視線計測による熟達度判別のための視線分析手法の検討

鎌原 淳三^{*1} 長松 隆^{*2} 廣江 葵^{*3} 伊藤 京子^{*4} 青柳 西蔵^{*5} 高田 健治^{*6}

^{*123} 神戸大学海事科学研究科 〒658-0022 兵庫県神戸市東灘区深江南町 5-1-1

^{*4} 大阪大学経営企画オフィス 〒565-0871 大阪府吹田市山田丘 1-1 共創イノベーション棟 401

^{*5} 東洋大学情報連携学部 〒115-0053 東京都北区赤羽台 1-7-11

^{*6} Ask-Prof オーラル・ヘルス・コンサルティング 〒530-0001 大阪府大阪市北区梅田 2-2-2

E-mail: ^{*1} kamahara@maritime.kobe-u.ac.jp, ^{*2} nagamatu@kobe-u.ac.jp, ^{*3} hiroe@stu.kobe-u.ac.jp,

^{*4} ito@lserp.osaka-u.ac.jp, ^{*5} uoziass@gmail.com, ^{*6} director@ask-prof.com

あらまし 医療における目視の外観検査である視診は診察の最初の段階に行われる重要なプロセスである。熟達した医師が行う視診の際の視線と経験が十分でない医学生や一般人では視線の動きが異なることが予想される。本研究では視線の動きから視診技能に関する熟達度を推定できるか検討するため歯科矯正医（以下、専門医）と歯科矯正に関する知識・経験を有さない成人（以下、一般人）を被験者として矯正患者の顔画像を閲覧する視線計測実験を行い、機械学習の手法によって熟達度の判別が行えるか検討を行った。検討の結果、専門医が部位別に設定した AOI(関心エリア)に基づく視線の頻度を用いることで専門医か一般人かを比較的良好に判別できることが分かった。

キーワード 視線計測, AOI, 線形判別分析, LDA, 視線分析, 機械学習

1. はじめに

医療における目視の外観検査である視診は診察の最初の段階に行われる重要なプロセスである。熟達した医師は最初の視診を的確な治療方針につなげるが、この技能が不足している場合には診断ミスや治療期間の長期化などの問題となる場合がある。このような視診技能の差異は視線に現れ、熟達した医師が行う視診の際の視線と経験が十分でない医学生や一般人では視線の動きが異なることが予想される。

本研究では視線の動きから視診技能に関する熟達度を推定できるか検討するため歯科矯正医（以下、専門医）と歯科矯正に関する知識・経験を有さない成人（以下、一般人）を被験者として矯正患者の顔画像を閲覧する視線計測実験を行い、機械学習の手法によって熟達度の判別が行えるか検討を行った。視線の分析に際して、関心グリッド頻度と関心エリア頻度という2つの視線位置集計方法を用いた。

2. 視線を用いた熟達度の判断

視線はその人の興味や暗黙的な知識によって同じ対象物を見るときでも異なった動きをすることが知られている[1]。筆者らは視線の動きに着目して視線情報を用いた暗黙の技能知識獲得と学習を行うことができるシステムの開発[2,3]などを行ってきた。その他、視線の動きによって閲覧者がどのような背景知識を持っているか評価する研究が行われている[4]。

一方、医療分野では視診技能の重要性が指摘されており、臨床の教育において合理的根拠に基づく効果的な診断技能の教授法や専門知識の獲得の度合いを測る方法が求められている。そこで我々の研究グループでは

熟達した診断プロセスの形式化に向けた視線の活用について研究を行っている[5,6]。

これらの研究の一環として、本稿では視診技能の熟達度として視線データから専門医か一般人かを判別できるかどうかの検討を行った。

3. 視線計測実験のデータ

視線の動きデータの計測は、眼球運動を計測する機器である Eye-tracking System を内蔵した Tobii Pro X2-30 [7]（トビー・テクノロジー株式会社製）を用いて行われた（サンプリングレート 30Hz、精密度 0.4°、正確度 0.32°）。実験用 PC の OS は Windows 10、ディスプレイは 23 インチ（解像度 1920 × 1080）であった。Tobii Pro X2-30 は、ディスプレイの中央下部に設置され、ディスプレイと眼球との距離が 62.5cm となるよう実験環境は設定されている。頭部固定が必要な場合は、顎台を利用した。図 1 はシステムの設置例であるが実際の実験風景ではない。



図 1 システムの設置例

Fig. 1 Experimental Settings.

