

A Proposal for Object Detection System Using YOLO v3 in Embedded System Development

Hiroataka YONEDA^{*}, Erina MAKIHARA^{**}, Takeru BABA^{**}, Yuya MAEDA^{*}, Mitsunori MIKI^{**}

(Received January 21, 2020)

In the embedded systems development that needs to consider many factors simultaneously, it is expected that the learning efficiency will be improved by passing experts' implicit tips on to beginners. In this paper, we focus on gaze information that reflects implicit tips and propose an attention object detection system using a wearable eye tracker. We constructed a system for automatically recognizing an object by machine learning using YOLO v3 and identified four elements in the embedded systems development; display, paper, Arduino and circuit. As the result of the experiment on the evaluation of the recognition accuracy, the target object checked by humans and the target object detected by the system matched 88% on average. When it comes to the information which we needs to focused on as a ratio, such as the attention time of each object in the entire developing time and the ratio of attention transitions between objects, our system should contribute to the analysis of the gaze information of a developer's attention object.

Key words : embedded system development, education, gaze, object recognition

キーワード : 組込みシステム開発, 教育, 視線, 物体認識

組込みシステム開発における YOLO v3 を用いた注目物体検出手法の提案

米田 浩崇, 榎原 絵里奈, 馬場 建, 前田 侑哉, 三木 光範

1. はじめに

IoT 市場規模は拡大しており, IT 人材の中でも特に IoT 技術者の需要が増加すると予想されている¹⁾. IoT における開発では組込みシステム開発に必要な知識に加え, ネットワーク構成や通信に関する知識も必要であり, これら技術を総合的に習得することは容易ではない. したがって, IoT の基礎技術である組込みシステム開発を取り巻く学習環境の整備は今後の課題になると考えられる.

組込みシステム開発では, ソフトウェア (以下, SW) 開発に加えてハードウェア (以下, HW) 開発が必要である. システムが正常に動作しない原因, すなわち誤りの箇所は, SW だけでなく HW にも発生しうる. HW の誤りとして, 配線の間違いや電源不足, HW そのものの故障などがある. 組込みシステム開発の初学者にとって多くの場合, 誤りが SW と HW のいずれに起因するかは見分けがつかない. HW の誤りが原因でシステムが正常に動作しない場合でも, その誤り箇

^{*} Graduate School of Science and Engineering, Doshisha University, Kyoto

Telephone:+81-80-3866-7382, Fax:+81-790-45-0043, E-mail:hyoneda@mikilab.doshisha.ac.jp

^{**} Faculty of Science and Engineering, Doshisha University, Kyoto

Telephone:+81-774-65-6780, Fax:+81-774-65-6716, E-mail:emakihar@mail.doshisha.ac.jp

所を把握できなければ SW の誤りと判断する可能性がある。SW と HW のどちらに誤りが存在するかの推定には、開発者の経験則が大きく関わっている。したがって、組込みシステム開発の初学者にとって、システムの誤り箇所を適切に推定し修正することは容易ではない。

本研究では、組込みシステム開発中の作業者の視線移動に注目し、熟練者が持つ組込みシステム開発特有の暗黙的知識を定量的に分析する。具体的には、アイトラッカーという視線を検出する機器を用い、熟練者がどのような要素に注目しながら開発を行っているのかを調べる。しかし、アイトラッカーを用いた注目物体の判定は人間が手作業で行うことが多い。暗黙的知識の定量的分析には多量の実験データが必要であるため、実験から得たデータから注目物体を手で解析することは膨大なコストがかかる。したがって本稿では、組込みシステム開発中の作業者の注目物体を自動的に検出するシステムを提案し、その妥当性を検証する。

2. 視線分析

2.1 視線の特徴

人の視線は、その人物の能力や思考プロセスによって動きが変化するという特徴がある²⁾。ある作業中の視線の動きには、作業者が意識的にやっている視線移動に加えて、作業者が言語化できない暗黙的知識による視線移動が反映される³⁾。そのため、視線を分析することで、熟練者が暗黙的にやっている伝達困難な行動を明らかにできると考えられる。

人間の視線は主に 2 つの動きに分類できる⁴⁾。1 つ目は saccade とよばれる視線の高速移動で、注目物体の移動速度が 30 deg/s を上回るときや、離れた位置の物体に視線を動かす際に起こる。2 つ目は fixation とよばれる視線の停留で、注目物体の移動速度が 30 deg/s を下回るときや、静止物体を見る際に起こる。人間は saccade と fixation を交互に行い、視覚認知の多くを fixation で行っている。そのため、fixation に注目することで注目物体の検出が可能である。

