

現実空間への漫画エフェクト投影システムの提案

小林 龍成¹ 田代 琉人¹ 長谷川 稜馬¹ 久野 文菜² 谷口 航平² 濱川 礼¹

概要: 本論文では、現実空間への漫画エフェクトを投影するシステムについて述べる。近年では Web 技術の発展やデバイスの普及により、電子漫画化が進んでいる。電子漫画のほかにも「漫画動画」や「VR 漫画」等の新しいメディアで提供されるようになり漫画の読み方が多様化している。また、漫画を利用したコンテンツも数多く登場しており、キャラクターとして漫画世界に入り込むことが可能になっている。しかし、それらのコンテンツは現実の一瞬を漫画として出力しているだけであり、常に漫画世界に浸ることは出来ない。そこで、現実世界に漫画のような効果表現を用いて、人が見る日常の風景全てが漫画に見える世界を目指す。具体的には、漫画の構成要素である、イラスト・吹き出し・オノマトペの表示により、漫画のような効果を表現する。イラストは画像処理、吹き出しは音声処理、オノマトペは機械学習を用いて実装し、現実空間と仮想空間を複合できる MR デバイスを用いて表示する。

1. はじめに

漫画は、字義的に「気の向くままに漫然と描いた画」という意味であり、語源は随筆を意味する漢語「漫筆」が「漫筆画」を経て「漫画」になったという説がある [1]。「漫画」が普及した現在では技術の発展に伴い、読み方や楽しみ方が多様化している。現在の漫画について以下に述べる。

1.1 現在の漫画

21 世紀に入り、Web 技術の発展及びスマートフォン・タブレット等のデバイスの普及により広まってきたのが「電子漫画」である。電子漫画の販売金額は年々増加しており、2017 年には紙の漫画の販売金額を上回った [2]。定額で漫画が読み放題となるアプリやサービスなども登場し、電子漫画は著しく成長している。また、電子漫画だけでなく、「漫画動画」「VR 漫画」等の新たなメディアを通して提供される漫画も増えてきた。漫画動画は、声や効果音、BGM とともに漫画をスライド形式で流した動画のことであり、VR 漫画は、VR で見る漫画のことで、360°漫画世界が描かれており、漫画世界に入り込むことができる。技術の発展により漫画の楽しみ方が多様化してきている。

1.2 漫画を利用したコンテンツ

一方で、漫画を利用した新しいコンテンツも製作されている。例えば、人気漫画「ジョジョの奇妙な冒険」[3] を題材にした『ジョジョの奇妙なスタンド体験』や、アート展

「マンガ マッピング！」[4] で公開された『onomatopée』である。『ジョジョの奇妙なスタンド体験』では、AR 技術を用いて作中に登場する“スタンド”の能力を体感することができる。『onomatopée』では、自分が発した音に応じて 5 種類のオノマトペがスクリーンへと映し出される。他にもスマートフォンアプリには『奇妙な写真メーカー』[5] や『漫画コミックカメラ』[6] など多数のシステムがある。

1.3 目的

1.2 で述べた様々なコンテンツでは漫画の登場人物として、一時的に漫画世界へ入り込むことができる。しかし現実の刹那を漫画として出力しているだけであり、常に漫画世界に浸ることはできない。我々は漫画内の登場人物として漫画世界に入り込むのではなく、現実世界そのものを漫画世界へと変換することで、漫画が日常に取り入れられ、面白い世界になるのではないかと考えた。

漫画ではキャラクターの言動に合わせて様々な効果が描かれる。キャラクターが話せば吹き出しが、動けば効果音が表示され、状況を表すオノマトペなども表示される。このような漫画の効果を現実世界に表現し、日常風景を漫画世界にするエンタテインメントシステム「セカマン (世界は漫画でできている)」について述べる。

2. 関連研究・関連システム

漫画を題材にした研究や開発はこれまでに多く行われており、様々なシステムが存在する。

Manga Generator[7] では、ユーザが漫画のキャラクターとなって、画面に映るシーンに合わせてポーズをとること

¹ 中京大学 工学部

² 中京大学大学院 工学研究科 情報工学専攻

でオリジナルマンガが生成される。赤外線カメラを用いてユーザの姿勢から感情を推定し、感情に応じて効果線を変更する。また、姿勢に対してレンジ分けを行うことで、段階に応じたマンガ効果を描画する。

