

視覚スイッチの開発 Development of Eye Gaze Switch

小林 秀輔[†]
Syusuke Kobayashi

ピトヨ ハルトノ[†]
Pitoyo Hartono

1. はじめに

人間は日常生活空間の中で、電化製品をはじめ、それぞれのインターフェースで操作する必要がある多くの機械に囲まれ、工場などの生産現場においても状況は同じである。近年、この状況を改善するために人間にとってより直感的なインターフェースの開発が盛んに行われている。その中で、音声と自然言語認識によるインターフェースの開発が進み、一般家庭に普及しつつある。これらにより人間と機械間で、人間が普段使うバーバルコミュニケーションを用いてインタラクションを行うことができる。多くの場面でバーバルインターフェースは有効であるものの、人間同士においてはノンバーバルコミュニケーションも盛んに行われている。ノンバーバルコミュニケーションは言語情報以外を使って行われるコミュニケーションであり、表情・視線・姿勢・しぐさなどを用いる意思の伝達手段である。人間は、バーバルとノンバーバル手段をうまく組み合わせながらコミュニケーションを円滑に進める。人間はこの組み合わせがうまくいかず、相手の意図が読み取れない時には、ノンバーバル要素を優先して読み取ろうとする[1]。このことから、ノンバーバルコミュニケーションは人間のコミュニケーションで重要な役割を果たしている。

音声と自然言語認識によるインターフェースは、騒音が多い工場や日常的な空間では非効果的である。会話を控える必要がある公共の場においても利用が限定的となる。

本研究は人間と機械間のインターフェースの柔軟性を高め、より自然なインタラクションを可能にするマンマシンインターフェースの開発を主な目的とする。ここでは、視線を用いるインターフェースの開発を試みる。視線には、Kendon の研究[2]により、4つの機能があることが分かる。その中の一つとして、認知機能がある。これは、自分が相手に注目を向けていること、もしくは意思疎通の希望があるということを示す働きである。その認知機能を人間間で有効にするために、視線を知覚するアイコンタクト効果[3]がある。本研究では機械に対して視線を送ることで、動作開始・停止指示ができるインターフェース、視覚スイッチの開発を行う。

本研究は初期段階にあり、本論文ではインターフェースを実現するためのシステム構築に関して主に説明する。また、このシステムの利用性と制限評価を初期実験によって確認し、その結果を本論文で報告する。

2. システム概要

本研究で開発する視覚スイッチは人間が機械を視覚的にとらえることで、その機械を作動または、停止させることができるインターフェースである。これにより、複数の機

械に対し、発話、ジェスチャー、また手の操作によって指示を行うのではなく、人間のアイコンタクトに相当するコミュニケーションが可能となる。

この視覚スイッチを可能にするためには、人間が何を視覚的にとらえるかをシステムで認識する必要がある。具体的には、システムが人間の視界の中の様々な物体を認識し、それらの相対的な位置を特定する必要がある。その後、人間の視線がある物体の位置の狭い範囲に収まる場合、その物体とアイコンタクトが確立できたとみなし、動作指示をすることができる。