2.2 アイトラッカー

アイトラッカーとは、人間の視線を定量的な座標で取得できるデバイスである。アイトラッカーは非接触型とウェアラブル型の 2 種類に分類できる。

非接触型アイトラッカーは、VDT 作業と同時に使用されることが多く、作業者がディスプレイ上のどの部分を注視しているかを高精度に取得できる。ディスプレイ上の座標と視線の fixation の座標を一致させて取得できるため、注目部位の判断が容易である。そのため、ソースコードリーディングやパッケージや広告などのデザイン、web 閲覧解析など幅広い分野で利用されている⁵⁻⁷⁾。

ウェアラブル型アイトラッカーは、作業者が装着して使用し、一人称視点の映像上のあらゆる物体に対する視線を取得できる。眼鏡型のデバイスが多く、視線の対象に制限はないが、読書中の文章の位置の検出のような高精度な視線情報の取得はできない。

2.3 アイトラッカーから得られる情報の定量分析

アイトラッカーを用いることで、fixation 時の座標を取得できる。fixation 時の座標から分析できる定量データとして、注目座標や停留時間、視線の移動距離などがある。加えて視線映像上の物体領域情報を組み合わせることでより多くの情報が取得できる。

視線には暗黙的知識が反映されるため、視線分析により暗黙的知識を定量的に表せるとされている³⁾。Busjahn ら⁸⁾ は、ソースコードリーディングにおいて注目した関数の順序に関してマルコフモデルを作成した。松本ら⁹⁾ は、プログラミング実装課題において、ソースコード上の高得点群の fixation に関するヒートマップを初学者に提示することで、低得点群の回答速度が向上したと報告している。ソフトウェア工学分野以外では、看護¹⁰⁾ やドラム演奏¹¹⁾ でも熟練者の暗黙的知識を教示する効果が報告されている。このように、熟練度による動作の違いを定量的に表すことができるため、視線情報は広い分野で活用されている。

2.4 課題

ウェアラブル型アイトラッカーには多くの場合カメラが内蔵されており、装着者の視線情報と視線映像

の両方を取得できる。視線映像上に視線情報の座標をマッピングすることで、装着者が注目している物体を判定することができる。しかし、装着者が頭部を動かすことで視線映像は常に変化するため、視線情報の座標が同様であっても注目物体が異なる可能性がある。そのため、ウェアラブル型アイトラッカーを用いた注目物体の判定において、注目物体の判断は人間の手作業で行われているのが現状である。

ウェアラブル型アイトラッカーを用いて被験者実験を行う場合、実験中の映像を実験者が全て目視で確認し、注目物体の判定を行う必要がある。多数の実験協力者を必要とする実験では実験者にとって大きな負担となり、加えてヒューマンエラーが発生する可能性から、これは取得データの信頼性の低下につながる。そのため、ウェアラブル型アイトラッカーの装着者の注目物体を自動的に判定するシステムが必要である。

3. 注目物体検出システムの構築

3.1 手法

ウェアラブル型アイトラッカーの装着者の注目物体検出を自動的に行うため、筆者らは機械学習を用いた手法を提案する。アイトラッカーで撮影した映像から目的の物体を検出するため、YOLO v3 という手法で機械学習を行い、物体検出器を作成する¹²⁾。アイトラッカーには Fig.1 に示す Tobii Pro Glasses2 を使用する。本アイトラッカーで取得できる視線情報のサンプリングレートは 100 Hz であり、視線映像のフレームレートは 25 フレーム毎秒である。視線移動が fixation であると判定する条件は、60 ms 間の視線移動距離が 30 deg/s 以下とする¹³⁾。アイトラッカーから得られる視線情報、視線映像および物体検出器を用いて注目物体の検出を行う。これにより注目している時間、注目物体の遷移順序を取得できる。物体検出の手順を以下に示す。

1. 視線情報と視線映像を入力
2. 視線の fixation が生じた映像フレームに移動
3. 物体検出器にそのフレームの画像を入力
4. 物体検出器から物体位置を長方形の領域として取得



Fig. 1. Eye tracker used in the experiment.

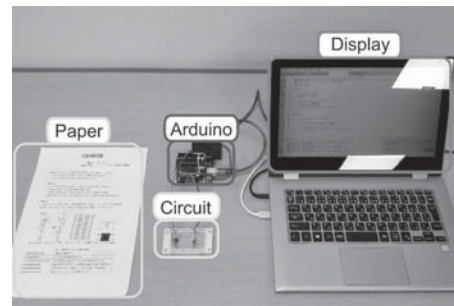


Fig. 2. Factors in the environment.

