

TERASU MAP

～ 視覚障害者の方の危険箇所マッピング ～

0304・群馬県立高崎高等学校 3年 高田悠希



研究の概要

本研究では、Image to Text AIを利用して、白杖デバイスで撮った画像から①現在の状況の説明を生成・音声通知し、さらに②自動で視覚障害者にとっての危険箇所をマッピングするというシステムを実装した。

視覚障害者の方の事故のリスクは自分の住む群馬県のような、インフラ整備が不十分な地域で特に大きい。その問題を解決するため、私は2年前からAIを搭載した白杖の研究を続けてきた[1]。このAI白杖は、独自のデータセットに基づいたAIにより、線路や横断歩道、歩行者や自転車など多くの危険を、白杖の視点からリアルタイムに検知することを可能としている。

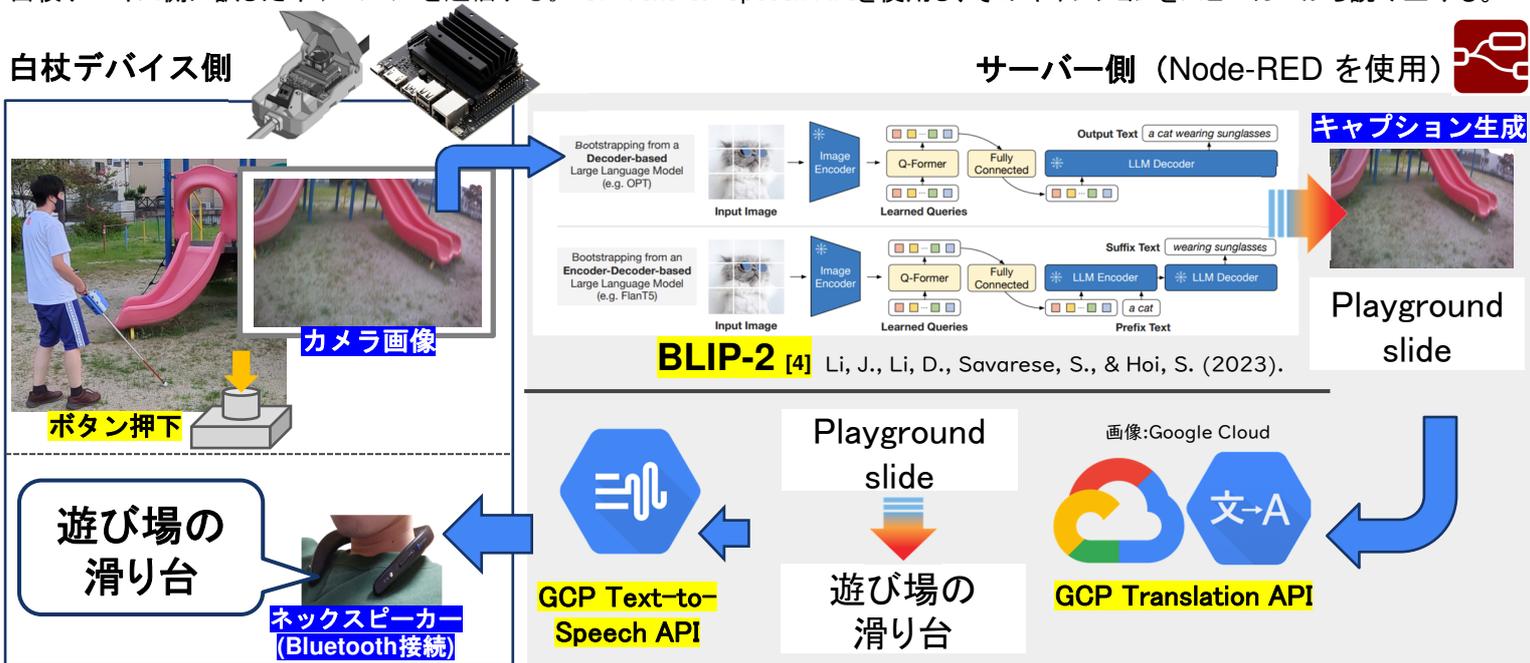
本研究はこのAI白杖の機能を向上させるとともに、インフラ整備のような新たな場面での活躍を見込むものである。



視覚障害者のモビリティ向上を目的とした電子白杖の開発事例はいくつか存在する[2][3]が、線路や横断歩道のリアルタイム検出や Image to text AIを用いたジェネレーティブな現在状況の説明、自動マッピング機能を実現したものはいずれも存在なく、本研究によるシステムは、視覚障害者の方への新たな支援の形として、大きな革新をもたらす可能性を秘めている。

① Image to text AI による状況説明機能の開発 - システム構成

現在の状況を説明するImage to text AIとして、事前学習済みの画像エンコーダー及び大規模言語モデル(LLM)に接続した「BLIP-2」[4]のAPIを使用する。ユーザーがボタンを押下すると、Jetson Nanoを搭載した白杖デバイスがカメラ画像とGPS座標をNode Redサーバーにアップロードし、画像をBLIP-2にキャプションさせる。BLIP-2からの応答は英語であるのでGCP Translation APIを用いて日本語訳し、白杖デバイス側に訳したキャプションを送信する。GCP Text-to-Speech APIを使用し、そのキャプションをスピーカーから読み上げる。



