

視覚障害者のための同形物体のリアルタイム認識システムの検討

張 文軒† 渡邊 裕司†

名古屋市立大学理学研究科‡

1. はじめに

VLEG/GBD 2020[1]によると、2020年に全盲になる人は30年前と比べて50.6%近く増加し、今後30年で2倍になると予測されている。これまでの視覚障害者支援システムの研究は、触覚認識や音声補助が中心である。触覚は形状でしか判断できないため、均質なものや手の届かないところにあるものを扱えない。触覚では牛乳パックや缶など日常生活で遭遇する同形物体を視覚障害者が認識することができない。一方の音声支援は、QRコードや点字などを必要とせず、健常者が音声で対象物を教える場合に最適な補助具であり、同形の対象物や色の変化にも影響されないのが特徴である。これは、健常者が対象物の特徴を抽出して記憶しているためである。ディープラーニングを用いたコンピュータビジョン技術の急速な発展により、この健常者の役割をコンピュータに置き換えることができるようになった。しかし、既存の視覚障害者向けの認識アルゴリズムのほとんどは道路検出に用いられており、色やテクスチャに基づくため検出速度が遅い。

本研究では、高速な物体検出が可能なYOLOv5モデル[2]を用いて、リアルタイムで同形物体を認識して音声で結果を通知するシステムを構築する。同形物体として三種類の牛乳ボックスを選び、さまざまな色の光源をシミュレートしつつ、最適な認識距離と性能を評価するための実験を行う。結果として、距離60cmで平均97%の認識率を達成し、各色の認識精度を向上できたことを示す。

2. リアルタイム認識システム

本研究で構築するリアルタイム認識システムの構成図を図1に示す。より幅広い環境で対象物を検出できるようにするために、カメラ映像やビデオやネット上の画像などの様々な画像データをWindows APIで受け入れ、必要に応じて画像合成や画質変換などを行うことにより様々な環境をシミュレートする。対象物識別にはYOLOv5

モデルを使用し、物体検出後に物体の種類のリストと各物体の位置を出力する。出力側の発音モジュールは、この物体リストを受け取り、対応する音声プロンプトを発話する。

YOLOv5モデルを用いる理由は、高速な物体検出ができるだけではなく、YOLOシリーズが従来の認識モデルに比べて低リソースでクロスプラットフォームに展開できることもある。モデルをコンピュータ上で学習させ、モバイルプラットフォーム上に展開することにより、視覚障害者が容易に持ち運びできる支援システムとなる。

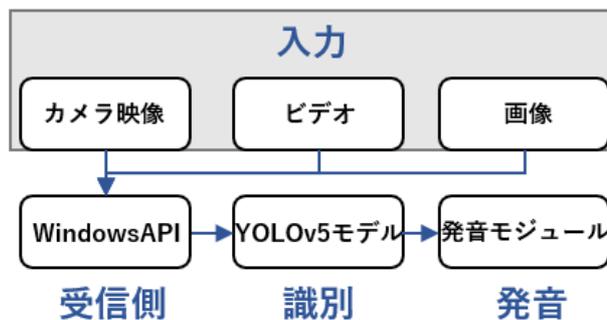


図1 リアルタイム認識システムの構成図

3. データセット作成

本研究の認識システムに対して、まず公開されているフルーツデータセットで学習させたところ、外観がほぼ同じであるにもかかわらず、異なる色の同類フルーツを識別できることを確認した。

次に、日常生活で必要とされるほぼ同形で異なるパターンの物体を使用してデータセットを作成する。具体的には、三種類の牛乳ボックス（明治、森永、北海道）をデータセット作成の対象とする。牛乳ボックスを対象とする理由は、円柱状の製品と異なり、角柱の牛乳ボックスを異なる角度から見たときにより複雑な形状を示し、表面のパターンがより多様で、同類の製品間の外形がより似ていて差異が少ないためである。

録画ビデオから画像を切り取る方法で6,513枚の画像を抽出し、4,027枚の画像を選択した。そして、これらの画像をインターネットから収集された画像と混合してデータのアノテーションを行い、6,265枚のデータセットを作成した。

A Study of Real-Time Recognition System for Isomorphic Objects for Visually Impaired People

†Zhang Wenxuan, Watanabe Yuji

‡Graduate School of Science, Nagoya City University

4. 実験方法

実験では、まず距離の変化（実際には物体が画像内で占める割合の変化）に伴って物体の認識精度がどのように変化するかを調べる。そのために、カメラを装備した移動可能なデバイスを設計して構築する。対象物から7-200cmまでカメラが移動する間にビデオ録画する。ルーラーを使用してカメラの位置と対象物からの距離を求める。