5. fixation の座標が物体領域内にあれば、物体領域の重心までの距離が小さいものを現在の注目物体とする
6. fixation の座標が物体領域内になければ、視線が注目物体の遷移過程であるとし、一つ前の注目物体を現在の注目物体とする

3.2 想定する環境

本研究では SW と HW の両方の要素を含んだ環境を構築し、複数の要素間における作業者の視線の移動を検出する必要がある。そのため、組込みシステム開発の要素として以下の 4 つを考える。

- ディスプレイ：SW 開発のための PC
- 問題用紙：HW 仕様や要件定義が記された用紙
- Arduino：回路を制御するマイコン
- 回路：デジタル入出力に関する回路

以上の 4 要素を Fig.2 に示す環境に配置し、物体検出器の学習を行う。

3.3 物体検出器の学習

システムが識別すべき組込みシステム開発の要素は、ディスプレイ、問題用紙、Arduino、回路の 4 つであ

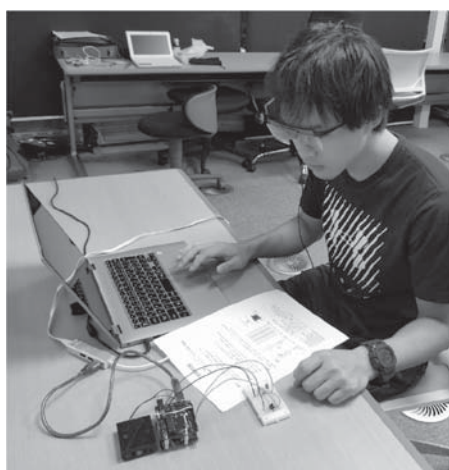


Fig. 3. Experimental environment.

る。物体検出器の訓練は 1800 枚の画像で行った。まず、作業者の視線映像から 2 秒ごとに 1 枚の画像を 100 枚抽出し、手作業でラベリングを行った。その後、各画像に対して加工を行い、さらにそれぞれを左右反転させたものを追加し合計 1800 枚とした。画像に施した加工はヒストグラムの平均化、平滑化、ガウシアンノイズ付与、ホワイトノイズ付与、コントラスト高化、コントラスト低化、明度高化、明度低化である。

4. 注目物体検出システムの評価実験

4.1 手法

注目物体検出システムの精度を評価するため被験者実験を行った。実験協力者は 20 代大学生の 3 人 (S1, S2, S3) である。本実験は提案手法の妥当性を検証することが目的であるため、実験協力者は 3 人ともに組込みシステム開発の経験がある学生とした。実験協力者はアイトラッカーを装着し、実際に簡易な組込みシステム開発の課題を行った。実験風景を Fig.3 に示す。実験中に取得した情報はアイトラッカーから得られる視線情報、視線映像および実験風景の映像である。

実験後、映像を目視で確認し実験協力者が注目している物体と注目した順序を記録し、これを正しい注目物体、すなわち真値とした。真値とシステムによって得られた注目物体を比較し、システムの妥当性を検証した。

Table 1. Performance of the proposed system.

sub.	precision	recall	F-measure	recognition rate
S1	0.73	0.79	0.76	82%
S2	0.71	0.85	0.77	86%
S3	0.66	0.96	0.78	95%

Table 2. Average error of a focused object.

sub.	recognition rate	ave.error	S.E.
S1	82%	1.6%	1.1%
S2	86%	2.8%	2.1%
S3	95%	2.9%	2.9%

4.2 結果

実験協力者が課題を完了するまでの時間は S1 が 20 分, S2 が 15 分, S3 が 32 分であった。実験で得られた適合率, 再現率, F 値および認識率を Table 1 に示す。ここでの認識率とは、実験全体において注目物体がシステムと真値とで一致した割合である。全被験者において適合率より再現率が高く, F 値の平均値は 0.77 であった。

要素間の遷移順序の検証として、真値とシステムの識別結果を比較する。本実験環境では識別すべき要素が 4 つであるため、各々の要素間の遷移は 12 通りある。真値とシステムの識別結果のそれぞれに関して、総遷移回数に対する 12 通りの遷移回数のそれぞれの割合を算出する。12 個の割合について、真値の割合とシステムの識別結果の割合との誤差から誤差の平均と標準誤差を求めた結果を Table 2 に示す。

4.3 考察

本実験では真値の判定を従来手法と同様に実験者が手作業で行っているため、実験協力者が実際に注目していた要素と真値にはある程度の誤差が発生する。しがたって、本システムの性能を測るために重視すべきは再現率、すなわち人間が判断した物体をシステムも同様に識別したときであると考えられる。Table 1 から、3 人全ての実験協力者において適合率より再現率が高く、提案システムの妥当性を裏付けている。また、真値と異なる判定をシステムが行った際の映像に注目すると、注目物体の判定範囲付近に視線はあるが微量

に判定範囲から外れているケースがあった。原因として、アイトラッカー自体の精度やキャリブレーションのずれなどが影響している可能性が考えられる。これは、各要素の判定範囲を広くし、各要素間の距離を十分に取ることで精度の向上が期待できる。