(赤い円内に視線計測装置がある)。

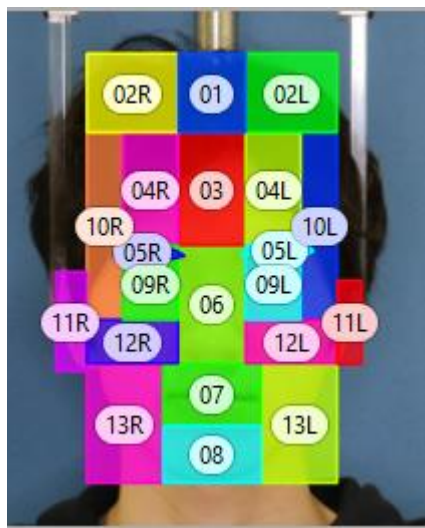


図2 正面顔 AOI
Fig. 2 Frontal Face AOI

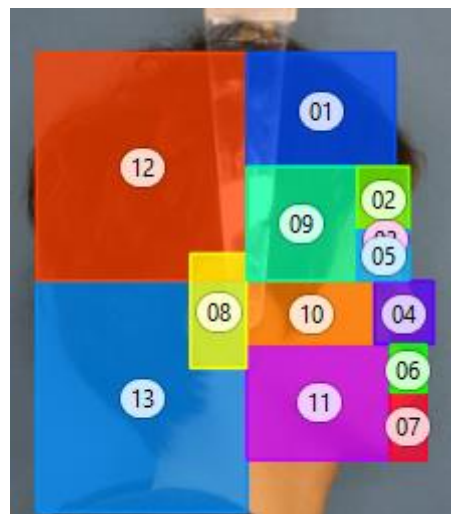


図3 横顔 AOI
Fig. 3 Profile AOI

被験者は専門医 15 名(男性 10 名, 女性 5 名;年齢範囲 24~67 歳), 一般人 15 名(男性 11 名, 女性 4 名;年齢範囲 20~25 歳)であった。視線計測に使用した顔画像は歯科矯正を要する患者 3 名の正面と横それぞれを合わせて 6 枚あり, 被験者毎に提示する患者の順序は入れ替えて平準化している。なお, 患者画像の利用について大学の倫理委員会の承認を得ている。

実験においては, 提示される顔画像についての情報は事前には提示せず顔画像を閲覧して判断(診断)し, 閲覧後口頭で回答するよう教示した, 実験中には口頭ではなく画面上の指示で操作を行ってもらうこととした。顔画像の閲覧時間に制限はなく, キーボードの操作で次の顔画像に進むように設定した。

Tobii Pro X2-30 では、Recording Time Stamp や Local Time Stamp, FixationPointX (MCSpx), FixationPointY (MSCpx), GazePointX(ADCSPx), GazePointY(ADCSPx), の他, 事前に装置に設定した AOI(関心エリア:Area Of Interests) にヒットしたどうかの情報が撮影フレーム毎に 1 行で記録される。Recording Time Stamp は視線計測を開始した時点からの時間(msec), Local Time Stamp は記録を開始した時点の時刻(時分秒.msec)であり, FixationPoint は注視したか(一定時間同じ場所を見たか)を GazePoint のデータを基に Tobii 社提供のソフトウェアで推定される座標である。今回視線のデータとしては生のデータに近い GazePoint の座標 (スクリーン座標) を用いた。サンプリングレートが 30Hz なので, 例えば 10 秒だと 300 行記録される。

AOI は特定の領域を関心エリアとして設定できるので, 共同執筆者の矯正歯科医の経験から本研究では以下の図 2, 図 3 のような領域を設定した。なお, Tobii のソフトウェアで AOI にヒットしたか出力されるが,

