

# 視線計測と機械学習を用いた不特定多数の屋内歩行者による デジタルサイネージへの注視傾向推定手法

高井峻\*・沖拓弥\*\*・福島佳浩\*\*\*・羽田優太\*\*\*\*・伊山潤\*\*\*

## Estimating the Gaze Tendency on a Digital Signage for Unspecified Number of Pedestrians inside a Building Using Gaze Measurement and Machine Learning.

Ryo TAKAI, Takuya OKI, Yoshihiro FUKUSHIMA, Yuta HANEDA, Jun IYAMA

**Abstract:** Nowadays, eye gaze measurement devices have become relatively easy to obtain, and analysis of people's gaze behavior characteristics and gazing tendencies has been attempted in a variety of situations. However, it is not realistic to acquire data from an unspecified number of pedestrians wearing eye tracking devices. In this paper, we develop a method for estimating the gazing tendency of an unspecified number of indoor pedestrians from videos captured by a single web camera installed on top of a digital signage system, utilizing machine learning and the results of a preliminary eye movement measurement experiment using a limited number of subjects.

**Keywords:** 歩行者 (pedestrian), 機械学習 (machine learning), 視線計測 (gaze measurement), 情報提示 (information presentation), 画像認識 (image recognition)

### 1. はじめに

#### 1.1. 研究の背景

昨今、屋内外を問わず様々な場所でデジタルサイネージでの広告物や掲示物を目にする。情報提供者にとって、提示した情報のどの部分にどの程度、人々が関心を持つかを把握することは、情報提示効果を高めるために重要である。

近年、視線計測技術の発展とともに視線計測機器（アイトラッカー）が入手しやすくなり、比較的手軽に人々の視線行動特性や注視傾向を把握できるようになりつつある。しかし、不特定多数の歩行者を対象とする場合、視線計測機器を着用してデータを取得することは現実的でない。

#### 1.2. 既往研究

Web 広告の提示内容の改善や地下鉄構内での視覚行動特性の把握（横山ら，2021）など、視線計測から被験者の視点座標を把握し、限られた被

験者のもとで情報提示方法の検討や情報提示効果の検証を行った研究事例が増えている。また、Sugano ら（2016）は、パブリックディスプレイの上部に設置した Web カメラで被験者の様子を撮影し、顔の検出と顔の特徴量抽出を行うことで、視線計測装置を使用せずに被験者のディスプレイ内注視領域の推定を試みている。しかし、ディスプレイから被験者までの距離が離れていて、顔の特徴量がうまく抽出できない状況下では有効ではない。

#### 1.3. 研究の目的

以上の背景から、本研究では、限られた人数の被験者による視線計測実験結果とその際の被験者の様子を撮影した動画の 2 種類を活用し、機械学習により、不特定多数の屋内歩行者の注視傾向を推定する手法を開発する。

---

\* 学生会員 東京工業大学環境・社会理工学院 (Tokyo Institute of Technology)

〒152-8550 東京都目黒区大岡山 2-12-1 E-mail: takai.r.ac@m.titech.ac.jp

\*\* 正会員 東京工業大学環境・社会理工学院 (Tokyo Institute of Technology)

\*\*\* 正会員 東京大学大学院工学系研究科 (The University of Tokyo)

\*\*\*\* 学生会員 東京工業大学環境・社会理工学院 (Tokyo Institute of Technology)