本研究で開発した視覚スイッチのシステムの概要を図1に示す。

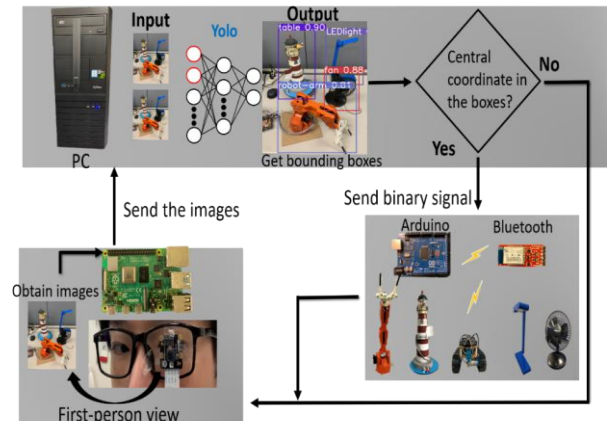


図 1: システム概要図

本インターフェースでは、メガネに小型カメラを装着することで人間の視覚をシステムにとらえることができる。作成したデバイスを図2に示す。

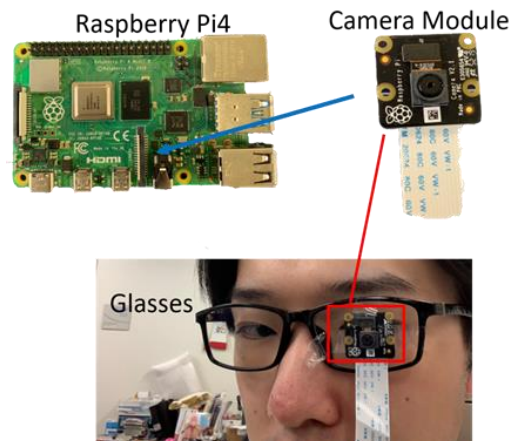


図 2: 作成したデバイス

[†] 中京大学 大学院工学研究科電気電子工学専攻
Dept. of Electrical and Electronic Engineering,
Graduate School of Engineering, Chukyo University

カメラは小型計算機, Raspberry Pi4 とつながり, 人間が視覚的にとらえたものを画像として保存し, 無線で計算機に送信する. 計算機では前もって物体検出を可能とするニューラルネットワーク(NN)が動作し, NNが学習した物体を検出する. ここでは, 物体検出アルゴリズムとしてYOLO[4]を用いた. YOLOはConvolutional Neural Network(CNN)に基づいた物体検出アルゴリズムであり, 数ある物体検出アルゴリズムの中で検出速度が速い. 本研究ではYOLOの物体識別の学習の対処として, 図3に示す5つの機械を用意した.

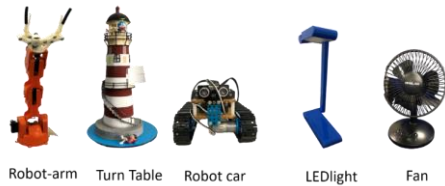


図3:動作対象例

ここでは, 各機械に対し, 様々な方向や環境で撮影し, 教師データとして774枚, テストデータとして124枚用意した. そして, 1枚の画像に対して物体の位置情報と分類を格納するアノテーションファイルも同数作成し, 学習を行った. 学習結果については3.1の予備実験で述べる.

Yoloによって物体を検出した後, 検出した物体の中に視線が収まっているかを確認する. ここでは, 人間が常に視界の中心を注視することを仮定し, 取得画像の中心座標を視線の位置として扱う. 計算機では, 取得画像の中心座標と, 検出した物体の位置座標を照合し, 人間が物体に対して視線を向けたことを確認する. 人間が物体に対して視線を向けたことを確認できた場合は, 無線で対象機械に対して信号を送ることで動作開始・停止指示をする. この流れをリアルタイムで動作させることで, 視覚スイッチシステムを実現した.

3. 実験

本インターフェースの利用性, 操作性, および制限を評価するために実験を行い, その結果を報告する.

3.1 予備実験

まず, 本システムの利用範囲に関して予備実験を行う. この予備実験の必要性は, 人間のアイコンタクトには知覚範囲[3]が存在し, この知覚範囲によって, アイコンタクトが正常に行われずコミュニケーションが円滑に行えないことがある. 従って, アイコンタクトを模した本システムにおいて知覚範囲は重要である. 物体認識率と, 距離に応じた認識精度がシステムの利用性に影響する. 予備実験ではこの二つの指標を用いてシステム利用範囲を評価する. 具体的には, 学習に用いた物体の認識率である mean Average Precision(mAP)と, 学習に含まれる物体で様々な距離に撮影した画像の認識率を有効範囲とする評価結果を図4にそれぞれ示す.