本稿では GazePoint の座標から算出した。

4. 視線分析手法

視線の分析には視線の位置に基づく位置測度 (Position Measure) や視線の経路に基づくスキャンパス (ScanPath) によるものなどがある[8]。

分析手法の検討においてスキャンパスによる分析も予備的に行ったが, スキャンパスは対象を見る順序を考慮するため「文章」のような見る順序に一定程度規則性があるものには適用できるものと思われるが, 本研究のように顔のどの部位を見るかについて特定の順序が想定できないこともあり, 予備的分析でスキャンパスでは専門医と一般人の分類が困難であった。

そこで本稿では位置測度に基づくエリアに視線がどの程度あったかという頻度に基づく分析を行った。

分析にあたっては, 画面上を等間隔の矩形領域 (グリッド) で分割したもので頻度を計測する方法 (関心グリッド頻度) と, 前記の AOI で頻度を計測する方法 (関心エリア頻度) を比較した。

1	2	3	10
11	12	13	20
						...	

図4 画面のグリッド分割 (例)
Fig. 4 Gridding of Screen (example)

4.1. 関心グリッド頻度

関心グリッド頻度では、顔が表示される画面を矩形領域のグリッドに分割する(図4)。グリッドは、左上から右下に順に番号が付けられる。各時間の視点は、いずれかのグリッドの1つに位置する。そこで、各グリッドは視線の位置となった数を数える。

グリッドの面積(大きさ)によって、グリッド毎の頻度も変わってくる。対象の顔の大きさとの関係もあり、そのためグリッドサイズを変化させながら分析する必要がある。使用したグリッドサイズは後述する。

例として、1名分の正面顔画像を対象とした視線データについてグリッドサイズを180(Grid=180)としたときの頻度マップを図5、図6に示す。図5は専門家、図6は一般人のものであり、各サブプロットの上の「Subject ?」の?は被験者番号となっている。なお、

図5の被験者番号の後のカッコ内は専門家の経験年数を示しており、nanは不明を意味する。顔画像の閲覧時間は被験者によって異なるため、各グリッドの頻度値を全グリッドの頻度値の合計で割ることにより正規化している。頻度マップでは頻度が低いところを青色で表し、頻度が高くなるにつれて赤っぽくなるように表している。

顔画像を表示する前には画面上の指示で中央を見ることとなっており、視線計測開始時点では中央から視線が開始するため、専門家・一般人いずれにおいても中央のグリッド頻度が高くなっている。

頻度マップを比較すると、一般人では頻度が分散する人がおり、顔画像を見る際に意図を持たずに閲覧することがあることが分かる。一方専門家の頻度マップでは、対象となっている顔画像が下顎部に治療が必要