Comixify[8] は動画から自動で漫画を生成する。「動画の自動要約」と「画像を漫画風に変換」の2段階の工程からなっている。「動画の自動要約」では、教師なし強化学習のDSN (Deep Sumarization Network) を用いてキーフレーム抽出を行う。一方「画像を漫画風に変換」ではGAN (Generative Adversarial Network) を使用して元画像から漫画風の画像に変換している。2つの工程を通すことで、任意の動画を簡単に漫画へ変換することができる。

漫画カメラ[9]は、写真を漫画風に加工するスマートフォンアプリである。人物や物体だけを切り抜き、漫画風に画像処理を行う。オノマトペや効果線が描かれているフレームが20種類以上あり、様々な漫画風の写真が作成できる。

本システムでは、漫画風の変換、吹き出し・オノマトペの表示をリアルタイムに自動で行う。そのため既存の研究やシステムとは違い、複数の漫画要素を取り入れた変換が可能であり、ユーザは常に漫画世界に浸ることができる。

3. システム概要

我々は現実世界を漫画世界にするシステム「セカマン」を開発した。漫画のコマは、キャラクターやセリフ、吹き出し、オノマトペ、効果線などから構成される[10]。オノマトペとは音や状態を表す擬音語・擬態語の総称であり、漫画では情景や雰囲気表現するために用いられる。住山[11]は漫画の主な構成要素として「イラスト」「吹き出し」「形喩・音喩」(オノマトペ)をあげている。そこで我々はイラスト、吹き出し、オノマトペの3つを漫画の構成要素と定義し、日常の風景に「漫画風イラスト(図1左)」、「会話内容が表示される吹き出し(図1中)」、「見ている情景に沿ったオノマトペ(図1右)」を表示することで、漫画としての印象を与えられるのではないかと考えた。それぞれ画像処理、音声処理、機械学習を用いる。機械学習は、漫画のようなオノマトペを自動で付与するために行う。漫画のコマの画像とオノマトペの対を学習させることで、画像からオノマトペを付与できると考えたからである。表示には現実空間と仮想空間を重畳表示できるMRデバイスを用いる。



図1: 表示する漫画の構成要素

4. システム構成

システム構成図を図2に示す。

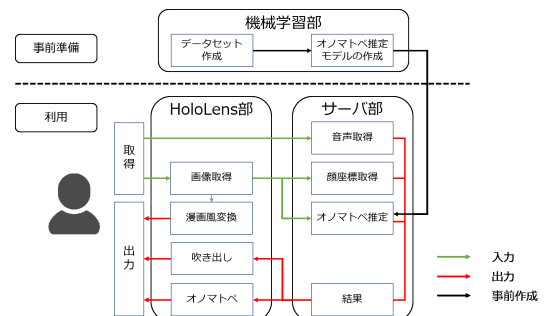


図2: システム構成図

「セカマン」は、機械学習部とサーバ部、HoloLens部からなる。機械学習部では、事前準備としてデータセットの作成、オノマトペ推定モデルの作成を行う。サーバ部では、音声取得、オノマトペの推定、顔の座標取得を行う。オノマトペの推定は機械学習部で作成したモデルを使用する。HoloLens部では、画像取得、画像の漫画風変換、吹き出しやオノマトペの出力を行う。サーバは、Python用のマイクロフレームワークであるFlask[12]を用いて構築した。HoloLensとサーバ間の通信では、HoloLens側は取得した画像を送信し、サーバ側は処理後の結果を送信する。

4.1 機械学習部

機械学習部では、画像を入力とし、オノマトペを出力とするモデルを作成する。

4.1.1 データセット作成

機械学習を行うためにはデータセットが必要であるため、画像にオノマトペラベルを付与したデータセットの作成を独自に行った。作成には、連載30周年で知名度があり、清水[13]の論文でオノマトペの活用事例として取り上げられるほど、オノマトペを豊富に使用している荒木飛呂彦の『ジョジョの奇妙な冒険』を用いた。『ジョジョの奇妙な冒険』は全63巻で、計12,600ページからなる。図3のように、漫画中に登場するオノマトペの描かれているシーンを抜粋し、データセットを作成した。

