

# Pepperを使用したテレグジスタンスロボットの色による 情動表現の開発

## Development of Emotional Expression by Color for a Telexistence Robot using Pepper

大海 悠太, 木俣 雄介, 仲野 匠, 辛 徳

Yuta Ogai, Yusuke Kimata, Takumi Nakano, and Duk Shin

東京工芸大学、243-0297 神奈川県厚木市飯山南5-45-1  
5-45-1 liyama Minami, Atsugi-shi, Kanagawa 243-0297, Japan

### 概要

テレグジスタンスロボットに情動表現を持たせるため、YOLOとCATRについてIAPS画像との対応関係を調べ、Pepperと連携して色などを変化させるシステムを構築した。また、Pepperをテレグジスタンスロボットとして改良し、動作実験を行った。

### 1. はじめに

アフターコロナにおいて、オンラインとオフラインを混ぜたハイブリッド環境の重要性が注目されている。テレグジスタンスロボットの利用が考えられるが、操縦者の全ての身体表現を伝えるのは難しい。IAPS(国際情動写真集)[1]は心理実験などに広く用いられており、人間が写真を見て感じる情動についての知見が蓄えられている。この知見を生かし、ロボットがカメラ入力に対して物体検出をした結果への反応をデザインすることで、ロボットが情動を持っているかのように振る舞わせることができるのではないかと考えている。そして、テレグジスタンスロボットを使用する際に、ユーザーが即座にコントロールできない感情的な反応を即座に出せるようになると期待している。

IAPS画像データと、物体認識アルゴリズムYOLO[2]及びキャプション生成アルゴリズムCATR[3]の対応関係を調べることで、ロボットの情動反応設計への利用可能性を調べる[4][5]。また、Pepperロボットへの搭載方法について検討を行う。

Pepperロボットを遠隔操作し、展示物の閲覧、及び画像データからの自動情動表現を色を用いて行うシステムを開発する。

### 2. 方法

#### 2.1 IAPSとYOLOとの傾向調査

IAPSの721枚の画像に対して、YOLOv4による物体検出結果を取得し結果の傾向を調べる。YOLOv4は事前学習済みのデフォルトのものを用いる。IAPSでは”man”、”woman”といった人間の姿が区別してラベルに記載されているが、YOLOでは手足などのパ

ーツであっても”person”として検出される。そこで、一致しているかどうかの整合性を調べるためのテーブルを作成し、傾向を調べた。

#### 2.2 IAPSとCATRとの傾向調査

IAPSデータ中、721枚の画像に対して、CATR(2021年5月22日版)によるキャプション生成結果の傾向を調べた。CATRは事前学習済みのデフォルトのものを用いる。傾向は著者が主観的に「ほぼ完璧に説明できている」「一部説明できている」「説明できていない」の3段階で評価した。評価の軸としては、”キャプション内の単語と画像内容が一致している”と”関係ない単語が含まれていない”というようにしている。



Fig. 1 Pepper robot with 360-degree camera.

#### 2.3 PepperとYOLO、CATRの連携

Pepperのカメラから画像を取得し、YOLOとCATRにかけて物体検出とキャプション生成を行う。検出した物体を高い順に並び替え、Pepperで物体名を信頼度の高い順に読み上げる。色について書く。

#### 2.4 Pepperのテレグジスタンスロボット化

Pepperの頭上にInsta360air全方位カメラを装着し、カメラと接続したGoogle Pixel2をPepperの脚部に取り付けた(図1)。またPCのキーボード入力により、前後移動と回転移動をできるようなCUIインターフェースを構築した。PCとPepperはWifiで接続されている。Google Pixel2からYouTube Liveに360度映像を送る。

PC上で移動命令を送ってから、YouTube Liveの映像内でPepperが動き出すまでの時間の計測を10回行い、遅延時間を求めた。また、実験参加者4名にYouTube Liveの画面を見ながら、Pepperに5m直進させてから右90度旋回させ、また5m直進させてから戻ってくるという往復20mのコースの操縦をさせた。

### 3. 結果・考察

#### 2.1 IAPSとYOLOとの傾向調査

IAPS画像721枚についてYOLOv4で物体検出させたところ、616枚で物体を検出できた。1枚の画像の中で最も信頼度の高い物体について、信頼度0.7以上は459枚であり大半が正答していた。情動反応設計の元としては十分に有用であると考えられる。

検出結果からIAPSとYOLOのラベルの対応関係を目で確認し、ラベルテーブルを作成した。これによりYOLOで検出した結果から、どのIAPSラベルが対応しているかを求めることができる。

#### 3.2 IAPSとCATRとの傾向調査

721枚の画像に生成されたキャプションを評価した結果を図5に示す。「ほぼ完璧に説明できている」のは221枚、「一部説明できている」のは249枚、「説明できていない」のは251枚だった。そのまま利用するには誤検出が多くなる恐れがあると言える。CATRはオプションにv1, v2, v3を付けることで、異なるバージョンの出力を得ることができる。それらの出力で共通する名詞を用いるなどすることで、精度を向上させられると考えている。また、YOLOと同じ画像を入力させた場合の比較では風景画像についてはYOLOよりも有用な出力を出す傾向が見られた。

#### 3.3 PepperとYOLO、CATRの連携確認

PepperとYOLO、CATRを連携するスクリプトを作成し、Pepperのカメラ画像に対し自動的に物体検出とキャプション生成をできるようになった。また、その結果によってPepperの色や動きを変化させることができた。

#### 3.4 Pepperのトレイグジスタンスロボット化

PC上でPepperに移動命令を送ってから、YouTube Liveの映像が動き出すまでの平均時間は6.57秒であった。また、実験参加者4名にPepperをYouTube

Liveを閲覧しながら遠隔操作をさせたところ、往復20mのコースを平均5分54秒で移動させることができた。映像は荒く遅延も大きいですが、Pepperの操作自体は可能なことは示せたといえる。

図2に示すようなメディアアート作品をPepperで体験するシステムを構築しており、この研究発表当日にデモを行う。



Fig. 2 Pepper robot with a media art.

### 4. まとめ

YOLOとCATRについてIAPS画像との対応関係を調べ、Pepperと連携して色などを変化させるシステムを構築した。Pepperをトレイグジスタンスロボットとして改良し、動作実験を行なった。

### 6. 参考文献

- [1] Lang, P.J., Bradley, M.M., & Cuthbert, B.N., "International affective picture system (IAPS): Affective ratings of pictures and instruction manual". Technical Report A-8. University of Florida, Gainesville, FL. (2008).
- [2] Bochkovskiy A., Wang, C., & Liao, H. M.: YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv:2004.10934 (2020)
- [3] CA:TR: Image Captioning with Transformers, <https://github.com/saahiluppal/catr>, Accessed 2022-12-20.
- [4] 木俣 雄介, 辛 徳, 大海 悠太, "IAPSと物体検出アルゴリズムによるロボットの情動反応設計の検討", 人工知能学会全国大会論文集(JSIAI2022), 4Yin245-4Yin245 (2022).
- [5] Yusuke Kimata, Duk Shin, Yuta Ogai, "A Study of Robots' Emotional Expressions Using IAPS, an Object Detection Algorithm, and a Caption Generation Algorithm", Proceedings of The Twenty-Eighth International Symposium on Artificial Life and Robotics 2023 (AROB 28th 2023), pp.117-120, (2023).