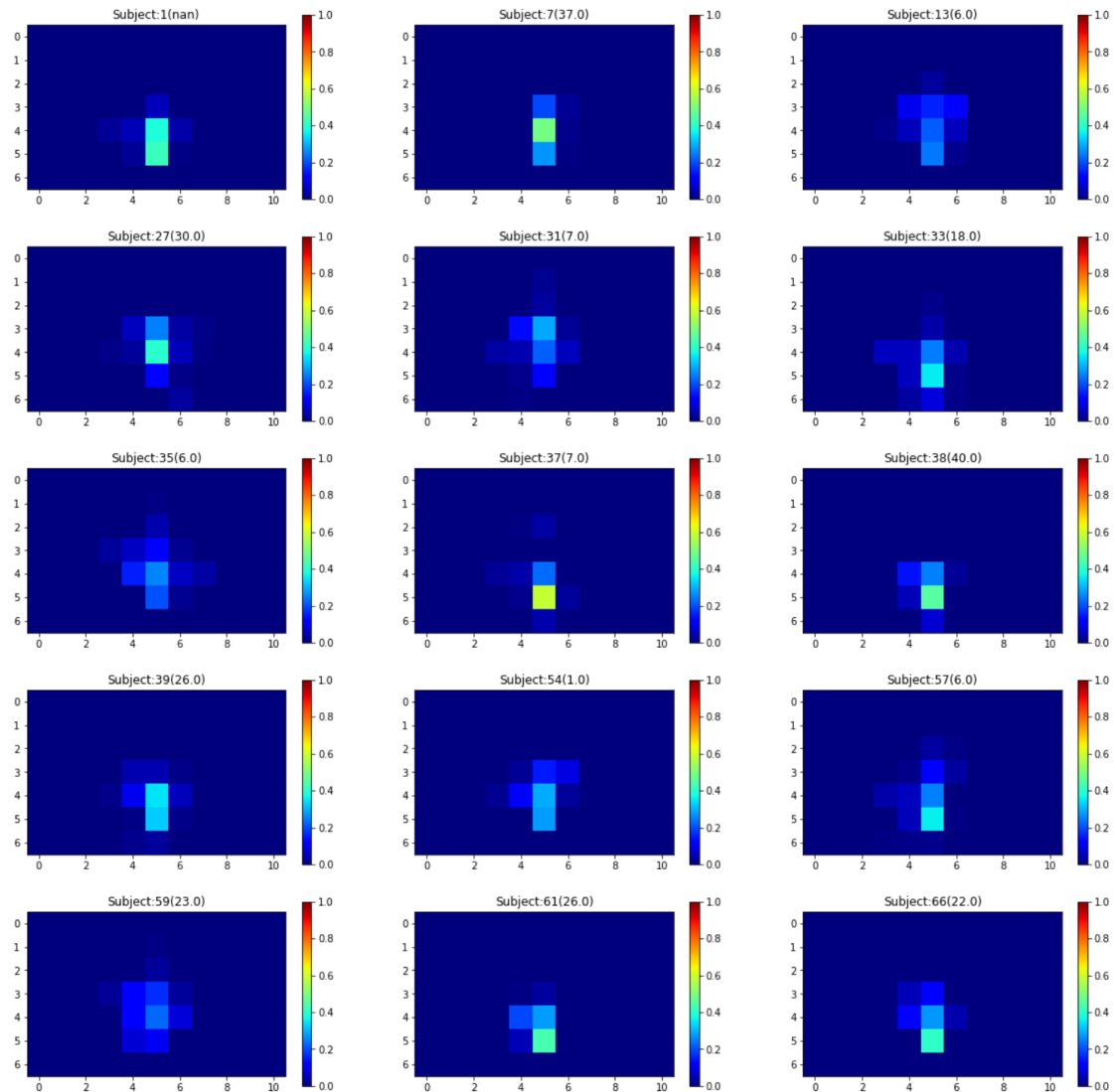


図5 専門家の頻度マップ(Grid=180)

Fig. 5 Frequency Map with Experts

な患者のものであるため、頻度は中央から中央下部に集中する傾向があることが分かる。なお、一般人でも同様のグリッドの頻度が高い被験者も複数おり、頻度の高いグリッドだけでは専門家と一般人を区別できないと考えられる。

4.2. 関心エリア頻度

関心エリア頻度を用いた分析では、視線データにおいて図2、図3で示したAOIにヒットした回数を頻度として数える。AOIが重複してカウントされることはない。関心エリア頻度でも各AOIの頻度は、その顔画像を閲覧した際の頻度値の合計で正規化する。

AOIには順に番号が振られており、左右の同じ部位に対応するAOIには顔画像の右の部位にR、左の部位にLの記号が番号の後ろに付されている。なお、正面顔画像と横顔画像で使われている番号は独立しており、

同じ番号が顔の同じ部位を示すわけではない。

なお、AOIの座標は対象の顔画像ごとに個々の部位に沿って設定するため、顔画像によって異なっているので掲載を省略する。

5. 専門家判別手法

顔画像閲覧時の視線計測データに基づく関心グリッド頻度、関心エリア頻度それぞれの正規化頻度ベクトルを対象に、専門家か一般人か判別することを試みた。

関心グリッド頻度の正規化頻度ベクトル $F_{grid}(t,s)$ は、対象画像を t 、被験者を s 、グリッドサイズに基づくグリッド数を N_{grid} 、 $i(=1 \dots N_{grid})$ 番目のグリッドの被験者 s の視線頻度値を $f_{s,t}(i)$ すると次のように表される。

$$F_{grid}(t,s) = \frac{1}{\sum_{i=1}^{N_{grid}} f_{s,t}(i)} (f_{s,t}(1), \dots, f_{s,t}(N_{grid})) \quad (1)$$

