

機械学習を用いた画像翻訳システムの構築

The Constructing of Image translation system with machine learning

陳 一帆¹⁾

指導教員 亀田 弘之¹⁾

1) 東京工科大学大学院 バイオ・情報メディア研究科 コンピュータサイエンス専攻 亀田研究室

キーワード：翻訳, 文字認識, マシンラーニング

1. 研究背景

今 21 世紀 20 年代に入り、国際コミュニケーションが増えてきた。外国語を使わなければならない機会が多くなった。さらに、AR 技術やカメラの発達により、誰でもカメラ持ちになり、どこでも写真を撮ることができるになった、本研究では、それを利用し、外国語の勉強が苦手な人の助けとなる、画像中の文字を翻訳できる機能を持つ言語学習サポートシステムを提案した。

現在、画像中文字翻訳技術を使い、スマートフォンで実装している電子辞書アプリもある。ある程度の画像文字翻訳ができるが、ほぼ目立たないところに隠していて、このような便利な機能があるの知らない人が多い。それはそれに関する技術はまだ実用化されていないからである。実際使ってみると、完全正しい認識していない、翻訳が間違っているのが多い。現在、文字認識、機械翻訳などの技術を研究しているのはグーグルなど大手企業しかない。そして、文字により、国々それぞれの文字の認識方法は完全に一致するわけではない。現存している画像文字翻訳システムは多くの制限があり、翻訳文字数制限、或いは有料になっている。さらに、その二つを組み合わせ、言語学習などで利用し、使いやすいシステムがほぼないのが現状である。

2. 目的

本研究は、主に中国語を学ぶ人や日本語を学ぶ

人のために学習サポートシステムとして構築するものであり、普及している電子辞書のような画像中の文字を認識して翻訳できる。誰でも手軽に使うシステムを構築し、みんなの生活に入る、実用化するのが目的である。

3. 実現方法

画像翻訳システムは画像の中の文字を翻訳する機能を実現するために、二つの部分がある。

- ・文字認識
- ・文字翻訳

具体的な手順まずは画像の取得。画像の取得は本や新聞など、自分の目で見たすべての文字、カメラで写真を撮るあるいはスクリーンショットで得たもの。そして、画像中にある文字を画像から抽出し、認識する。画像中の文字を認識するためには、その前に画像に文字にある区域を確定しなければならない。そのために、文字の検出が先である。検出した文字を認識して、最後に得た文字を翻訳することで、画像翻訳を実現する。

文字を認識する前に、画像のどこか文字であるかを確定する必要がある。まずは文字検出である。画像は画素で構成されているものである。文字検出はすべての画素から文字がある画素のみ切り出す仕組みである。文字がある画素を四角形の型でスキャンする。それは一行ずつで、文字を検出すること。文字がある区域を確定すれば、また文字を一個ずつでさらに小さい四角形で分けをする。それ

で文字検出を完成である。



図1 文字検出と認識

文字検出してから、文字が一個ずつで分けて認識する。認識方法はマシンラーニングを使い、システム自身で文字の特徴を探して、その特徴から文字を認識するである。認識不能の文字マシンラーニングによりシステムを訓練して、認識の正解率を高めることは可能であり、適度の訓練をすれば、文字認識の正解率はさらに上がる。実験環境はPython3.8+Tesseractで行うとなる。そして最後に、認識した文字を機械翻訳システムに送り出し、翻訳した文字をユーザーに表示する。翻訳したテキストは一つ一つの単語で調べられる。ユーザーは新しい学んだ単語をその形で覚える。

4. 画像処理

現在の方法で、複雑な画像から文字を正しく検出して、認識する正解率は60%しかない、その精度を高めるためには、画像の処理が必要である。本研究では、さらに認識正解率を高めるための追加した画像処理を提案した。まずは画像の取得で行うべき仕組み。それは、カメラのパラメータの調整である。カメラのパラメータを正しい認識しやすい数値に調整する。ISO、カメラの感光度、画面全体の輝度にコントロールできる。一般的の写真は画面全体の輝度が最優先に対して、文字のみをできる限りはっきり見える数値にすること。そして、フォーカスの調整。一般的の写真は人間の顔や一番近い物体にフォーカスする。調整で文字がある区域にフォーカスする。そしてホワイトバランスの調整。画面の色を文字が見やすい状態に調整する。

画像の取得から、画像の前処理を行う。二値化、輝度値、トリミング、ノイズ除去、角度、ピクセルなどある程度に調整すれば文字の検出と認識には

よい影響が与えられと考えられる。研究は現時点2020年10月では、一部の前処理が文字の検出と認識に与えられ影響はまだ検討中である。

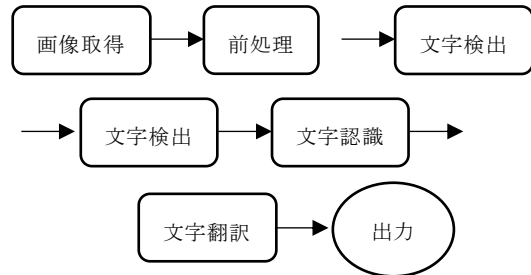


図2 提案システムの処理手順

5. 今後の課題と展望

いままでは翻訳の正解率の中心として、文字検出の成功率を上げるために、提案した方法の一部を一つずつで実行してみた結果、確かに最終の翻訳正解率により影響があることが明らかにした。しかし、画像処理の時間も共に伸ばしてしまった。そして、提案した前処理それぞれの順番と処理方法は最終の処理速度に関わると考えている。これからはそのバランスを調整することが必要。

そしてほかの機能について、誰でも使えることこそよいシステムだと思っている。そのために、色々便利な機能を考えている。メイン機能に基づいて、認識した文字を音声で流すと、目の不自由な人にも、見えない文字がわかるようになる。

これからのスマートフォンはダブルレンズ、さらに三つのレンズが載せているのが多いと思っている。多数のレンズがあれば、カメラは空間を認識することができる。レンズの角度差により、翻訳目標としての文字はさらに認識しやすいと考えている。そして、AR技術の発展により、スマートフォンでも簡単なARアプリを実行できる。それを利用して、この翻訳システムはさらに実用化できると考えている。