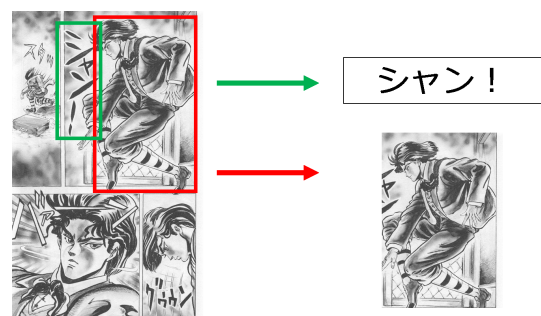


図3: データセットの作成例(『ジョジョの奇妙な冒険』1巻32ページ)

画像総数は3,176枚、ラベル数は465種類となり1つのラベルに対して平均6.83枚のデータが収集できたが、画像が1枚しかないラベルが113個もあった。深層学習は、大量のデータを与えられ特徴を自動的に学習していく手法であり、1枚しか画像がないデータでは特徴を学習できない。そのため、「アアア」と「アアアッ」や「ガァーン」と「ガーン」のように枚数が少なく似たような名前のラベルを統合した後、Data Augmentationを行った。Data Augmentationとは、画像に対して回転や拡大縮小などの変換処理を行い画像を複製するデータの拡張方法である。

本データセットには、図4の6種類の変換処理をランダムに組み合わせて行い、1ラベルあたりの画像枚数が50枚になるように拡張した。結果、ラベルの枚数に偏りがなくなり、画像総数は19,300枚、ラベル数は386種類となった。



図4: 変換処理の一覧

4.1.2 オノマトペ推定モデルの作成

作成したデータセットを用いて機械学習を行い、オノマトペ推定モデルを作成した。モデル構造を図5に示す。オノマトペ推定モデルの入力は150×150のカラー画像とし、出力は386種類のオノマトペの確率とする。このモデルを作成するために、大規模データで学習している高精度分類モデルのVGG-16[14]をFine-tuningして学習させた。VGG-16から全結合層(FC層)を1度消去し、386種類の分類ができるように全結合層を再構築した。構築後、VGG-16の14層目までを固定し、作成したデータセットで深い層の重みを再学習させた。

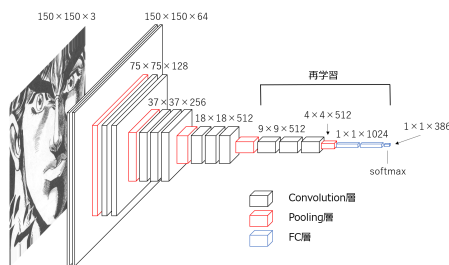


図5: 作成したモデルの構造

4.2 サーバ部

サーバ部では、HoloLens部から受信した画像を元にオノマトペの推定と顔の座標取得を行う。また、PCのマイクを用いて音声を取得し、音声認識を行う。なお、「セカマ

ン」は1対1での会話時の使用を想定しており、複数人の顔認識や話者の識別は行わない。

4.2.1 オノマトペの推定

HoloLens部から受信した画像を元に事前に作成したオノマトペ推定モデルを用いて、オノマトペを推測する。オノマトペの推測結果の一例を図6に示す。推定されたオノマトペのうち、確率が一番高いものをHoloLens部へ送信する。

キッ	39.20%
オゴオオオ	38.04%
ニカ	11.09%
ニヤリ	2.78%
ビシィ	2.33%

図6: オノマトペの推測結果の一例

4.2.2 顔の座標取得

HoloLens部から受信した画像を元に顔の座標を算出する。顔検出にはOpenCVのHaar-like特徴分類器を用いる。顔の検出ができた場合、顔を囲む矩形の座標情報をHoloLens部へ送信する。図8のように矩形の座標情報は左上を基準点とした x, y 、横方向の大きさ $width$ 、縦方向の大きさ $height$ からなる。