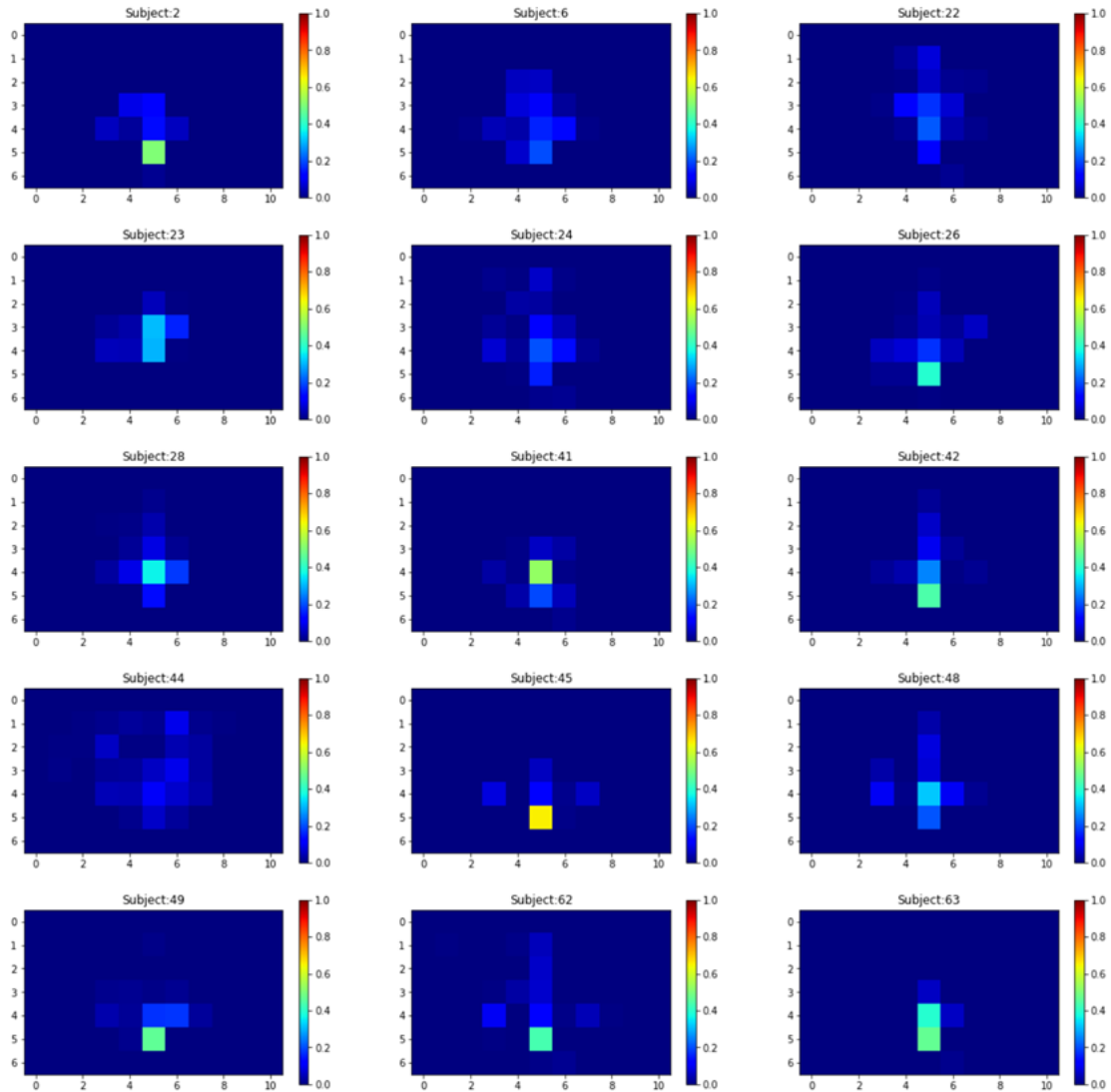


図6 一般人の頻度マップ (Grid=180)

