

視覚障害者支援は最新の ICT や AI 技術を必要としている -自動運転で開発された AI 技術から学ぼう-

釜江常好^{†1,2} 小出富夫^{†3} 野口哲夫^{†3}

概要: 情報技術で視覚障害者の支援を強化するには、自動運転の実用化で開発されたプログラムを参考にすべきと考える。「可能性を示す」だけでは実用化と考えられない厳しい環境でテストされたコードこそ、参考になる。例えば OCR は、カメラと書類を完全に位置合わせしなくても、照明に斑があっても、正しく読まねばならない。物体認識では、テーブル上に置いた本やパッケージ商品を認識しても、スーパーマーケットの棚にある商品は認識できないようでは、使えない。自動運転などで使われる AI 技術を見ると、瞬時に信号機を認識している。もう少し工夫すると、看板の文字を読み、棚の商品を認識し、通りの向かいにあるバス停や地下鉄への入り口も認識できるようにできるだろう。かなりの資源が必要となるかも知れない。しかし、このようなテーマは、新しいビジネスを立ち上げる訓練になる点で、大学院生には、魅力あるテーマとなるだろう。幅広い協力を得て、実用に耐える物体認識システムを作りたい。

キーワード: 視覚障害者支援、OCR、音声認識、物体認識、AI 技術、ビーコンの応用

Accessibility to visually handicapped calls for forefront ICT and AI - Lessons from AI-based autonomous driving programs -

TSUNEYOSHI KAMAE^{†1,2} TOMIO KOIDE^{†3}
TETSUO NOGUCHI^{†3}

Abstract: To advance accessibility for the visually handicapped with ICT, we should learn from AI-based autonomous driving projects. What considered as “existence proof” do not warrant real applications in the current assistive applications. For example, OCR needs to capture printed characters without camera alignment nor uniform illumination. Object recognition identifies books and packaged merchandise on a table but not on a shelf in a super market. AI-based automated driving technology can recognize traffic lights in real environments. We will be able to recognize signages, bus stops, and entrances to subway stations by learning this AI-based technology. Some adequate resources will be needed. However, these themes will be very attractive AI research topics and training grounds for the ambitious students and researchers looking for new business opportunities. We propose to form a broad coalition across disciplines and produce systems applicable to the real environments the visually impaired face.

Keywords: Assistive technology, OCR, Voice recognition, Object recognition, AI technology, Application of BLE-Beacon

1. はじめに

私たちは、長年 Android ベースのユニバーサル・アクセス携帯の開発し、視覚障害者に特化した、音声ベースで動くスマートフォンやタブレットを開発してきた。そして、障害者に、「技術はここまで進んでいます。試してみませんか」、と言えるものを頒布するとともに、講演やチュートリアルを開いてきた。また最近、DAISY 規格（視覚障害者向けの音声図書）で、アンドロイド携帯を使うための、マニュアルを制作し提供している。詳しくは、<http://universal-access.jp> を見て欲しい。この努力の中で、商品認識は、テーブルの上ではできて、スーパーの棚の中では不可能であること、OCR もカメラの位置を合わせて初めて精度が上がることなど、限界を感じている。一方、自動運転で AI 利用の技術を誇示する YouTube などを見ると、即時に信号機を認識して事を知る。これらを、資源が不足

している我々にも提供されれば、飛躍的な改良が見込まれると考えるに至った。協力の範囲を広め、最新の自動運転で実用化されつつある AI 技術を、利用できるようにしたいと願っている。

2. ユニバーサル・アクセス携帯 (UA 携帯) で使われている最新 ICT の紹介

UA 携帯では、通常の携帯機器が提供する機能に加えて、「技術はここまで進んでいます。試してみませんか」、と言えるアプリを開発し、使えるようにしている。その例は、音声とキーボードを融合した入力方式 (図 1)、音声コマンドの活用、書籍やパッケージ商品の認識、日本語 OCR 技術の応用、NFC の活用 (図 2) などである。

この他に、ブザー付きビーコンとスマホの組み合わせで障害者同士が待ち合わせたり、市民会館などへのナビを支援するシステムも開発した。

†1 東京大学 (名誉教授)
University of Tokyo (Emeritus)

†2 スタンフォード大学 (名誉教授)
Stanford University, SLAC/KIPAC (Emeritus)

†3 クリエイトシステム開発株式会社
Create System Development KK

持続的に障害者の携帯利用を支援することも課題の一つである。遠隔ミラーリングで携帯の画面を見て支援できる体制も整えた。DAISY 規格のマニュアルも用意している。



図 1 音声とキーボードを融合した入力方式

モードボタンで、音声入力とキーボードが切り替わる

Figure 1 Voice-keyboard combined input system.

