

# ウェアラブルカメラを利用したものの探しの検討

下田 寛子 (指導教員: 椎尾 一郎)

## 1 はじめに

TrackR, Inc. の探しものに関する調査<sup>1</sup>によると、人は年間で約 145 時間をもの探しに費やしており、失くしたものの総額は年間約 1.7 兆円にも上るといふ。もの探しは日常において頻繁に行われる行為であり、ものを失くすことによる損失は大きい。

もの探しをする際、人は視覚を利用して失くしたものを探す。目を動かさずに見える範囲とされる視野は、水平方向では約 100 度あるという。対して、焦点が合い認識できる範囲は約 20 度と狭く、これを有効視野と呼ぶ。

一方、コンピュータビジョンにおいては画像全体が処理の対象となる。一般的な web カメラの画角は 70 ~ 80 度程度、広角レンズを搭載し 120 度以上の画角をもつものや、2 枚のレンズを用いて 360 度撮影できる機器も流通している。特定の範囲を注視し認識する人と、画像全体に対して処理を行う計算機では画像の認識方法が異なると言える。加えて、人の有効視野よりも広角なレンズを搭載した web カメラで撮影した画像から探索することで、人よりも広範囲を探ることが可能となり、もの探しの効率向上が期待できると考えた。

そこで本研究では、人と計算機が同時にもの探しを行い、コンピュータビジョンによって物体を見つけるまでの時間を人と比較することにより、計算機のもの探し支援が有効であるかを検証する。

## 2 関連研究

もの探しに際して物体認識を利用した先行研究は多く見られる。もの探しを効率化するウェアラブルシステム I'm Here! [1] は、日々の生活で物体を把持した際の視点映像を記録することで、探索対象の物体を最後に手にした際の映像をユーザに提示する方法を提案している。また、原田らの AI Goggles[2] は、カメラを備えたゴーグルと Head Mount Display(HMD) を装着し、ユーザの視界を撮影したカメラ画像にリアルタイムでタグ付けを行い、その結果を HMD 上に出力する。物体認識技術には、事前に学習データとして登録された物体の画像に加え、このデバイスを使用中に手に持った物体の画像データを追加学習に充てることを可能とした。しかし、[1],[2] では、生活する上で常時装着する必要がある、視点映像上で最後に確認された場所にもものがある可能性が高いという提案方法であるのに対し、本研究では、オンタイムで視界のどこにあるかを提示するデバイスを目指す。

## 3 実装

### 3.1 システムの概要

今回の検証に使用するデバイスのシステム構成を図 2 に示す。本システムは、PC<sup>2</sup>、Web カメラ<sup>3</sup>、ワイド



図 1: デバイス外観と装着例

コンバージョンレンズ<sup>4</sup>を照射口に付属させたモバイルプロジェクター<sup>5</sup>で構成される。PC 上で動作するプログラムにおける画像認識技術は Google Cloud Vision API を利用した。Web カメラとプロジェクターを肩に固定し、Web カメラでキャプチャされた画像を PC 上のプログラムにて画像認識を行う。ユーザーの入力した検索ワードと一致する物体が検出された場合、プロジェクターからその物体の位置にマーカーを照射する。

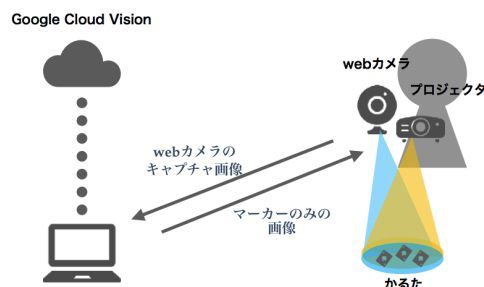


図 2: システム構成図

Web カメラから取得したキャプチャ画像を処理し、その画像から得られた物体の座標を元にプロジェクターからマーカー映像を照射するため、Web カメラとプロジェクターの光軸はなるべく近づくよう接着を行った。また、機器の焦点距離が異なり、Web カメラに映り込む範囲に対してプロジェクターの照射距離が長い場合、プロジェクターの照射口にさらにワイドコンバージョンレンズを付属させ、近くに照射できるようにした。Web カメラとプロジェクターを装着する部位に関して、撮影範囲が広がる点において頭部など対象物から遠い箇所がふさわしいと想定される。しかし、撮影画像がぶれにくく対象物を発見した際に照射するプロジェクターも合わせて身に付けることを考慮し、頭部より動きが少なく安定する箇所として肩を提案する。

### 3.2 テキスト検出

本デバイスでは、実世界にある探索対象をより早く見つけだすことを目的とし、文字認識を利用したテキスト検索を実装した。画像認識技術は Google Cloud

<sup>1</sup><https://prtimes.jp/main/html/rd/p/000000006.000022312.html>

<sup>2</sup>Apple 社 MacBook Air, macOS 10.12.6, 1.6 GHz Intel Core i5, メモリ 8GB

<sup>3</sup>Logitech 社 HD Pro ウェブカメラ C910

<sup>4</sup>Kenko SD-05W, 0.5 倍

<sup>5</sup>サンワサプライ社 400-PRJ014

Vision API の TEXT\_DETECTION を用いた。なお、現在リリースされているバージョン<sup>6</sup>では、テキスト検出以外に、一般物体のラベル、ロゴ、ランドマーク、顔検出に対応している。