Fig. 6 Frequency Map with Layperson

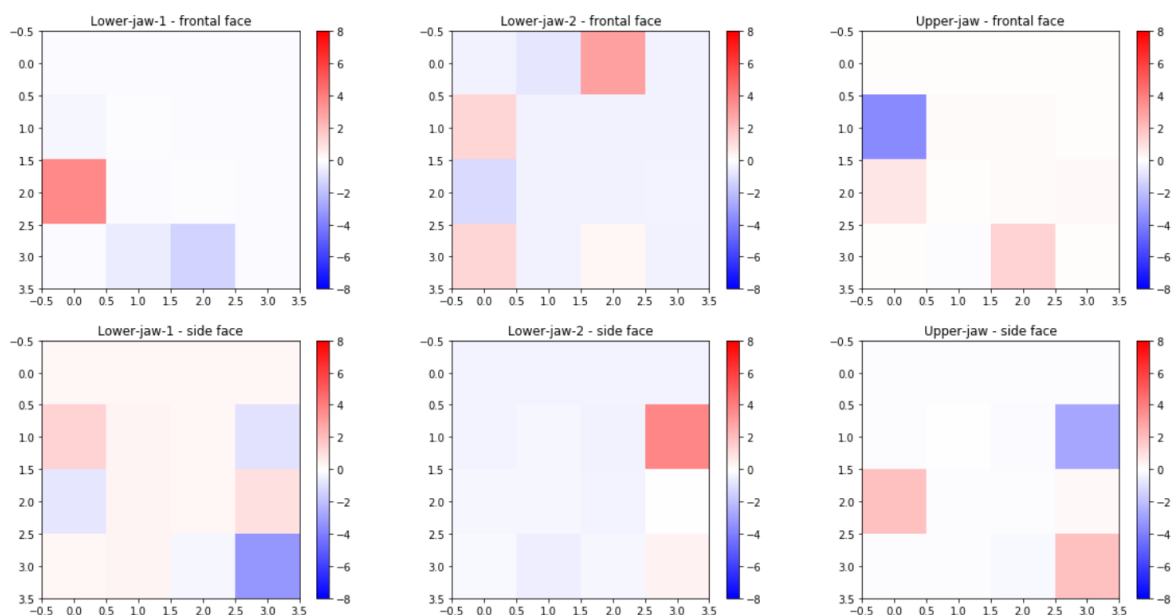


図7 LDAで学習した変換ベクトル W
Fig. 7 Learned transform vector W by LDA

同様に，関心エリア頻度の正規化頻度ベクトル $F_{AOI}(t,s)$ は，対象画像を t ，被験者を s ，AOIに基づくエリア数を N_{AOI} ， $j(=1 \dots N_{AOI})$ 番目のグリッドの被験者 s の視線頻度値を $g_{s,t}(j)$ とすると次のように表される．

$$F_{AOI}(t,s) = \frac{1}{\sum_{j=1}^{N_{AOI}} g_{s,t}(j)} (g_{s,t}(1), \dots, g_{s,t}(N_{AOI})) \quad (2)$$

5. 1. 線形判別分析(LDA)

専門家か一般人かの判別を行うにあたって，複数の機械学習手法を適用して検討した．MDS(多次元構成尺度)やPCA(主成分分析)などいくつかの手法を検討した中で，教師付き学習手法であるLDA(Linear Discriminant Analysis: 線形判別分析)で分類可能性があったと思われるので，LDAによる検討を行った．

LDAは2つのクラスの平均値の差が最も大きくなるようにしながらクラス内の分散は最も小さくなるような射影を見つける手法である．

判別では正解データとして一般人を1，専門家を0として2クラスの分類をおこなった．分析はPython 3.7でscikit-learn 0.21.3を使用した．LDAはscikit-learnのLinearDiscriminantAnalysisを用いて計算を行った．

5.1.1. 関心グリッド頻度のLDA

関心グリッド頻度をLDAで学習した際に，各グリッドに対応する線形変換ベクトル W をヒートマップで出力すると図7のようになる(Grid=500)．色が赤い部分が正の影響が大きく，青いところが負の影響が大きい部分となる．なお，図7は平均値が0になるように標準化している．図の列は対象となる患者，図の上段は正面顔，下段は横顔で，被験者30名分を学習した場合である．