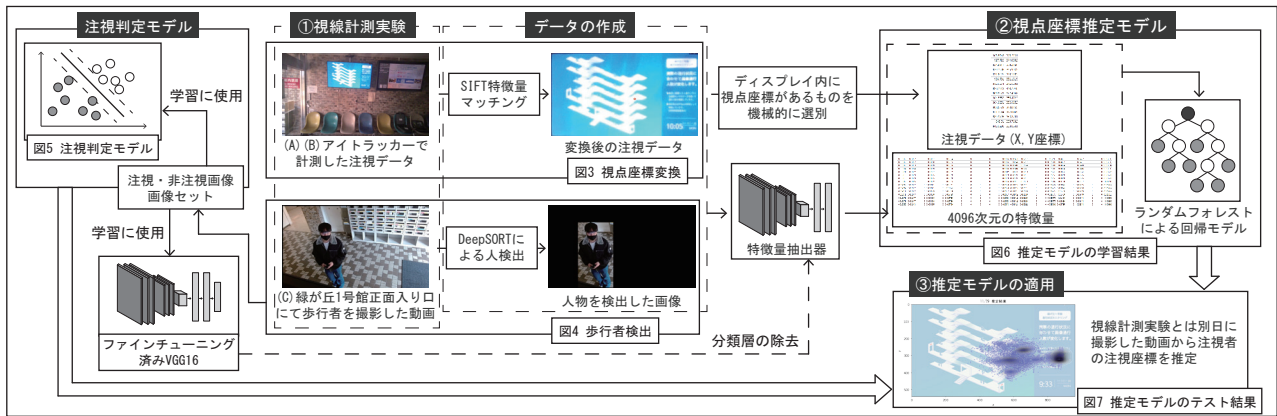


図1 分析の流れ

## 2. 提案する注視傾向推定手法の概要

### 2.1. 分析の流れ

分析の流れを図1に示す。最初に、Webカメラで撮影した動画のみから対象物への注視・非注視のみを判定するモデル（注視判定モデル）を構築する。次に、注視時点について、より詳細に対象物のどの点を注視しているかを推定する「視点座標推定モデル」を構築する。具体的には、まず、少人数の被験者の協力の下で、視線計測実験を行い、被験者の視点座標と被験者の視線を撮影した動画、および、その際の様子を撮影した動画を取得・処理する。そして、変換した視点座標とWebカメラの動画から、視点座標を推定する回帰モデルが視点座標推定モデルである。さらに応用として、別日に撮影した歩行者の動画のみから、注視判定モデルと視点座標推定モデルを使用し、その歩行者の注視傾向を推定を試みる。

### 2.2 歩行者検出とマスク処理の方法

Webカメラで撮影した動画データの処理方法を図2で示す。動画から被験者のみを抽出するために、物体検出および追跡アルゴリズムの一つであるDeepSORT (Wojkeら, 2017) を用いて人物を検出する。そして、バウンディングボックスで囲まれた人物の範囲にマスク処理を施し、その他の部分（背景）を黒とした画像（図2-②）を生成する。

### 2.3. 注視判定モデル

2021年11月26日に撮影したテストデータの動画から図2の処理を行い、画像内の人物のバウ



①Webカメラで撮影した動画のキャプチャ画像例。  
②TrackingアルゴリズムであるDeepSORTを使用し、人物を検出する。その際に検出された人物以外の背景となる部分を黒に塗りつぶす。  
(※ここで人物が抽出されたフレームの画像を用いて、通行人数を算出する。)

図2 Webカメラ画像からの歩行者検出とマスク処理の方法

表1 視線計測実験の概要

実験日	2022年3月10日, 11日
実験場所	東京工業大学大岡山キャンパス緑が丘1号館入口付近
被験者数	計4名(20~30歳の男女)
方法	<p>実験は被験者1名ずつ行う。</p> <p>(1) 視線計測装置(※1)を着用し、キャリブレーションを行う。</p> <p>(2) 以下の①~⑤の手順で被験者に歩行してもらい、その際の時刻、視線座標、瞳孔径、被験者から見た映像を視線計測装置で記録するとともに、ディスプレイ(※2)側のWebカメラ(※3)1台で被験者の動きを録画する。</p> <p>①ディスプレイ前の通路を等間隔の3列に分ける。ディスプレイに遠い列のエレベータ付近から開始。ただし、被験者がディスプレイに表示されている文字が読める場所から開始するように指示をする。また、実験中は顔だけでなく身体もディスプレイに向けて、注視するように指示をする。</p> <p>②立ち止まって画面上に表示している画像(※4)の文字や図を目で追う。その際に表示している画像内の指定した場所である(1)から(6)までを各々約10秒間注視する。→50cmほど横に移動→再度約10秒間画面上の文字や図を目で追う→50cmほど横に移動→を繰り返す。</p> <p>③入口ドアを通過したら同じ列で転回し、先ほどとは逆方向に同じ動作をエレベータ前に到達するまで行う。</p> <p>④再度転回後、ディスプレイに近い側の列に移動し、②~③の流れを繰り返す。</p> <p>⑤文字が読めるぐらいの歩行速度で歩いてもらい、注視をする。その際、被験者に伝えた注視箇所(画像内の(1)から(6)のいずれか1つ)を歩行中に注視する。注視箇所を変えて、6回歩行中の注視をしてもらう。</p> <p>※1) Tobii Pro Glasses 2を使用。視線計測間隔は50Hz。          ※2) 55型ワイド液晶ディスプレイ(パナソニック製TH-55EQ1J)。          ※3) G920n HD PRO WEBCAM (Logicoool製、フルHD)。          ※4) 羽田ら(2021)の研究結果を表示画面として使用。</p>