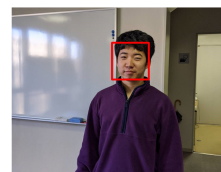


図7: 顔認識の例

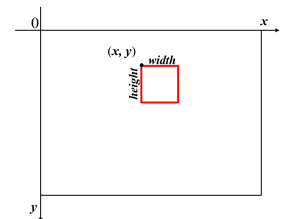


図8: 顔座標の情報

複数の顔を検出した場合は、先に検出した顔座標を送信する。検出した座標は一時的にサーバ内で保存する。顔が検出されなかった場合も、「セカマン」では会話での使用を想定しているため、会話相手はその場にいると仮定し、保存した座標を送信する。

4.2.3 音声取得

図9のように音声の取得には「セカマン」ユーザと会話相手の中間に設置したPCのマイクを使用する。取得した会話音声はWAV形式で出力し、出力されたファイルをバイナリ形式でエンコードする。そのバイナリ形式のデータをGoogleのCloud Speech-to-Text[15]に送信し、音声認識の結果を受け取る。Cloud Speech-to-Textは、日本語対応のクラウド型音声認識サービスの中でも認識精度が高いサービスであるため選択した。Cloud Speech-to-Textで取得した音声認識の結果はセリフとしてHoloLens部へ送信する。

4.2.4 HoloLens部への送信

4.2.1で取得したオノマトペと、4.2.2で取得した顔の座標と、4.2.3で取得した音声認識結果のセリフをHoloLens部



図 9: 「セカマン」利用時の構図

へ送信する。通信には Flask を用いる。送信は JSON 形式で行う。送信する JSON の内容は顔の座標 $(x,y) \cdot width \cdot height$, オノマトペ, 音声認識結果のセリフである。

4.3 HoloLens 部

HoloLens 部では、画像の取得、漫画風変換、および出力を行う。HoloLens のシステム開発のために Unity を使用した。

4.3.1 画像取得

画像は HoloLens 内蔵のカメラから取得し、バイト列として変換する。そのデータを Unity の WWW クラスを使用し、web サーバへ送信する。

4.3.2 画像の漫画風変換

漫画風変換は OpenCV for Unity を使用して行う。漫画風変換で一般的に用いられている、エッジ検出と 3 値化を用いた方法を使用した。4.3.1 で取得した画像とスクリーントーン画像を使用して変換する。変換手順を図 10 に示す。

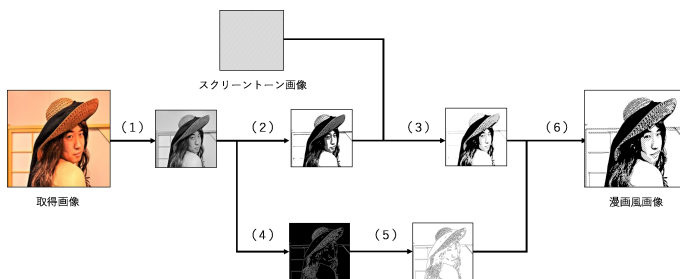


図 10: 漫画風変換の手順

- (1) グレースケール処理
- (2) 3 値化処理
- (3) 灰色の領域とスクリーントーン画像を入れ替え
- (4) エッジ検出処理
- (5) 白黒反転処理
- (6) 合成

また、例として図 11 (左) を漫画風変換した結果を図 11 (右) に示す。

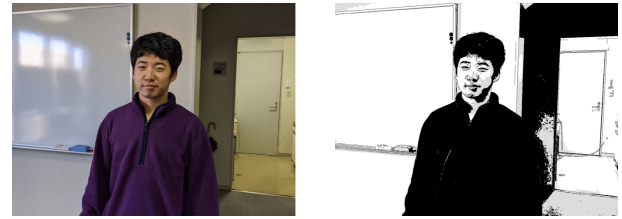


図 11: 漫画風変換の例

4.3.3 オノマトペ・吹き出しの表示

オノマトペ推定モデルで出力する全 386 種類のオノマトペの画像を作成した。作成には 2000 種類以上のオノマトペの画像が提供されているマンガ文字サイト「ドドド Fonts[16]」の画像を利用した。「セカマン」で使用するオノマトペ 386 種類のうち 189 種類は同サイトで画像が提供されている。提供されていない 197 種類のオノマトペは図 12 に示すように、似ているフォントの画像から文字を切り出しバランスを整え作成した。Unity で HoloLens のシステムを開発した場合に黒色部分が透過されてしまうため、白黒反転を行うことで文字の透過を防ぐ。作成した画像の一例を図 13 に示す。オノマトペの出力は、サーバ部から受信したオノマトペに対応する画像を張り付けたオブジェクトにより、切り替えを行う。