図7を見ると中央部の頻度が判別に与える影響は小さく，周辺部のグリッドの頻度の差異が専門家か一般人かの判別に影響すると考えられる．

なお，精度の評価ではStratified 10 Fold cross validation (層化抽出法を用いた10分割交差検証)を行い，平均値を用いている．症例と顔の向きで6通りあるが，精度を評価するのに被験者30名では少なすぎるため，症例についてはまとめて90名分として処理し，顔の向き(正面・横顔)を分けて精度の計算を行った．精度は専門家を専門家，一般人を一般人と判定する割合である．

関心グリッド頻度では，グリッドサイズによって判別精度が変化する．そこで表1にグリッドサイズ

表1 グリッドサイズ変更による精度変化
Table 1 Accuracy effected by Variety of Grid Size

Face/Grid	100	150	200	250	300	350	400	450	500	550
正面顔	0.503	0.613	0.675	0.660	0.680	0.738	0.620	0.665	<u>0.783</u>	0.673
横顔	0.525	0.613	0.615	<u>0.765</u>	0.638	0.758	0.678	0.65	0.718	0.713
調和平均	0.514	0.613	0.644	0.709	0.658	0.748	0.648	0.657	<u>0.749</u>	0.692

(pixel)を 100 から 550 まで 10 段階変化させたときの正面顔、横顔、調和平均の値を示す。表中の下線は最も高い値を示す。

グリッドサイズを大きくするに従って、精度 (Accuracy)が上昇するが、グリッドサイズが 500 を超えると精度が下がる (表では示していないが 600 以降減少傾向が続く)。精度は専門家を専門家、一般人を一般人として判定する割合である。

精度の調和平均では Grid=500 が最大となるが、Grid=350 の時と値にほとんど差がなく、正面顔・横顔どちらの値も 2 番目に高い。

そこでグリッドサイズが 350 の時と 500 の時の正面顔の混同行列 (Confusion Matrix)を表 2 に示す。その結果、Grid=500 ではほとんどの場合に専門家と判定することで精度が上昇していたが、一般人を判別できていないことが分かった。一方 Grid=350 の場合には専門家を一般人と誤判定する割合が増えるものの、一般人を正しく一般人と判定する割合が Grid=500 より増えた。しかし一般人について見ると、一般人を専門家と判定する割合が大きいことが分かった。なお、横顔も同様の結果であった。

表 2 関心グリッド頻度の混同行列

Table 2 Confusion Matrix for Grid base

(a) Grid=500

	判定：専門家	判定：一般人
専門家	45	0
一般人	37	8

(b) Grid=350

	判定：専門家	判定：一般人
専門家	39	6
一般人	24	21

表 3 関心エリア頻度の精度

Table 3 Accuracy for AOI base

正面顔	0.755
横顔	0.738
調和平均	0.746

表 4 関心エリア頻度の混同行列

Table 4 Confusion Matrix of AOI

(a) 正面

	判定：専門家	判定：一般人
専門家	33	12
一般人	10	35

(b) 横顔

	判定：専門家	判定：一般人
専門家	33	12
一般人	11	34

5.1.2. 関心エリア頻度の LDA

関心エリア頻度についても同様に LDA を用いて学習し変換ベクトルを計算した。3 症例をまとめて 90 名分のデータとして精度を計算したものを表 3 に示す。精度の調和平均では関心グリッド頻度の値よりも低いものの、正面顔・横顔それぞれの精度は Grid=350 の時と同様の高い値を示している。

そこで混同行列を確認したものが表 4 である。専門家を一般人と誤判定する率は増えているものの、一般人を一般人として正しく判定する率は増えており、専門家・一般人どちらについても同じ程度の正答率となっている。

6. 考察

関心グリッド頻度と関心エリア頻度という視線データの集計の仕方、専門家か一般人かどうかの判別の性能が変わるかどうかに着目すると、関心グリッド頻度を用いた方が精度だけで言えば高いが、一般人を専門家と誤判定する率が高い。関心エリア頻度の場合には一定の割合で専門家か一般人かを判定できることが分かった。