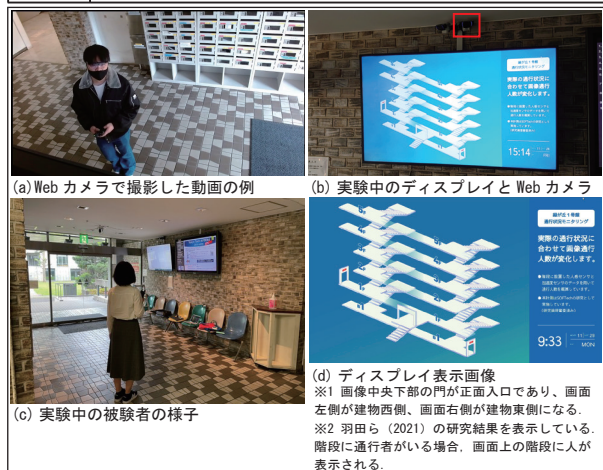


図3 視線計測実験の様子

ンディングボックスのみを切り取った画像を用いる。その画像から、注視・非注視に分類した画像セット（注視：7,966枚、非注視：29,237枚）を作成し、その画像セットを使い、機械学習アルゴリズムであるサポートベクターマシンによる学習を行い、注視判定モデルを作成する。

#### 2.4. 視線計測実験

東京工業大学大岡山キャンパス緑が丘1号館<sup>1)</sup>正面入口横に設置されているディスプレイ（入口ドア側）付近で視線計測実験を行う。トビー・テクノロジー社のウェアラブルアイトラッカーである Tobii Pro Glasses 2（以下、アイトラッカー）を着用した被験者4人に対して、個別にディスプレイ前を往復しつつ、ディスプレイの表示内容の指定箇所を見てもらう実験を行った（表1）。その結果、アイトラッカーで計測した被験者の視点座標（図1（A））とアイトラッカーに搭載されたカメラで撮影した視線動画（図1（B））、およびディスプレイ上部に設置したWebカメラから撮影した動画（図1（C））を取得した。

#### 2.5. 視線計測実験で取得したデータの処理

視点座標データ（図1（A））と視線動画データ（図1（B））の処理方法を図4に示す。被験者の身体や頭部の移動に伴い、被験者が見ているシーンは変化する。そのため、異なるフレーム間で同じ視点座標であっても、同じ領域を見ているとは限らない。そこで、（B）の動画フレームからディスプレイの四隅の座標を取得し、SIFT特徴点マッチングにより、次フレームのディスプレイ領域を抽出する処理を逐次的に行うことでディスプレイ領域を追跡しフレーム間で比較できるようにした。抽出したディスプレイ領域を実際のディスプレイの縦横比（16:9）へと変形するために、射影変換を用いて動画フレームを変形し、その際に求めた変換行列を視点座標に適用する。その結果、ディスプレイの範囲を基準とした視点座標に変換し、フレーム間でディスプレイ上のどこを歩行者が見ているかを把握できる。

#### 2.6. 視点座標推定モデル

マスク処理済み画像のうち、ディスプレイ内に

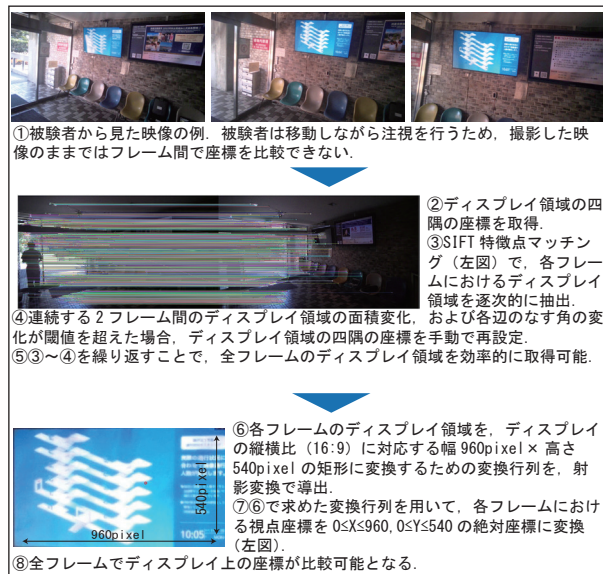


図4 被験者の視点座標の変換方法

視点座標がある画像を対象に、特徴量抽出器を通して得られる各画像の特徴ベクトルから、図3の手順で変換済みの視点X・Y座標を推定するランダムフォレスト回帰モデルを学習させる。ここでは、2.3節の注視・非注視画像セットでファインチューニングしたVGG16（Simonyan and Zisserman, 2015）の学習モデルを特徴量抽出器として用いる。具体的には、マスク処理を施した画像を、このモデルの全結合層（FC1）で出力される4,096次元のベクトルを、各画像の特徴ベクトルと定義する。

