

# 画像解析による観光者行動の分析の試み

—イル未来と 2023 イベント会場での基礎実験—

An Attempt to Analyze Tourist Behavior by Image Analysis.

-Fundamental experiments at the “Illumilight 2023” site-

神谷 達夫

Tatsuo Kamitani

## 要旨

著者らは、観光イベントの運営とマーケティングのために観光者の行動分析を研究してきた。焦電センサーの問題を解決するために、画像認識を用いることを検討したが、このためのセンサーを用いることにより、観光者の動きの検出ができるのではないかと考えた。

この可能性を確認するため、京都府福知山市で開催された”イル未来と 2023”イベントにおいて実証実験を実施した。実験では、HOG と SVM を用いた方法と深層学習による方法により人が居るかどうかの判定を試みた。その結果、深層学習による画像識別による方法が容易に実現できた。

キーワード: 行動分析、観光者、画像処理、HOG, SVM, 深層学習

Keywords: behavior analysis, sightseer, image processing, HOG, SVM, deep learning

## 1. はじめに

著者らは、観光イベント等の企画立案のために、観光者の行動パターンを理解することは、効果的な観光地の運営やマーケティング戦略の策定に有用であると考え、これまで WiFi パケットセンサー<sup>(1)</sup>やソーシャルディスタンス手持ち行燈(ディスタンス行燈)<sup>(2)</sup>による観光者の行動分析に取り組んできた<sup>(2,3)</sup>。

一方、観光者を楽しませるために観光者が特定の場所に立てば、動作の変化するような展示物を展示してきた。例えば、“イル未来と 2022”で展示した人が接近すると 3D CG を再生させるような内容である(図 1)。この展示では、焦電センサーを用いて人の検出を試みた。しかし、焦電センサーの検

知範囲は広く、人が存在するかどうかの概略を検出できるのみで、誤検出が多発した。このため、この時の展示では人による手動操作が必要になった。

この焦電センサーの問題を解決するためのセンサーを検討した結果、画像を用いる方法が考えられた。画像を用いると、対象に接触することが無く検出でき、センサー用のハードウェアも比較的安価に実現できると考えられた。さらに、人の検出が用意であるのなら、展示物の制御のみでなく、イベント会場の状況把握にも使用できるのではないかと考え、システムの開発を試みた。

この研究・開発により、イルミネーション会場における観光者の動向をより正確に把握することが可能になると考えられ、実現すれば観光地の運営、安全対策、および顧客体験の向上に寄与することが期待される。



図 1 焦電センサー検出と組み合わせた 3D CG の展示

焦電センサーでは誤検出が発生し、手動で操作することになった。

## 2. 画像による観光者行動の検出

### 2.1 イル未来と 2023 イベント会場での実験

京都府福知山市で開催されるイベントであるイル未来と 2023 会場にてカメラを用いて照明や展示

物の制御が可能か確認するため、実験によって確認した。実験は、2023 年 10 月 28 日, 10 月 29 日, 11 月 3 日, 11 月 4 日, 11 月 5 日, 11 月 11 日, 11 月 12 日の 7 日間であった。

## 2.2 カメラの配置

本研究では、会場の観光者の動きを捉えるために、前節で述べたように、合計 4 台のネットワークカメラを使用した。これらのカメラを会場の主要な入口や通路、および集客ポイントに設置できることが望ましいのであるが、カメラということで設置が許可されなかったため、学生の展示の周辺のみで実験することになった。したがって、学生展示の全体を各所から撮影するカメラと、展示場所に入りする人々を撮影するカメラを設置することにした(図 2)。

学生の展示物は、”竹灯籠”と呼ばれる竹に発光ダイオードを入れた電飾である。この竹の中の LED は、コンピュータコントロールによって色が変わる。図 2 の会場部分に竹灯籠が配置されている。図 2 の番号がカメラを示しており、1 番のカメラが受付の前に向けられている。その他の 2 から 4 番のカメラは、会場の方向に向けられている。