Table 1 に示す通りアイトラッカーの認識率は82%から95%であり、平均して全体の約12%を認識できていない。そのため、提案手法を用いて視線情報を定量的に用いる場合、注目回数や注目時間などの絶対的な数値に着目すると認識率が実験結果に影響を与える可能性がある。しかし、本研究では実験時間全体における各物体への注目時間や、実験全体における各物体への遷移回数といった割合が意味を持つ。Table 2 の標準誤差に示すように、各物体に対する標準誤差は3%未満に留まっている。そのため、真値や未検出による誤差を考慮しても、これが遷移確率の検定に与える影響は軽微であると考えられる。したがって、提案手法には物体への注目時間の割合や遷移順序を計測する妥当性があると考えられる。

5. おわりに

多数の要素を同時に考慮する必要がある組込みシステム開発において、熟練者が持つ暗黙的知識を初学者に伝えることで学習効率の向上が期待できる。本稿では、暗黙的知識が反映される視線情報に注目し、ウェアラブル型アイトラッカーを用いた注目物体検出システムを提案した。YOLO v3 を用いた機械学習により注目物体を自動的に認識するシステムを構築し、組込みシステム開発におけるディスプレイ、問題用紙、Arduino、回路の4要素の識別を行った。提案システムの識別精度評価実験の結果、人間が目視で確認を行った注目物体とシステムが判定した注目物体は平均88%一致した。作業時間全体に占める各物体への注目時間や、物体間の注目遷移の比率など、情報を比として捉える必要があるとき、提案手法は作業者の注目物体をめぐる視線情報の解析に寄与できると考えられる。

参考文献

- 1) 塩野正和, “IoT 市場の拡大と、日本における IoT 活用のありかた”, NRI Management Review **38** (2017),

- <http://www.nri.com/-/media/Corporate/jp/Files/PDF/knowledge/publication/m-review/2017/nmr38/nmr38-1.pdf> (参照 2020-1-20).
- 2) 大野 健彦, “視線から何がわかるか”, 認知科学, **9**[4], 565-579 (2002).
- 3) 檜山敦, 浅田和宏, 並木秀俊, 宮廻正明, 廣瀬通孝, “伝統技能継承のための主観視点を含んだ支援映像の生成”, ヒューマンインタフェース学会論文誌, **12**[3], 249-258 (2010).
- 4) Tobii Technology, “The Tobii I-VT Fixation Filter”, <https://www.tobiipro.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/tobii-pro-i-vt-fixation-filter.pdf> (参照 2020-1-20).
- 5) Olsen Anneli, “The Tobii I - VT fixation filter”, Tobii Technology AB (2012).
- 6) 應治紗織, 上野秀剛, “コードレビュー時の読み方指示によるレビュー効率の変化”, 電子情報通信学会技術研究報告, **114**[128], 1-8 (2014).
- 7) Tobii, “売上 15%増に結びついたパッケージデザイン調査”, <https://www.tobiipro.com/ja/fields-of-use/marketing-consumer-research/customer-case/strategir> (参照 2020-1-20).
- 8) Tobii, “Eye tracking TC commercial helped Team Foods recover market share”, <https://www.tobiipro.com/fields-of-use/marketing-consumer-research/customer-cases/yanhaas>. (参照 2020-1-20).
- 9) T. Busjahn, C. Schulte, B. Sharif, Simon, A. Begel, M. Hansen, R. Bednarik, P. Orlov, P. Ihanola, G. Shchekotova and M. Antropova, “Eye Tracking in Computing Education”, In *Proceedings of the tenth annual conference on International computing education research*, 3-10 (2014).
- 10) 松本光稀, 若原徹, “視線情報の分析に基づくプログラミング教育支援システムの提案”, 情報処理学会第 80 回全国大会講演論文集, **2018**[1], 721-722 (2018).
- 11) 西方真弓, 西原亜矢子, 定方美恵子, 清野由美子, 井越寿美子, 笠井美香子, 佐藤富貴子, 川合功, 坂本信, 小浦方格, 田邊裕治, “新人看護師の‘観察・判断への気づき’を育てる視線解析を用いた教育プログラムの評価:臨床経験豊富な看護師の‘DVD教材’視聴による気づきの分析”, 新潟大学保健学雑誌, **11**[1], 25-32 (2014).
- 12) 早川和輝, 長谷川大, 佐久田博司, “主観視点の 3D 手動作教材提示によるドラム演奏学習支援および熟練者視線情報を利用した教材による学習効果”, 知能と情報, **28**[1], 511-521 (2016).
- 13) J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement”, arXiv preprint arXiv:1804.02767 (2018).