### 3. 注視傾向推定手法の精度検証

#### 3.1. 注視判定モデルの精度

注視判定モデルの学習精度は98.6%と良好であった（図5（a））。ここで、2021年11月29日に図3（a）と同じ仕様で撮影した動画データから、人物が検出されたフレーム画像をランダムに500枚抽出し、注視・非注視を目視で判断した結果と、同じ500枚を注視判定モデルで分類した結果を図5（b）に示す。Accuracy（正解率）、Precision（適合率）、Recall（再現率）、F値はそれぞれ0.942, 0.519, 0.871, 0.651であった。この結果は、やや過大となる可能性を示している。実験で使用したディスプレイの隣に設置している別のディスプレイを見ている場合や、ディスプレ

(a) 注視判定モデルの学習時の精度		(b) 11/29 データでの混同行列とモデル評価指標				
	精度	目視での判別		モデルでの判別		精度
				注視	非注視	
訓練データ	0.9970	注視	27	4	精度	0.942
検証データ	0.9856	非注視	25	444	適合率	0.519
					再現率	0.871
					F 値	0.651

図5 注視判定モデル

イ前に集まって会話をしている場合に「非注視」であるにもかかわらず、「注視」の判定がされている。実験環境の特性や判定モデルの学習の際に想定していなかった状況の教師データの不足が注視の画像を分類する上での課題と考えられる。

### 3.2. 視点座標推定モデルの精度

本稿では、2.5 節の手順で変換済みの視点座標が、ディスプレイ領域内にある時点のみを学習データとし、それを訓練データ：検証データ = 75% : 25% となるように分割した。その結果、検証データの精度は 92.5% であり、Y 座標の検証データの精度は 61.5% であった。

## 4. 注視傾向推定手法の応用

### 4.1. 使用するデータ

視点座標推定モデルに、2021 年 11 月 26 日および 11 月 29 日から 12 月 3 日の計 6 日間に、第 2 章と同条件で撮影した動画データを入力し、それをもとに屋内歩行者の画面注視状況を推定することを試みた。ここでは、学習済みの注視判定モデルで、「注視」と判定された画像を用いた（ただし、11 月 26 日のみ、比較のため目視で「注視」と判断した画像セットも使用した）。

推定結果を図 7 に示す。いずれの計測日でも、ディスプレイ領域の中央部にある建物東階段付近、および右中央部に記載された実験の説明文に視線が集中している。当該建物では、東側の階段が頻繁に使われており<sup>1)</sup>、画面表示の変化が大きいことから、対応する部分に視線が集中したものと推察される。