The two modes toggle each other by the “Mode” button



図 2 切込みが付いた NFC タグと専用の設定アプリ

Figure 2 NFC tag with notches and setting for various setting

最近、力を注いでいるのは、物体認識である。UA 携帯には、「カメラトーク」と命名したアプリが実装されている。図 3 で示すように、日本で出版された本やパッケージ化して売られている商品は、単独で一様な照明の下に置くと、ほぼ完全に認識する。この例では、Docomo の商品認識 SDK を使っている。

この他に、QR コードを読んだり、QR コードを名刺に印刷することで、読むと「連絡帳」に入力される仕組みも、含まれている。さらに、アプリ開発者の協力を得られた場合、アプリのパラメータを渡し、入力の手間を軽減する試みも、導入している。ぜひ試して欲しい。



図 3 Docomo の商品認識 SDK を使ったカメラトーク

Figure 3 Recognition by “Camera talk” on Docomo’s SDK

我々のカメラトークの特徴は、自動的に焦点を合わせながら、物体を認識するまで試行を繰り返すことである。すなわち、障害者には難しい物体とのアラインメントに寛容で、ゆっくりカメラを動かすと、適当なところで認識できたと音声で教えてくれる。あとで述べる、グーグルの物体検出では、写真を撮り、それを決まった Web ページにドロップすると認識が始まる。位置合わせができないと、try-and-error を何十回も繰り返すことになる。

我々が実装しているアプリは、かなりの使ってもらっているが、それでも使い易いものではない。その理由を、以下に列挙する。

- OCR は、照明が一様でない場合や、ミスアラインしている場合には、文字認識の確度が、大幅に下る。
- 文字検出も、背景がある街中では極めて難しい。
- 道路脇の信号機、バス停などは認識が困難

この状況を打開するため、自動運転、ロボットなどで AI 技術を駆使している研究者の力を借りたい。

3. グーグルの物体検出サービスを試す (1/2)

Google の検索窓に追加する物体検出で、街をあるきながら撮影した画像をテストした。その結果から、何を、どのように改良すべきか見えてくる。

簡単に成功するだろうと考えたものから、検出から始めた。横断歩道のゼブラゾーンは、多くの人が歩いている中でも、検出できた。しかし、雨覆がついた東京都バスの停留所でも、かなり近づかないと、検出できない。通りの反対側にあると、建物にまぎれて、認識しない (図 4)。



図4 横断歩道（左：成功）とバス停（中：成功、右失敗）
Figure 4 Zebra zone (left:OK), bus stop (cntr:OK, right:Fail)

簡単に認識できると考えたが、完全に失敗するのが、切符売り場と改札である（図5）。データベースが整備されていないのではないかと考える。現状では、「機械」として認識される。JR のロゴがはっきり読み取れる駅でも、「駅入り口」と認識されない。KIOSK も同様である。これもデータベースが整備されていないのではないかと。

後で述べるように、数万項目程度のデータベースを用意し、GPS の位置を組み合わせれば、簡単に「千駄ヶ谷駅」や「千駄ヶ谷駅改札」と認識できるはずである。



図5 切符売り場と自動改札：ともに認識失敗
Figure 5 Ticket vendor and ticket gate: failure

4. 自動運転技術で見る信号検出

自動運転に関係した YouTube を見ると、リアルタイムに信号機を検知し、その色を正しく認識していることが判る。自動車にどの程度の CPU が搭載され、いくつのカメラが使われているか、はっきりしない。その中で、Ricoh の ThetaS で撮影しながら、歩行者用の信号をリアルタイムで検出し、赤か青かを認識している YouTube が加藤真平博士から投稿されていたので[文献 2]、そのクリップを利用して、性能を紹介する。

4.1 加藤真平博士のグループの成果から

かなり離れたところにある、歩行者信号を、**Ricoh ThetaS**で検出

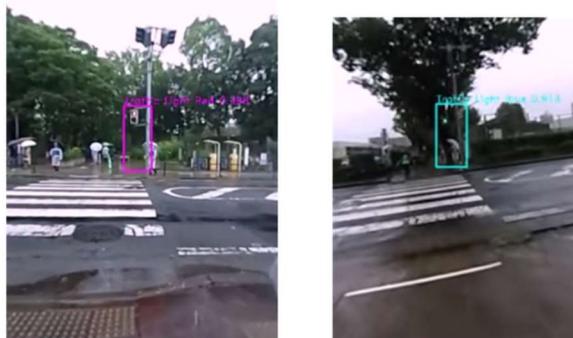


図6 加藤真平氏の YouTube[1]から：Ricoh ThetaS でライブに歩行者信号を検出。Faster R-CNN アルゴリズムを利用。
Figure 6 From YouTube by Shinpei Kato[1]: Pedestrian traffic light is recognized real-time with Ricoh ThetaS and algorithm known as Faster R-CNN

ちなみに Ricoh の ThetaS は、画素数は、最新のスマホ程度で、遥かに広い視野をカバーしている。画素密度では、スマホより一桁程度悪い。使われているアルゴリズムを調べてみた。