図 2 会場とカメラ配置

地理院地図に受付とカメラ配置を追加

## 2.3 カメラの選定

使用したカメラは、通常監視カメラ用として使用されるネットワークカメラと呼ばれる LAN インターフェースでデータを転送するタイプのカメラである。このカメラを選定した理由は主にコストの

面での利点にある。ネットワークカメラは 1 台 8,000 円程度と比較的安価である。また、LAN を用いているため、LAN プロトコルを用いて画像を送ることになり、キャプチャするためのプログラムが不要となりプログラミングが容易である。

また、これらのカメラは Power over Ethernet (PoE) に対応しており、電源供給のために別途の配線が不要である。これにより、設置工事が簡単になる上、設置の柔軟性が高まる。

## 2.4 カメラからのデータ取得

カメラがネットワークカメラであるため、RTSP (Real Time Streaming Protocol)によるデータ伝送に対応している。したがって、今回は、RTSP プロトコルを用いて静止画データを取得した。

画像データ取得用のソフトウェアは、Python で記述された。Python を用いて 1 台のカメラからの画像を定期的に取得できるようにし、そのソフトウェアを 4 プロセス動作させ、4 台のカメラからデータを取得した。データ取得に用いたコンピュータは、Raspberry Pi 4 である。



図 3 使用したネットワークカメラ

H.View 社製 HV-500G2

## 2.5 画像処理に使用する技術

本研究では、イベント会場における観光者の行動を分析するために、HOG と SVM を用いる方法<sup>(4)</sup>と深層学習を用いる方法の 2 つの異なるアプローチを検討した。これらの方法は、それぞれ異なる特徴抽出手法と分類アルゴリズムであり、それぞれに利点がある。

HOG と SVM を用いる方法は、リアルタイム処理やリソースが限られた環境に適しており、比較的小規模なコンピュータで実現可能である。一方、深層学習を用いる方法は、複雑な画像特徴の学習と高い精度が求められる場合に優れた選択となりうるが、モデル作成のために GPGPU のようなハードウェアを使用する必要があり、小型の機器では実現が容易ではない。

本研究では、これらのモデルの強みと弱みを比較し、特定の要件に最適なモデルの選択基準を検討する。

## 2.6 HOG+SVM 法

HOG と SVM を用いる方法では、画像からの特徴抽出に HOG (Histogram of Oriented Gradients) を利用し、これらの特徴を基に SVM (Support Vector Machine) で分類する。この方法の実現には、scikit-learn と OpenCV が用いられた。

scikit-learn ライブラリは、様々な分類、回帰、クラスタリングアルゴリズムを持っており、SVM の実装において高い柔軟性と利便性がある。また、OpenCV ライブラリはリアルタイム画像処理に適しており、HOG 特徴抽出などの機能を効率的に実行できる。

## 2.7 深層学習法

深層学習法では、CNN (Convolutional Neural Network) を主要なアーキテクチャとして採用した。この CNN モデルの構築には TensorFlow とその高レベル API である Keras が使用された。TensorFlow は大規模な数値計算に適したオープンソースのライブラリで、深層学習モデルの開発に広く使用されている。

Keras は TensorFlow の高レベル API として使用でき、モデルの設計と実験を迅速に行うためのインターフェースを提供する。

# 3. 実証実験の結果

## 3.1 HOG と SVM による人物認識

OpenCV には、事前に訓練された HOG (Histogram of Oriented Gradients) ベースの人物検出器が含まれている。それを用いて SVM (Support Vector Machine) と組み合わせて人間の存在の検出を試みた。その結果、人の居ない場合の画像でも人の形のようなものがあつた場合人と認識するため、背景のみの画像でも何人かの人間が存在すると識別された(図 4)。図 4 は、その一例であるが、画像中の矩形で囲まれた部分に人間が存在すると識別された。この画像では 9 人の人間が存在するという判定であつた。

このような結果になったため、HOG と SVM を組み合わせた人の認識はできなかった。今回は、scikit-learn ライブラリの中のモデルをそのまま使用しているため、モデル等のチューニングはして