### 3.3 ソフトウェア

本システムのアプリケーションは、ユーザーから任意の検索ワードの受け取り、画像認識サービスの呼び出しと結果受け取り、検索ワードにヒットする物体の位置を実世界の光景に照射する際の補正を担う。以下、今回の実験で行うかるた取りを想定した文字認識に絞った実装を記す。

#### 3.3.1 マーカー照射位置補正

ローカル環境での Web カメラのキャプチャの速度<sup>7</sup>に対し、インターネット<sup>8</sup>を介した画像認識サービスとのやり取りは約 1~2 秒ほどの実行時間がかかり、物体の認識結果が約 2 秒前のカメラ位置から見た座標位置が得られることとなる。本システムの目的である動きながら探しものをしている最中に使うことを踏まえ、遅延のない地点に照射できるように補正する。具体的には、ポストした際のキャプチャ画像と、画像認識結果が返ってきた時点でのカメラのキャプチャ画像の差分から逆向きに移動させた座標にマーカーを照射させることで実現する。

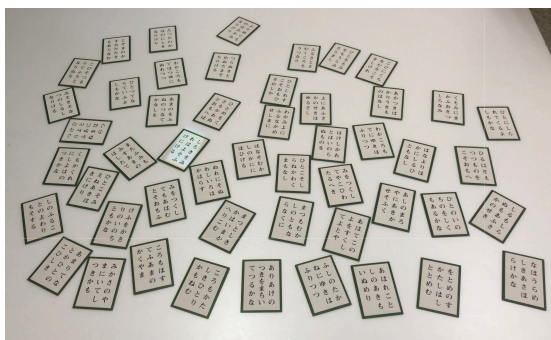


図 3: 照射例

## 4 実験

### 4.1 実験概要

web カメラと計算機を利用したもの探し支援の有用性を検証するための以下の実験を行った。

広範囲にあるテキストを探す環境としてかるた取りを想定した実験を行った。表向きに散らばった状態のかるた(図3参照)の中から、指定された語句ではじまる 1 枚を見つける作業を、デバイスを装着した人物に行ってもらい、人と計算機が対象の語句を含むかるたを見つけるまでにかかった時間を比較する。図4は探索時間のヒストグラムである。

### 4.2 結果

図4にあるように、人の探索時間が10秒あたりに集中するのに対し、デバイスの探索時間の最頻値が0.3秒にきている。認識できた文字に関しては非常に高

速な反応がみられることがわかる。

また、探索語句が”ひとの”や”ひとや”など頻出の文字列(この場合は”ひと”)を含む場合、人が時間のかかる場合にも認識できている場合は高速な反応が得られる。同時に複数箇所の照射も可能のため、現実世界での検索コマンドとして効率的なキーワード検索の可能性が伺える。

一方で、図4のグラフの右端にみられる探索に長い時間がかかるケースもみられ、全体では人よりも先に計算機が見つかる試行は全体の62%ほどに留まった。見つかるのに長時間を要した場合の探索画像の多くでは、文字がぼけて”わ”, ”ね”などの似た文字や”ふ”といった画数の多いひらがなが正しく読み取れていないことがわかった。この点に関しては、照明でより明るくする、より高解像度のカメラを利用するなどで精度の向上は可能だと考えられる。また、キャプチャした画像の中でよりぶれの少ないものを選別する工程を加えることでもより効果的な画像認識結果が得られると考えられる。

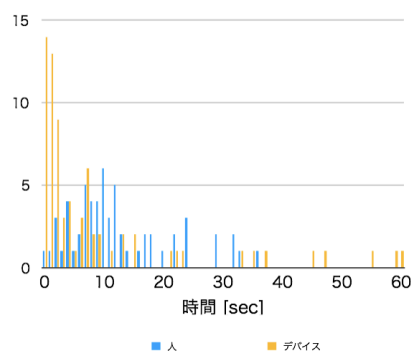


図 4: 探索の所要時間

## 5 まとめ

今回は、画像認識を利用したデバイスによるもの探し支援の検証を行った。精度面での改善点が挙げられるが、この先インターネット速度やカメラの視野角や解像度などの機器の性能が向上することで、本システムが有効な場面が増えると考えられる。例えば、紙媒体での書類をざっと見渡すだけで、欲しいキーワードが記載された書類を探し出すといった応用が考えられる。今回は文字認識に限った実装となったが、画像認識技術として採用した Google Cloud Vision API では、物体認識も提供されているため、一般物のもの探しへの発展も期待できるだろう。

### 参考文献

- [1] 上岡隆弘. I'm here! : 物探しを効率化するウェアラブルシステム. ヒューマンインタフェース学会論文誌, Vol. 6, No. 3, pp. 19-29, 2004.
- [2] 達也原田, 英樹中山, 康夫國吉. Ai goggles : 追加学習機能を備えたウェアラブル画像アノテーション・リトリバルシステム. 電子情報通信学会論文誌. D, 情報・システム = The IEICE transactions on information and systems (Japanese edition), Vol. 93, No. 6, pp. 857-869, jun 2010.

<sup>6</sup>2018年9月28日更新

<sup>7</sup>約 24fps 程度

<sup>8</sup>