4.2 物体認識アルゴリズム、Faster R-CNN

Faster R-CNN について調べてみた。これは 2015 年ごろから使われ始めたアルゴリズムで、クラウド上の GPU サーバーで稼働しているようだ。論文 1 によると、図 7 に示す例の場合、200ms 程度で認識している。すなわち十分実用になるスピードである。もとは、Matlab で書かれているようだが、Python に移植されている。

このアルゴリズムを実装し検証した論文[2]には、筆頭著者はマイクロソフトにインターンとして働いていた期間に、図 7 に示すようなレベルまで完成させたと記されている。やる気がある学生なら、卒業研究で十分やれるはず。修士課程の大学院生なら、歩行者信号やバス停のサンプルを集め、プログラムを十分トレーニングすることも可能だろう。

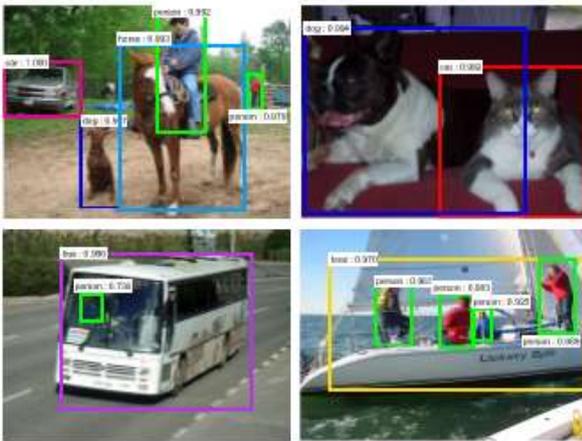


図7 参考文献[2]の図3を転載。約200ミリ秒で認識。
Figure 7 Taken from Figure3 of [2]. Recognition in ~200ms.

5. グーグルの物体検出サービスを試す (2/2)

グーグルの物体検出が、比較的簡単だと思う物体が検出できなかったことから、テンプレートあるいはデータの充実度に大きく依存している可能性を疑うことになった。逆に、データが充実していれば、高い確度で認識できる例を示したい。その例が、歴史的建造物と芸術作品である。図8の左に示す三重の塔が、どこのものか判る人は少ないと思うが、グーグルの画像検索は、見事に飛騨国分寺の塔であることを言い当てた。また絵画や版画も、近代のものまで、データベース化され、図8の右の作品でも、認識する。これらの背景には、各国が文化遺産の映像をデータベース化し、世界の美術館が所蔵品を写真データベース化し、公開していることがある。

すなわち、日本にある全ての駅と改札口を数か所から撮影し、データベース化すれば、文化遺産程度に認識できると考える。3-4万の写真で済む話である。バス停留所は、会社ごとに似ているので、数千枚で済むだろう。各駅、各駅の入口のGPSのデータと照合すれば、確度を上げることは難しくないだろう。全国のボランティアに協力を依頼すれば、達成できそうだ。

6. 視覚障害者へのプログラミング教育

近い将来、取り組まなければ大きな課題は、大学や大学院で、視覚障害者に対する、プログラミング教育である。スクリプト言語を選ぶことになると考え、PythonやJupyter Notebookの音声化への試みを始めているところである。この課題でも、情報学を学ぶ学生には、音読(TTS)や音声認識を実践する良い機会となる。Windows上のPythonでは、SAPIのTTSが使えるし、Linux上では、eSpeakなどが使える。音声認識は、クラウド上のグーグル音声認識を使うJavascriptが用意されている。

ここでも、分野間の交流を蜜にすることで、多くの魅力

あるプロジェクトを提案すべきと考えている。



図7 グーグル画像検索が正しく認識した、飛騨国分寺の三重の塔(左)と、川瀬巴水の版画「池上本門寺」
Figure 7 Tower in Hida Kokubunji temple (left) and wood-block print by Kawase Hasui (right), both correctly identified by Google Photo Search

7. 終わりに

多くの課題が残るが、何とか見通が立てられるようになってきた。しかし限られた資源の中で、具体的な成果を得るには、以下のような施策が必要だろう。

- 学生に、最新情報技術、特にAIを実践するチャンスと捉え、論文のテーマとして取り組んでもらう。
- 情報処理学会のAI分科会に協力してもらい学生の指導への参加を要請する。
- どこかのクラウド上で、Faster R-CNNなどのアルゴリズムを試せる場所を確保する必要がある。
- GoogleやDocomoなどの技術者にも、参加を促すことも、重要となるだろう。
- 国、地方自治体、民間の資金を確保したい。

謝辞

情報処理学会のアクセシビリティ分科会には、このような機会を与えてもらったことを感謝します。

参考文献

- [1] <https://www.youtube.com/watch?v=OblbhaRwA7o>
<https://www.youtube.com/watch?v=1Tn1Oin5q1Q>
- [2] S. Ren, K. He, R. Girshick, J. Sun;
arXiv:1506.01497v3[cs.CV]