いない。

モデル等のチューニングをすると、使用できる可能性もあるが、チューニングをするための手間を考えると次節で述べる深層学習型の認識の方が手軽であると考え、今回はこれ以上のチューニングを実施しなかった。

今回は使用できる状態にはできなかったが、HOG+SVMによる人物の識別にかかる計算機資源の量は小さく、どうしても能力の高いコンピュータを用いることができない場合は、この方法の使用も考えられる。



図4 scikit-learn の HOG と SVM を用いた人物認識

9 人の人間を誤検出している。矩形で囲まれた部分に人が居ると識別している。

### 3.2 深層学習による人物認識

TensorFlow と Keras を用いて判定モデルを作成した。実験によって取得した画像は大量にあるが、学習データとするためには手作業による分類が必要であるため、現時点では部分的な画像データによるモデル作成しかできていない。

実証実験のトレーニング用に使用した画像は人の居ない時の画像が 99 枚、人の居る場合の画像が 87 枚の合計 186 枚の画像である。これに検証用に、人の居る場合が 6 枚、人の居ない画像が 6 枚の

合計 12 枚を用いた。

モデル作成の結果、このような多くない学習量によっても、十分な精度で人の居る居ないを識別できるモデルが作成できた。モデルのパラメータ等は、ChatGPT にて出力した画像判定用のプログラムそのままである。この条件で、概ね 30 弱ぐらいの epochs でモデルが完成している(図 5)。

プログラムを作るためには、従来は Keras などを用いる知識が必要であった。しかし、今回の実証実験では ChatGPT によって生成したプログラムを基にプログラムを作成しただけで十分使用可能なモデルが作成可能であった。このことは、計算機能力が十分高ければ、人間がほとんど考えない状況で識別プログラムを作ることが可能であることを示唆している。したがって、計算機資源が潤沢であれば、他の方法よりも深層学習による画像認識を用いる方が適しているといえる。



図 5 モデル作成時の学習の進行

## 4. まとめ

本稿では、観光イベント会場において画像解析による観光者行動の分析の可能性を実験により確認した。実験では人の識別に HOG と SVM を用いた方法と深層学習による方法の 2 種類を試みた。

実験の結果、チューニングをしない場合の HOG+SVM 法は、人の居ない状態でも人が居るように識別してしまい、使用に耐えなかった。一方、深層学習を用いた方法では、モデルやプログラムのチューニングをすることなく、ChatGPT のような AI の出力したプログラムをほぼそのまま利用した状態でも使用に足る識別精度を得ることができた。

この結果から、計算機資源が十分であれば、深層学習による方法を用いることが良いということが分かる。

今回の実証実験では、技術的には容易に画像による人物識別が可能であるということが確認できたが、カメラの設置が許可されにくいということが分かった。照明や展示物制御のためのセンサー目的であっても、カメラの設置の許可を得ることは難しい。カメラを用いる方法は、非接触で状況を把握する方法としては優れていると考えられるので、今後はどのような形態にすればカメラやカメラ様のセンサーの設置が許可されやすいのか検討する必要がある。

## 謝辞

データ取得のための機会を提供していただいた”イル未来と”学生プロジェクトのメンバー及び、データ取得に協力いただいた前田尋翔氏の貢献に深く感謝する。

## 《参考文献》

- (1) 神谷 達夫, 位置情報データを活用した観光地指標：海の京都観光圏 Wi-Fi パケットセンサーの情報量解析から, 日本観光学会誌, No. 59, pp.41-48 (2018)
- (2) 神谷 達夫, コロナ禍における観光者の行動分析に関する試み—イル未来と 2021 イベントにおける電子行燈の運用—, 福知山公立大学研究紀要, Vol.6, No. 1, p.63-76 (2022)
- (3) 神谷 達夫, 自動取得された移動軌跡による観光者行動の分析—コロナ禍に対応した機器の実証実験—, 日本観光学会誌, Vol.23, No. 1 (条件付き掲載) (2023)
- (4) Monocular Pedestrian Detection: Survey and Experiments, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(12) 2009