

映像のマルチモーダル指標に基づく情報抽出手法の開発

Development of an Information Extraction Method Based on Multimodal Indicators in Video

○鎗田一輝¹, 塚本新²Kazuki Yarita¹, Arata Tsukamoto²

Abstract: The purpose of this paper is to propose a new interface for video. We developed a system to extract information from video based on multimodal indicators, and showed a concrete example of a content representation interface for video.

今日、機械学習や情報処理技術の発展により、様々なデータの価値が高まっている。その中で映像は、画像の時系列的つながりや音声や効果音等が含まれており、また見る人によって印象や得る情報が変化する媒体である。そこで、映像からマルチモーダル指標をもとに情報を抽出するシステムを用いた、映像に対する新しいインターフェースの提案を行うことで、指標に基づく情報処理的利用ができるのではないかと考えた。本稿では指標に文字認識、物体認識、物体分類モデルを利用し、映像に対する情報抽出手法の開発及び評価を行った。また、映像に対する内容表現インターフェースの具体例として一瞥で映像の全体像を把握するシステムの開発を行った。映像に対する情報抽出手法のシステム概要をFigure1に示す。

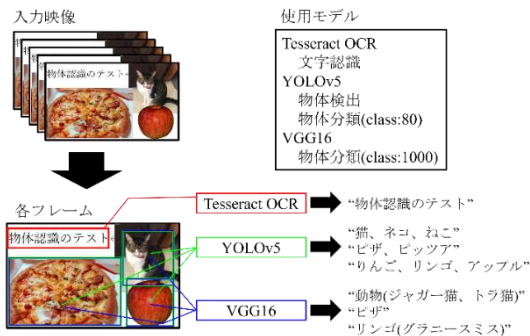


Figure1. System for extracting information

入力映像を文字認識モデル(Tesseract OCR^[1])及び物体認識モデル(YOLOv5 の学習済みモデル^[2])に適用し、その後物体認識モデルで検出された領域をさらに物体分類モデル(VGG16 の学習済みモデル^[3])に適用することで、映像内に含まれる文字及び物体ラベルを抽出した後に、物体ラベルを日本語に翻訳したもの及び検出文字を辞書として格納するシステムを開発し、評価を行った。文字認識モデルの評価には、光による影響や歪みが少なく、文字背景が半透明や様々な色である特徴を持つゲーム映像を使用し、著者がその映像におい

て主と判断した文字列を辞書内で検索したところ、26映像中18映像において該当文字列が検出されたことを確認した。また、物体認識及び物体分類モデルの評価には、分類クラスに含まれている物体を撮影した映像を使用し、物体ラベルを辞書内で検索したところ、動物や果物、野菜を含む計14クラス中全てのクラスにおいて90%を超える検出精度を確認した。

次に、映像に対する内容表現インターフェースの具体例として、上記システムにて作成した辞書から形態素解析器である janome^[4]を使用して名詞のみを取り出し、記号等を削除した後に、wordcloud^[5]を使用して頻出単語を可視化するシステムの開発を行った。動物を多く写した映像と果物や野菜を多く写した映像を使用した場合の出力をFigure2に示す。



Figure2. Visualization by wordcloud when using animal-based(left) and fruit-based(right) videos

本稿では、映像からマルチモーダルな指標をもとに情報を抽出するシステムを開発し、それを用いた映像に対する新しいインターフェースの具体例として、複数画像データより、動物、野菜、果物の種類という指標に基づいた単語及び出現頻度を可視化することで、一瞥で映像の全体像を把握するシステムを示した。

参考文献

- [1] R.Smith "An Overview of the Tesseract OCR Engine" ICDAR2007(2007)
- [2] G.Jocher "YOLOv5" <https://github.com/ultralytics/yolov5>(2020)
- [3] K.Simonyan,A.Zisserman "Very Deep Convolutional Networks for Large-Scale Image Recognition" ICLR2015(2014)
- [4] T.Uchida "Janome" <https://mocabeta.github.io/janome/>(2015)
- [5] A.Mueller "word_cloud" https://github.com/amueller/word_cloud (2016)