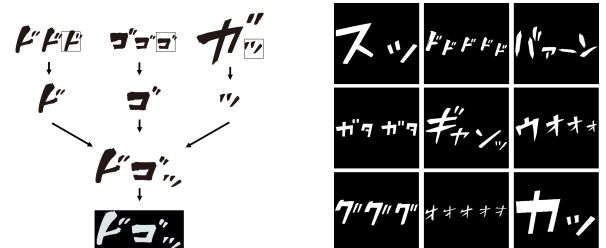


図 12: オノマトペ画像の作成 図 13: オノマトペ画像の一例

吹き出しは、吹き出しの画像とテキストの 2 つのオブジェクトで構成される。別々のオブジェクトとしたのは、今後吹き出しの画像やフォントの自動変更などを実装する場合に、互いに影響を及ぼさず変更が容易であると考えたためである。サーバ部から受信したセリフをテキストに配置し、縦書きに変更する。Unity の標準テキストは横書きとなっているが、一般的に漫画は縦書きのため変更を行う。テキストは文字数によって大きさを変化することで、文字数が多い場合でも吹き出し内に収まるようにする。半角は最大で 50 文字、全角で 25 文字まで枠内で表示する。文字数が超えた場合はそれ以降表示されない。

4.3.4 出力

4.3.2 で変換した画像を現実空間へ張り付け、4.3.3 で生成したオノマトペと吹き出しを配置する。オノマトペと吹き出しの配置位置は、サーバから受信した顔座標を元に顔と重ならないように調整した。実際の使用例と HoloLens を通した映像を図 14 に示す。