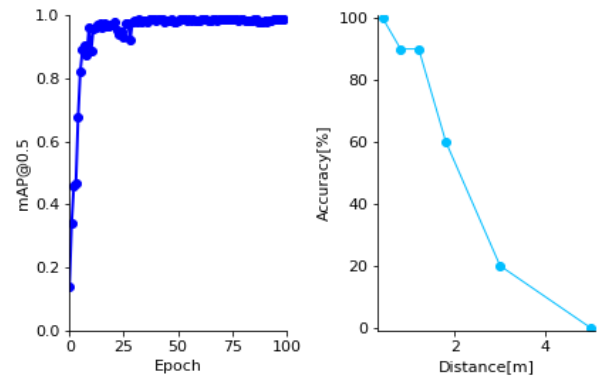


図4:知覚範囲結果

図4から, mAPは0.9を超え, NNの学習により物体の識別ができること分かる. 有効範囲に関しては, 人間と物体の距離が増えるに従い認識率が大きく低下することが分かる. このことから, 本システムの有効な知覚範囲は約2mであることが分かる.

予備実験後, 被験者によるシステムの評価実験を行う. 実験の前に, 被験者に対して本インターフェースの説明する. その後, 被験者自身がインターフェースを使って機械を動かす練習を5分程実施する. 実験では被験者に対して動作対象をランダムに指示し, 視覚を用いてそれを動かす. 一定時間内に動作対象を動かすことができるかどうかを評価する. 動作範囲に関する予備実験結果を踏まえ, 被験者と物体との距離は2m以内に収まるよう実験環境を整えた. 実験において, 被験者に10回の指示が行われ, 図3に示す各動作対象が2回ランダムな順番で指示される.

3.2 実験1

本インターフェースの利用性を確認するために行う. 図5に示す設定で実験1を行った. ここでは, モニターは被験者に動かす対象物を指示するために用いた.

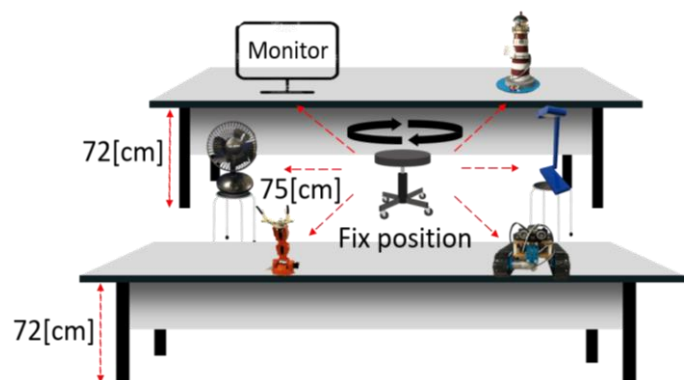


図5:実験環境1

動作対象の画像がモニターの画面に映されたら, 10秒間以内に対象物を直視し動かした後に, モニターに視線を戻すよう指示した. 10秒間までに動作できない場合には実験は続行するが, そのタスクは失敗と見なす. 実験中, 被験者はFix positionにある回転イスに座るよう指示し, 回転イスを使って体の向きを変え, 視覚を動かし物体をとらえ

るように指示した。表1に対象物体を動かす平均成功率を示す。本実験は10人の被験者に対して実施した。

表1:実験結果1

	Robot-arm	Fan	LEDlight	Turn Table	Car	Average
Accuracy[%]	94	83	89	100	61	86

全体の平均成功率は86%となり、被験者は指示された物体の方向に視覚を向ける事で、動作させる事ができた。しかし、各機器の平均成功率を見た時には、Carが低い。これは、Carの動作中に所定の位置から動き回り、OFFにしようとする際に動き続ける物体に対して視覚にとらえる必要があるため、所定の位置から離れない他の対象物より、制御しづらいことに理由がある。