街中での使用結果

上記のシステムを自作の白杖デバイス内に組み込み、Image to text AIを用いた状況説明機能を実際に街中の場面で使用した。以下に使用風景と読み上げられたキャプション結果の一部を示す。



①②のように、歩道脇にある植物や自動販売機、もしくは散歩している犬などを説明することができた。また、③④のように、視覚障害者の方が何なのかを特定しにくいポールやゴミ袋などの説明も可能。さらに、⑤のように川といった危険な場所も説明可能であり、楽しさの向上と安全面の両方で有用性を示した。

リアルタイム危険検知と組み合わせたモビリティ向上



現在状況の随時通知
(サーバー上でAI処理)

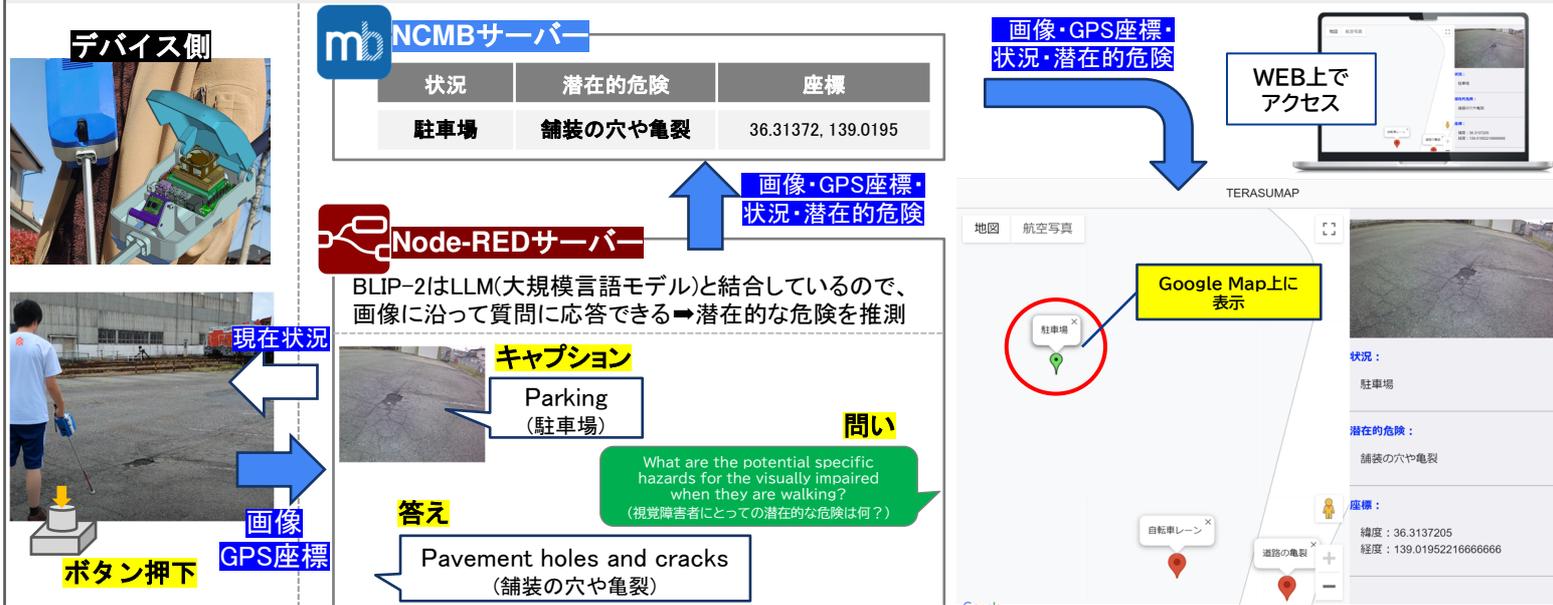


危険のリアルタイム検出
(ローカルでAI処理)



処理をサーバー上で行うため、ローカルで動作する画像認識AI(昨年が開発[1])と併用が可能
組み合わせることでより安心で、そして**楽しい**歩行が可能になる

② 自動マッピングシステムの構築 - システム構成



WEBアプリ上での動作



コーン(ポール)やごみ袋のような、歩行中の視覚障害者の方にとっての危険を示せる

今後の活用



展望 - 互いに照らしあう社会



参考文献: [1] 高田悠希(Haruki Takada), Michishirube: AI Assist Cane for Visually Impaired, 国際学生科学技術フェア(ISEF)2023 (Regeneration International Science and Engineering Fair 2023), <https://projectboard.world/isef/project/robo023-michishirube-ai-assist-cane-for-visually-impaired>, 2023年発表
 [2] 「Wewalk Smart Cane」We Walk社(<https://wewalk.io/en/>), '23年5月閲覧
 [3] 吉岡学、中川浩文、諸菱正典、保高拓哉、「RFID技術によるスマート白杖を用いた視覚障害者用車輪式ホーム検出システムの評価」、2022年発表、https://www.istage.jst.go.jp/article/jie/58/3/58_114/pdf/-char/jia
 [4] Li, J., Li, D., Savarese, S., & Hoi, S. (2023). BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. ArXiv preprint arXiv:2301.12597.