3人称視点の画像 HoloLensを通した映像

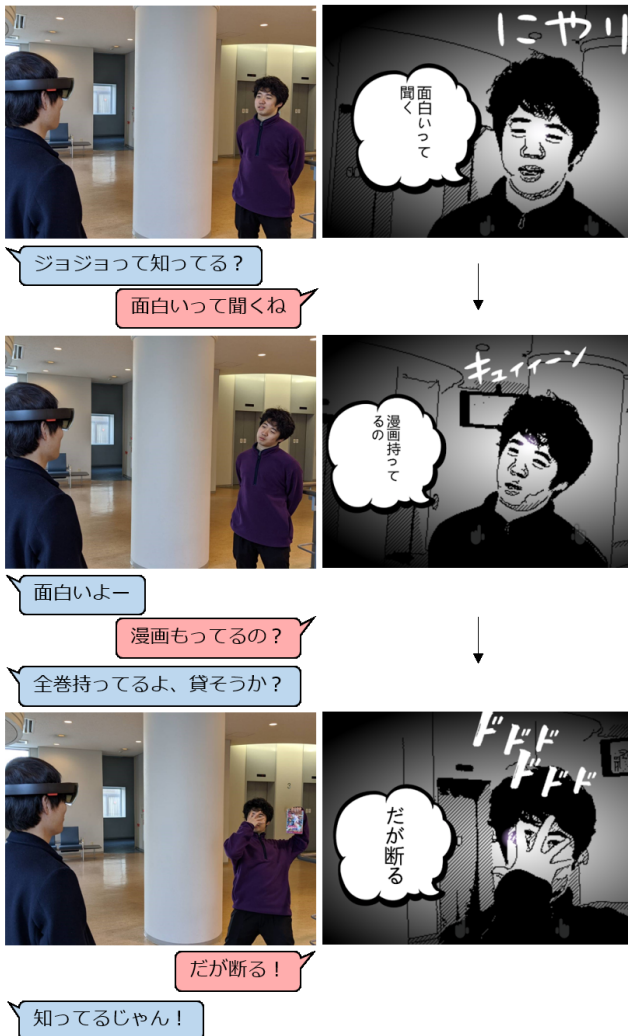


図 14: 実際の使用例と HoloLens を通した映像

5. 評価・考察

本章では、機械学習部で作成したオノマトベ推定モデルの評価と、提案手法を実装したシステム「セカマン」の評価を行う。

5.1 モデル評価

5.1.1 評価方法

作成したモデルについて精度評価を行う。全 386 種類のオノマトベに対して、我々が独自に定めたクラスでクラスタリングを行った。クラスは「attack」「emotion」「mood」「motion」「noise」「shout」の 6 種類からなり、それぞれ、「攻撃」、「感情」、「雰囲気」、「人間の動作」、「物体から発生する音」、「叫び声などの音声」を表現するオノマトベと定義した。これは、金田一が分類した「擬態語」「擬声語」「擬音語」「擬容語」「擬情語」[17]に、『ジョジョの奇妙な冒険』で多く登場していた攻撃に関するオノマトベを加えたものである。推定されたオノマトベのクラスが、正解オ

ノマトベの属するクラスかどうか検証する。『ジョジョの奇妙な冒険 Part6 ストーンオーシャン』1 巻から、4.1.1 と同様にして 29 枚のテストデータを作成した。

5.1.2 結果

検証結果を混合行列のヒートマップとして図 15 に示す。また、比較対象としてランダムにオノマトベを選出した結果も同様に図 16 に示す。

結果はオノマトベ推定モデルを用いた場合、29 枚中 9 枚が正解となり精度は 34.48% だった。一方、ランダム選出の場合は 29 枚中 4 枚が正解となり精度は 13.79% だった。

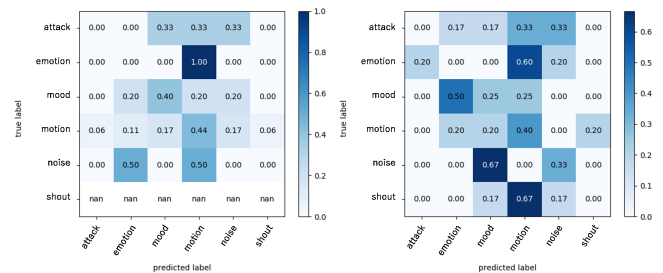


図 15: 本モデルの推定結果 図 16: ランダム選出の結果

5.1.3 考察

作成したモデルはランダムに出力するよりは精度が高くなったが、必ずしも満足のできる結果ではなかった。本モデルは漫画から切り抜いたものをそのまま学習させている。漫画には多数のキャラクターが登場しているため、図 17 のようにオノマトベが同じでも画像がかなり異なっていた可能性がある。



図 17: 異なる画像の例 (ドドド)

5.2 「セカマン」の評価

5.2.1 評価方法

評価は学生 14 名に HoloLens を着用し、「セカマン」を使用した対話を体験してもらい、アンケート調査を行った。項目は以下のとおりで、評価は 5 段階評価である。

5.2.2 結果

それぞれの項目の平均値および分散値を表 2 に示す。

表 1: アンケート項目

Q1. 体験して面白かったか	1:面白くない~5:面白い
Q2. 「セカマン」で漫画のように見えたか	1:見えない ~5:見える
Q3. 画風は漫画の要素として適切か	1:思わない ~5:思う
Q4. 吹き出しは漫画の要素として適切か	1:思わない ~5:思う
Q5. オノマトペは漫画の要素として適切か	1:思わない ~5:思う
Q6. 吹き出しとオノマトペの位置は適切か	1:思わない ~5:思う
Q7. 表示の遅延は少なかったか	1:思わない ~5:思う
Q8. オノマトペは情景に合っていたか	1:思わない ~5:思う

表 2: アンケートの平均値・分散値

	平均値	分散値
Q1	4.71	0.20
Q2	4.43	0.39
Q3	4.43	0.39
Q4	3.64	0.66
Q5	2.93	0.64
Q6	4.14	0.69
Q7	3.14	0.84
Q8	2.79	0.45