3.3 実験2

次に、2つ目の実験を図6に示す設定で行った。

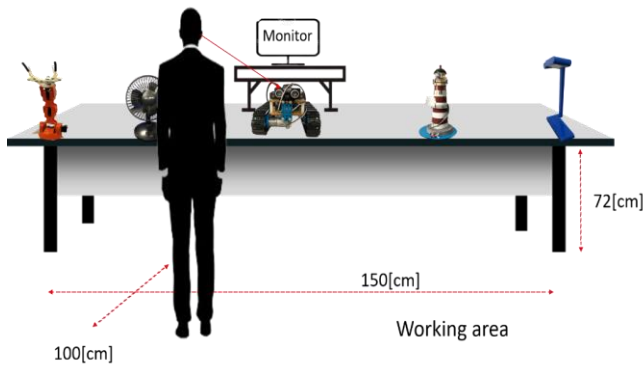


図6:実験環境2

1つ目の実験と同様のタスクを行うが、この実験では図6のような被験者が自由に動くことのできる空間を用意した。より日常空間に近い環境で本インターフェースを用い、その操作性を評価するために行う。この実験の被験者は9人である。表2に実験2における平均成功率を示す。

表2:実験結果2

	Robot-arm	Fan	LEDlight	Turn Table	Car	Average
Accuracy[%]	100	100	83	94	94	94

実験1と比べて良好な結果となった。特に、Carの成功率が大きく上昇した。これは、Working areaを与える事によって、被験者各々が本インターフェースを使いやすいように体を動かすことができるからである。例えば、Carを停止させる場合には、カメラの画角にCarが大きく収まるように近づけば、物体の位置座標範囲が大きくなることで、座標範囲に中央座標を収めることが容易になる。

4. 結論及び課題と今後の展望

本研究では視覚スイッチのハードウェア platform の開発を行い、その利用範囲を確認した。実験においては、本インターフェースを用いて物体の方向に視覚を向けることで、その物体が動かせることを確認した。現段階では、利用範囲がまだ狭いものの、機械とのノンバーバルインタラクションを実装することができた。

課題としては、人間の視界とカメラでとらえた視界の間に乖離があることで操作性が低下することが挙げられる。解決方法は、カメラでとらえた視界を人間の視界に近づけることで考えている。具体的には、カメラを二つ用いて画像を取得する。このカメラは、人間の視野角が全て確保できるようにメガネに装着する。次に、その二つのカメラの画像を入力として、NNによって人間の視界の画像を生成する。NNには、人間の視界の画像を生成するように学習させる。教師信号には、メガネと装着者の目の距離を仮定し、そこから撮影した画像を教師データとして学習させる。これによって実現する。

また、利用範囲の狭さも課題である。これに関しては、本システムで用いたYOLOの構造を変えることで解決することを考えている。NNは、パラメータ数が少ないと計算回数が少ないために高速で出力することができる。しかし、特徴量を細かに抽出できなくなるために利用範囲が狭くなる。本システムの実装にあたって、物体検出速度が操作性に影響すると仮定したため、YOLOの中でもパラメータ数が一番少ないTinyYOLOを用いた。そして、入力画像サイズも低く設定した。これにより、利用範囲を広く認識するように、検出速度との兼ね合いを考えながら、適切な構造を選び解決していきたい。

人間はコミュニケーションを、それぞれの場面に応じて適切な方法を取ることで円滑に行う。機械とのコミュニケーションも同様に、適切な場面に応じてそれぞれのインターフェースを使用することができれば、便利である。今後の展望として、これを実現できるように、本研究で作成するインターフェースと他のインターフェースを組み合わせ、よりマルチモーダルなインターフェースを目指す。そして、人間と機械との関係がより自然になる事を実現したい。

参考文献

- [1] Mehrabian, Albert, "Communication without words.", *Communication theory*, pp.193-200 (2007).
- [2] Kendon, Adam, "Some functions of gaze-direction in social interaction.", *Psychologica*, 26:22-23 (1967).
- [3] Senju, Atushi and Johnson, Mark, "The eye contact effect: mechanisms and development.", *Trends in Cognitive Sciences*, Vol.13, No3, pp.127-134 (2009).
- [4] Redmon, Joseph, et al., "You only look once: Unified, real-time object detection.", *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.779-788 (2016).
- [5] Kim, Jung-Hwa, Choi, Seung-June and Jeong, Jin-Woo, "Watch & Do: A smart IoT interaction system with object detection and gaze estimation.", *IEEE Transactions on Consumer Electronics*, 65.2:195-204 (2019).