## 5. まとめ

本研究では、少人数の注視データと注視者の様子を Web カメラで撮影した動画データから視点座標推定モデルを作成し、不特定多数の歩行者の

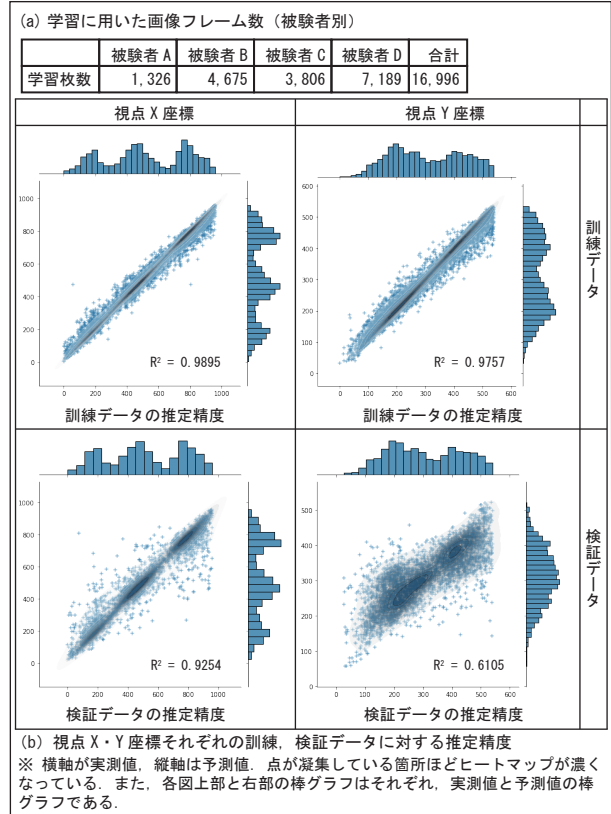


図6 視点座標推定モデルの精度

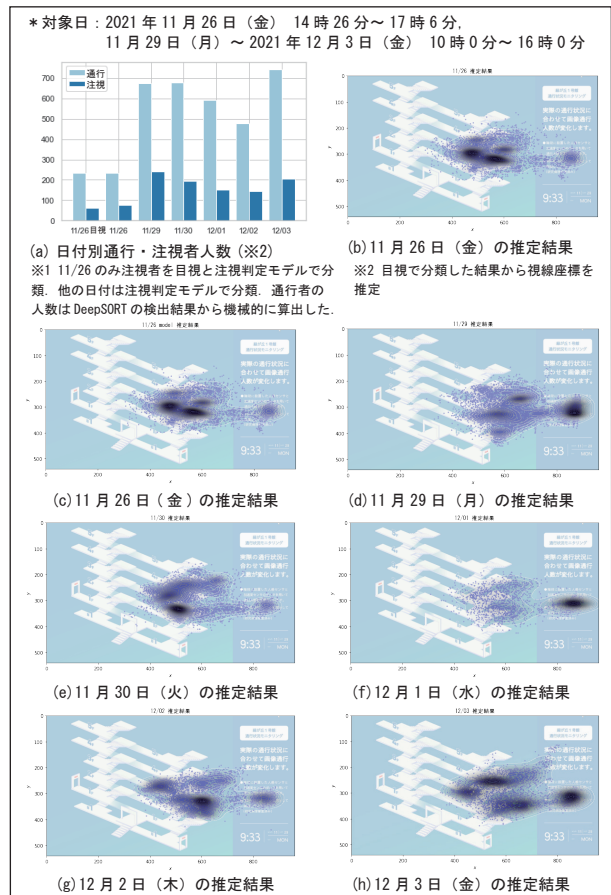


図7 屋内歩行者による画面注視状況の推定結果

注視傾向を推定する手法を開発した。

今後は、推定精度向上に向けた各モデルの改善、ディスプレイ表示内容の変化に伴う注視傾向の関係分析、および本稿とは異なる状況、場所における注視傾向の推定手法への拡張などを予定している。

## 謝辞

本研究は、JST 産学共創プラットフォーム共同研究推進プログラム（JPMJOP1723，領域統括：吉敷祥一）の一環で行った研究の一部である。共同研究者である東京工業大学の大佛俊泰教授と永岑光恵准教授からは、本研究を進めるにあたり多くのご意見を頂いた。関係者各位ならびに視線計測調査、動画撮影にご協力頂いた方々へ謝意を表す。

## 注

- 1) 実験を実施した東京工業大学大岡山キャンパス緑が丘1号館は、正面入口が主な出入口であり、建物西側の階段よりも東側の階段やエレベータが頻繁に使用されている。
- 2) 本研究は、東京工業大学「人を対象とする倫理審査委員会」の承認の下で実施した（承認番号：第2019204）。

## 参考文献

- トビー・テクノロジー株式会社（2022）消費者行動・ユーザービリティテストの事例。（2022年8月20日参照）<<https://www.tobiiipro.com/ja/applications/marketing-user-research/case-studies/>>
- 羽田優太，大佛俊泰，田頭まき，伊山潤，福島佳浩（2021）加速度センサを用いた階段利用者人数及び歩行方向の推定，地理情報システム学会講演論文集，Vol.30
- 横山広充，西應浩司（2021）大型デジタルサイネージ視認時の視覚行動特性に関する研究，日本デザイン学会研究発表大会概要集，68巻，日本デザイン学会 第68回春季研究発表大会，

セッション ID P1-24，344-345.

- D.G.Lowe（2004）Distinctive Image Features from Scale-invariant keypoints, Journal of Computer Vision, 60(2), 91-110
- K.Simonyan and A.Zisserman（2015）Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv:1409.1556
- Y.Sugano, X.Zhang, and A.Bulling（2016）AggreGaze:Collective Estimation of Audience Attention on Public Displays. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16). Association for Computing Machinery, New York, NY, USA, 821–831.
- N.Wojke, A.Bewley and D.Paulus（2017）Simple online and realtime tracking with a deep association metric, 2017 IEEE International Conference on Image Processing (ICIP), 3645-3649.