

Google の VisionAPI を用いた AR 教育コンテンツの作成と 学習評価に関する考察

富山県立大学 工学部 電子・情報工学科
1415048 福嶋瑞希

指導教員：奥原浩之

1 はじめに

自身の研究テーマは数日前まで「ドローンからの映像を 360 度で確認できるバーチャルリアリティの作成、人間をマーカーとする拡張現実システムと組み合わせた気づき支援のための 3D-AR システムの開発」であった。しかし何度もトライアンドエラーを繰り返すたびに実現が難しいことが判明した。AR の性質上 VR 空間に AR を出すことが困難であることが判明したため、テーマの方向を変えることとした。なお、この過程で Android, MOVERIO 共に簡単な AR アプリの作成 (Vuforia を用いてマーカー認識 AR) と, theta のリアルタイム VR についての技術を得た。ただ、EPSON に問い合わせたところ Vuforia は MOVERIO に対応していない (Vuforia の SDK に x86 用のネイティブライブラリが含まれないため), 従って Wikitude での AR 作成を検討している。

2 新たな研究テーマ

2.1 研究テーマの決定

本研究では、Google の VisionAPI を用いた AR 教育コンテンツの作成を研究テーマとする。HMD で目の前のものに手をかざすと、その物体の英語音声が聞くことができるコンテンツの作成と学習速度の比較を行う。

2.2 システム全体の流れ

- 1, 手の甲に AR マーカーを付着させ, HMD のカメラで認識して手の位置を把握
- 2, 手の横にオブジェクトを表示
- 3, オブジェクトが表示されている画像領域を切り取り Vision API に渡す
- 4, 結果を音声で返す

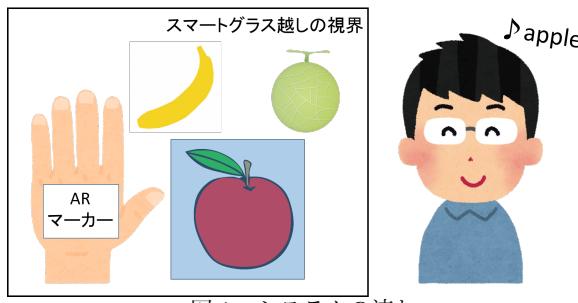


図 1 システムの流れ

3 システム関連事項

3.1 音声での英語学習

英語でのコミュニケーションスキルは現代において非常に重要な技術である。ビジネスやアカデミックなどの様々なシーンでグローバル化が急進展する今日、様々な国籍の人々とディスカッションや協同作業を行う機会が増加している。このような、異なるバックグラウンドを持つ人々とのコミュニケーションの技術として、英語は不可欠であると言える。このため、教育の現場では英語コミュニケーション学習のための様々なプランが検討されている。しかしながら、英語の習得は依然として容易ではない。前述したように英語教育において様々な施策が行われているにもかかわらず日本人の英語能力は低い。高校 3 年生を対象にした文部科学省の調査によると、その多くが CEFR における A1 A2 という初学者のレベルに収まっており、特に話す能力と書く能力が低い [1]。話す能力と書く能力とは、つまり自分の伝えたいことを英語で表現する能力である。英語を話すことへの不安であったり単語力の低さが挙げられると考えた。

3.2 Google の VisionAPI

Google の VisionAPI とは、Google Cloud Platform(GCP) が提供する Google の機械学習モデルを使用した画像認識サービスのことである。このサービスを利用することで、Google の持つ画像に関する機械学習モ

デルを使い、対象となる画像から様々な情報を取得することができる。取得できる情報の例として、猫やヨットなど、画像に含まれている物体の検知、画像内に含まれている文章やロゴの取得、人物の顔認識やその表情が示す感情に関する情報の取得、コンテンツがアダルト、暴力的など不適切な要素を持つかの判定等である。また、API とは自己のソフトウェアを一部公開して、他のソフトウェアと機能を共有できるようにしたものである。また、機械学習 (machine learning) とは、「機械にたくさんのデータを読み込ませ、解析することでアルゴリズムを作ること」である。最大の長所は、人の手でデータを分類するよりも早く、正確に、大量にさばくことができることであり、新しいテクノロジーだと思われがちだが、1950 年代にはすでに登場している技術である [2]。

3.3 MOVERIO を使用した MR・AR 環境

シースルーモバイルビューアー MOVERIO は HMD の一種である。HMD (ヘッドマウントディスプレイ) は頭部に装着する表示装置であり、特に密閉型の HMD はその没入感から VR (Virtual Reality) の代表的なデバイスとして知られている。両眼に対応したものはステレオ視による立体表示が可能である。Robinett らは HMD における自然な 3D 表示のための表示ゆがみやステレオ視のモデル化を行っている [3]。HMD は手持の必要がなく、顔の向きや体勢によらずつねに目の前に仮想画面を表示可能であることから、作業支援のウェアラブルデバイスとしての利用も數多く研究されている。MOVERIO はユーザが見ている現実空間に対してコンピュータが生成した仮想オブジェクトを重畳表示する、いわゆるシースルーディスプレイを行なう。シースルーディスプレイの実現にはビデオシースルーワーク方式と光学シースルーワークの 2 通りがあり、それぞれ長所と短所を持っている。MOVERIO は光学シースルーワークである。密閉型の HMD を用いてカメラ画像と CG を合成した画像を表示するビデオシースルーワークでは、正確な合成ができる反面、ユーザは直接外界を見ることはできず、視界の制限や表示遅延のため安全に使用できる場所は限られる。これに対し、光学シースルーワークは肉眼の視界に対して CG を重畳する方式であり、視界が広く、現実視界部分には表示遅延がない。この性質は移動中の使用や現実の物体を用いた作業時の使用に適している。一方で、肉眼の視界とカメラの視界にはずれがあるため、CG の表示位置とユーザの視界を合わせるために表示座標を変換しなければならず、カメラ取得画像の座標へ直接合成すればよいビデオシースルーワークよりも難しいと考えられている。

4 今後の展開

システムの流れ 3,4 である「オブジェクトが表示されている画像領域を切り取り Vision API に渡す」と「結果を音声で返す」が難関であると考える。カメラの位置からオブジェクトとの距離の調整、API から返答される言語を音声にすることが現在の難関であるため、まずは Wikitude での AR 作成から作業開始をする。同時にこのシステムに付随できるコンテンツの検討を必要とする。

参考文献

- [1] 文部科学省：平成 26 年度 英語教育改善のための英語力調査事業報告 (2015), http://www.mext.go.jp/a_menu/kokusai/gaikokugo/1358258.htm (2017.11.16 確認).
- [2] 小高知宏 (2011)『はじめての機械学習』株式会社 オーム社.
- [3] Robinett, W. and Rolland, J.: A computational model for stereoscopic optics of a head-mounted display, Presence: Teleoperators Virtual Environ, Vol.1, No.1, pp.4562, MIT press (1992).
- [4] 北山 史朗 (2017)「聴覚遅延フィードバックを用いた英会話学習支援手法の研究」,[online]<https://dspace.jaist.ac.jp/dspace/bitstream/10119/14132/3/paper.pdf>(2017.11.16 確認).
- [5] 入江英嗣ほか (2014)「AirTarget : 光学シースルーワーク HMD とマーカレス画像認識による高可搬性実世界志向インタフェース」,『情報処理学会論文誌』50(4), p.14151427, 情報処理学会.