関心グリッド頻度は事前に専門家にエリアを設定してもらうなどのコストは必要なく、グリッドサイズなどの調整で性能が変化するが、スクリーン座標を等間隔に分割するだけであるため顔の部位と関係がない。一方、関心エリア頻度の場合には専門家が各画像を見て部位ごとに設定するため事前準備のコストは高いが、重要なエリアかそうでないエリアかの差が出ることで安定した判別につながっているのではないかと考えられる。

なお、LDA を用いた分析では 3 症例分をまとめて 90 名で学習と判別を行ったが、その結果を専門家・一般人 15 名ずつとして、同一人物について 3 症例のうち 2 例以上について専門家とされた場合には専門家、専門家とされたのが 1 例以下の場合には一般人として判定した際の混同行列を表 5 に示す。

30 名中 25 名について正解となるので精度としては 8 割程度の結果となった。正面と横顔を合わせて判定した場合でもほぼ同様の結果となった。

7. まとめ

熟達度を判定するため、まず専門家か一般人が判定できるか分析方法の検討をおこなった。実験で治療が必要な患者の顔画像 (正面、横) の 3 症例 (計 6 画像) を閲覧した際の視線の動きを計測したデータを集め、これを分析した。

関心グリッド頻度と関心エリア頻度それぞれについて LDA の判別分析を行った結果、専門医が部位別に設定した AOI (関心エリア) に基づく視線の頻度を用いることで専門医か一般人かを比較的よく判別できるこ

表 5 専門家判定の混同行列

Table 4 Confusion Matrix of Analysis

(a) 正面

	判定：専門家	判定：一般人
専門家	14	1
一般人	4	11

(b) 横顔

	判定：専門家	判定：一般人
専門家	11	4
一般人	1	14

とが分かった。

今後の課題として、専門医については経験年数などもデータとしてあるが熟達度としてその経験年数を推定するまでには至っていない。また AOI の設定は専門医による判断が必要であるが顔認識により部位などを判定することにより自動的に AOI が設定できるかなど検討が必要である。

参 考 文 献

- [1] Yarbus, A. L.: Eye movements and vision, New York: Plenum Press (1967).
- [2] 鎌原淳三, 長松隆他, “ウェブとデータベース” 視線を利用した技能継承支援システムにおけるマルチメディア拡張,” 電子情報通信学会技術研究報告・PRMU, パターン認識・メディア理解 107(115), 19-24, 2007-06-21
- [3] Junzo Kamahara, Takashi Nagamatsu, et al., “Instructional Video Content Employing User Behavior Analysis: Time Dependent Annotation with Levels of Detail,” Proc. of 18th International Conference for User Modeling, Adaptation, and Personalization, (UMAP 2010), pp.87-98, Big Island, USA, June 20-24, 2010.
- [4] 原田紘人, 中山 実, “プログラムコード読解時の視線特徴と理解度に関する検討,” 信学技報, vol. 119, no. 39, HIP2019-4, pp. 19-22, 2019 年 5.
- [5] 伊藤京子, 佐尾明輝, 廣江葵, 長松隆, 青柳西藏, 永田 順子, 高田健治: 熟達した診断プロセスの形式化に向けた視線計測実験, ヒューマンインタフェース学会研究報告集, Vol. 21, No.1, pp.11-16, 2019.
- [6] Kyoko Ito, Haruki Sao, Takashi Nagamatsu, Junko Nagata and Kenji Takada : Formulation of Diagnostic Expertise in Oral Health Care and Its Application to Clinical Education, Proceedings of the 19th International Conference on Human-Computer Interaction, pp. 532-539, 2017.
- [7] トビー・テクノロジー株式会社, Tobii Pro X2, <https://www.tobiipro.com/ja/product-listing/tobii-pro-x2/> (閲覧日 2020 年 1 月 9 日)
- [8] Kenneth Holmqvist, et al., “Eye Tracking: A Comprehensive Guide to Methods and Measures,” Oxford University Press (2011).