5.2.3 考察

Q1 に対して体験した全員が「面白い」「やや面白い」と評価した。よってシステム全体の評価はとても良かったといえる。また Q2 より、「セカマン」を通すことで現実世界を漫画世界に変換できたといえる。漫画の構成要素については、Q5 のオノマトペ部分が悪い評価となった。これは Q8 からわかるように、情景とオノマトペが合っていなかったためである。原因として、学習データはイラストであるのに対し、取得した画像は通常の画像であり、イラストと現実の画像とでは特徴が大きく異っていたと考えられる。一方で、Q3 から漫画風変換の評価は良かった。しかし、「想像より漫画世界の範囲が狭い」という意見が複数あった。HoloLens の処理性能の関係上、ディスプレイ全てを変換してしまうと正常に動作しなかったため、変換の範囲を中心から一定範囲としたことが原因だと考える。

6. 課題

オノマトペの精度向上

学習方法を変えることで精度向上が可能であると考えられる。学習データをキャラクターの骨格線とすることで体の形を学習させ、HoloLens からの入力も人の骨格線を使用することで余分な情報を省いた推定ができるのではないかと考える。

オノマトペの決定方法の変更

「セカマン」では、オノマトペ推定モデルによって推定された一番高い確率のオノマトペを出力する。ユーザが見ている景色に大きな変化がない限り出力されるオノマトペはほぼ同一である。そこでモデルによって

推定された確率を出力する確率とすることで、出力の幅を広げられると考える。

7. 展望

多人数への対応

「セカマン」は 1 対 1 での会話時の使用を想定して開発した。しかし、現実では複数人での会話の場面も多くある。そこで、音声の話者識別や複数の顔を認識することで多人数への対応を考えている。現時点では取得画像に複数人数が認識された場合、最初に認識された顔の座標にオノマトペと吹き出しが表示される。これは顔認証の結果を一人分しか送信していないためである。そのため会話をしていない人に表示される可能性がある。また音声も話者認識をしていないため、取得した音声は会話している人以外も文字起こしされてしまう。これらを解決したいと考える。

漫画の構成要素追加

新たに「漫符」「効果線」「吹き出しの切り替え」の実装を考えている。「漫符」とは怒りマークや汗マーク、眠っている際の「ZZZ...」のような感情や、状態を現した記号である。「効果線」とは表現効果などを付与する描線であり、人物の感情や雰囲気、動きを表現するために描かれる線である。

参考文献

- [1] 漫画, <https://ja.wikipedia.org/wiki/漫画>
- [2] 出版科学研究所, 出版月報 2018 年 2 月号
- [3] 『ジョジョの奇妙な冒険』 荒木飛呂彦, <https://ja.wikipedia.org/wiki/ジョジョの奇妙な冒険>
- [4] 『マンガ マッピン!』, 2017, <https://naked.co.jp/comic/>
- [5] 奇妙な写真メーカー, <https://dotapps.jp/products/com-objectgraph-BizarreComicPhotoFree>
- [6] 漫画コミックカメラ, <https://app-liv.jp/578506865/>
- [7] 小出雄空明, 國富彦岐, 藤村航, 奈良優斗, 白井曉彦 (2013), 「マンガ没入型 VR エンタテイメントシステムにおけるコンテンツ制作手法」, 第 18 回日本バーチャルリアリティ学会大会論文集
- [8] MaciejPko, AdamSvystun, PawełAndruszkiewicz, PrzemysławRokita, TomaszTrzcinski, (2018), 「Comixify: Transform video into a comics」, arXiv:1812.03473 [cs.CV]
- [9] 漫画カメラ, <http://tokyo.supersoftware.co.jp/mangacamera/>
- [10] 松下光範 (2019) 「コミック工学—漫画を対象としたコンテンツ指向研究」, システム/制御/情報 63 巻 3 号 p. 113-117
- [11] 住山晋一 (2004), 日本心理学会第 68 回大会ワークショップ
- [12] Flask, <https://a2c.bitbucket.io/flask/>
- [13] 清水祐一郎 (2016), 「音象徴に基づくオノマトペの印象評価システムと生成システムの設計」, 電気通信大学
- [14] VGG-16, http://www.robots.ox.ac.uk/vgg/research/very_deep/
- [15] Cloud Speech-to-Text, <https://cloud.google.com/speech-to-text>
- [16] ドドド Fonts, <https://dddfont.com/term/>
- [17] 金田一春彦 (1978) 「擬音語・擬態語概説」 浅野編『擬音語・擬態語辞典』 所収角川書店