次に撮影方向の違いの影響を調べるために、対象物から15cmの距離で8方向の角度から撮影した画像に対して認識率を求める。

また視覚障害者の家庭環境では、極端な光源の変化が起こりやすいと考えられるため、極端な光環境下での検出の適応性と安定性を評価する必要がある。そのため、撮影画像に対して幅広い明暗とコントラスト調整を行い、6種類（RGBとCMY）の強度の異なる色を画像に適用して、極端な光源環境をシミュレートする。

さらに、公開されているデータの識別結果に基づいて、データセットを補充し、モデルを再トレーニングし、認識精度が低い物体に対するデータ拡張後の状況も検証する。

5. 実験結果

本研究のリアルタイム認識システムは、強光や弱光の環境での画像をトレーニングセットに追加せずとも、明暗とコントラストの大幅な変化に対しても良い適応性を示した。暗い環境や高コントラストの環境であっても、60cm以内で認識率を90%以上に維持した。しかし、画像の明暗が特定の閾値に達した場合、認識率は急激に減少し、明暗が減少または増加し続けると認識不能の状態に陥った。この現象は、後述の色実験でも現れた。

図2の破線に示す色強度を変化させた場合の結果では、RGBの3色に対して良い適応性が見られたが、反色のCMYでは相対的に悪い性能であった。特に、シアン環境では色強度の変化に伴い認識精度が最も急激に悪化した。この状況を改善するために、データを色強化、HSVランダム着色アルゴリズムの使用に加えて、データセットに600枚の着色画像を追加した。図2の実線で示す改善後は、シアンの環境を除くと、環境の色が大幅に変化しても、高い認識精度を維持し、対象物を見失わないことが分かった。

また、YOLOv5モデルは小さい対象物を検出しにくいという弱点が確認された。ただし、小さい対象物の識別能力が低いという特性を逆用することで、複数の対象物が画面内に存在する場

合に発音時の優先度の問題を解決することができる。つまり、システムは距離の近い大きな対象物に特化し、主要物の背景にある物体を無視することができる。

最終的な距離に対する識別率を図3に示す。7cmから始めて各ファイル間の間隔は約0.2cm～0.4cmである。明治の牛乳ボックスの結果からは、異なる距離で撮影された画像をデータセットとして使用した場合でも、小さな物体の検出は弱いままであることが分かる。小さな物体を検出したい場合や、検出を広い範囲で適用できるようにしたい場合は、データセットに多数の小さな物体の画像を追加する必要がある。森永と北海道は、遠距離の画像データが比較的不足しているため、認識率は近い範囲（1～90番目の画像、すなわち7cmから60cmほどの距離内）でより高くなる。この状況は視覚障害者が対象物を手に持った状況での検出に使用でき、背景のより遠い物体を効果的に無視できるといえる。

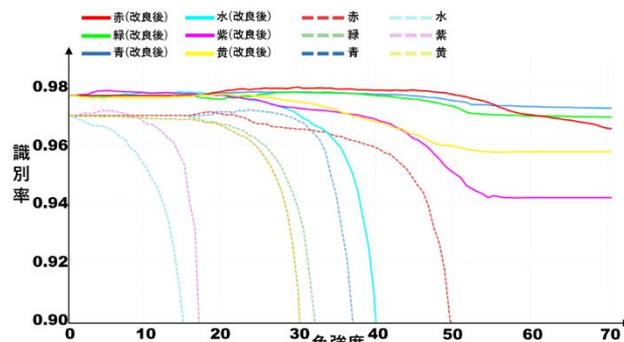


図2 色強度を変化させた場合の識別率

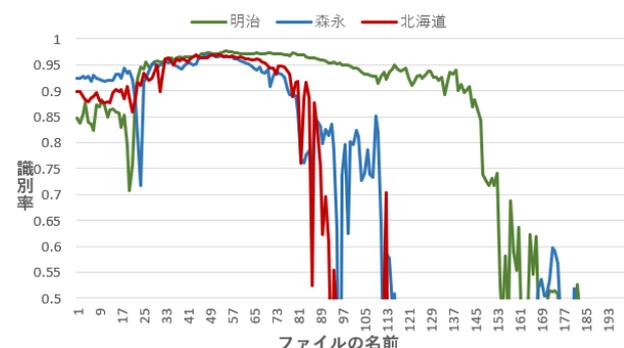


図3 距離に対する識別率

参考文献

- [1] GBD 2019 Blindness and Vision Impairment Collaborators, "Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study", 2020.
- [2] G. Jocher, <https://github.com/ultralytics/yolov5>