

# Web コンテンツを用いた 訃報映像の自動編集に向けた物体認識による画像選択法

國代 京花<sup>1,a)</sup> ナック フランク<sup>2</sup> 井手 一郎<sup>1</sup> 川西 康友<sup>1</sup> 出口 大輔<sup>1</sup> 村瀬 洋<sup>1</sup>

## 1. はじめに

著名人が死亡するとニュース映像などを通して訃報が伝えられる。その多くは、その人物の生前の功績や様子をまとめた映像の形で伝えられる。本研究ではそのような映像を「訃報映像」と呼ぶ。訃報映像は視聴者に対し、故人に関する情報を視覚的に伝えられる。しかし、訃報は速報性が求められることが多いために、訃報映像の内容は近年の限定的な話題で構成されることが多い。

そこで我々は、充実した内容の訃報映像を自動で生成することを目的とし、映像に用いるための素材を Web 上のコンテンツから抽出することにした。現在、我々はインターネット上の様々な媒体から情報を得ることができる。文字媒体ならば Wikipedia のようなオンライン百科事典から、映像や音声ならば Youtube のような映像共有サービスや画像検索エンジンから、必要な情報を容易に得ることができる。これらの Web コンテンツは、一般ユーザが盛んに投稿することで莫大な量の情報が日々追加、更新されているため、著名な人物に関わった出来事について、広範かつ最新の情報が得られる。しかし、大量の情報からある人物に関する出来事を網羅的に抽出して編集するのは大変な作業である。

そこで、本研究では大量にある情報から、個人に関する出来事を自動で選択する手法について検討する。本デモ発表では、実際に名前入力のみで自動編集を行った結果を紹介する。次章では、そのために必要な個人に関する情報の抽出方法と、映像に用いる画像の選択方法について述べる。

## 2. 提案手法

提案手法の処理手順を図 1 に示す。訃報映像は故人の人生を要約したものであり、幼少期、功績、私生活の 3 部で構成する。はじめに故人に関する特徴的な出来事を表現したキーワードを Wikipedia から収集する。次に、このキーワードをもとに画像検索を行う。一般に画像検索で画像

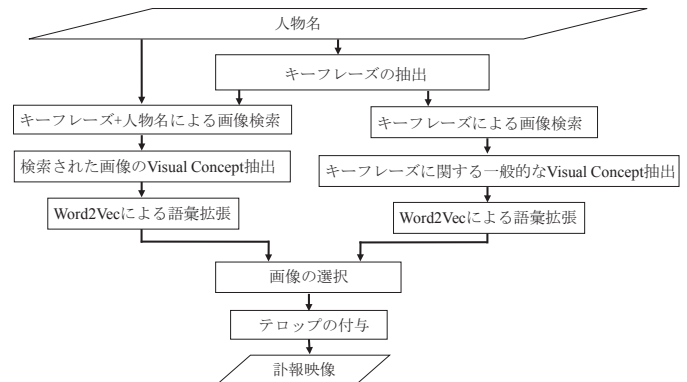


図 1 訃報映像生成の処理手順

を収集する場合、上位の画像ほどクエリの内容と一致しているとされる [2] が、クエリと視覚的な内容が一致しない画像が現れることもある。そこで画像の視覚的な内容を分析し、キーワードを表す画像の内容と最も一致するものを選択する。最後に、取得した画像に対応するキーワードをテロップとして付与し、スライドショーの形式で映像化する。

### 2.1 キーワードの抽出と分類

人物情報の収集は、Wikipedia とその内容を構造化したデータベース DBpedia [1] から収集する。DBpedia は Wikipedia よりも情報が整理されているため、これから生年月日、職業、出生地など定型的な情報を抽出する。しかし、功績や私生活に関する情報の分類や構造化は不完全であるため、それらについては Wikipedia から直接抽出する。

Wikipedia では、執筆者は強調したい語句をアポストロフィなどを用いた Wiki 記法により強調する。また文章中から張られるリンクのアンカテキストは固有名詞や重要な語句であることが多い。そこでこれらの語句をキーワードとして抽出する。一般に Wikipedia 記事冒頭の概要では、その人物の特徴が簡潔にまとめられているため、概要内の文章からキーワードを抽出する。しかし、概要には功績と私生活の内容が混在して記述されているため、それぞれに該当するキーワードを分類する必要がある。そこで、Wikipedia 記事中の見出しに対してテンプレートマッチングを適用する。このとき、予め功績と私生活に関連あ

<sup>1</sup> 名古屋大学 〒464-8601 愛知県名古屋市千種区不老町

<sup>2</sup> アムステルダム大学 情報学研究所 オランダ アムステルダム市

a) kunishirok@murase.m.is.nagoya-u.ac.jp

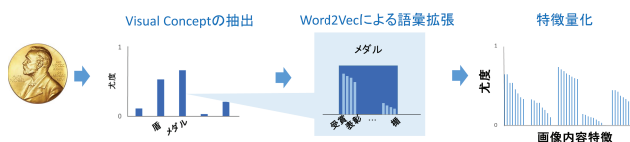


図 2 特徴抽出の様子

る語句を含む表現をテンプレートとしてそれぞれ用意しておき、それらと一致した見出しがあれば、その内容は功績と私生活のいずれかに関係する内容として分類する。分類されたそれぞれの内容の中に表れるキーワードと、概要文章内のキーワードとを比較し、一致するものを取得することで功績と私生活のキーワードを分類する。

## 2.2 画像の選択

取得したキーワードを用いて、故人とキーワードの両方に関係のある画像を選択する。その際に、故人の人名とキーワードをクエリとした画像検索で得られる画像の中から、特にその視覚的な内容がキーワードの内容と一致したものを選択したい。そこで、キーワードの内容を表す画像を分析し、内容を把握したうえで画像検索結果の中から選択を行う。

### 2.2.1 画像内容特徴の抽出

画像の内容を把握するため、本研究では Visual Concept を各画像から抽出する。Visual Concept は画像内に写っている複数の物体を認識し、認識した物体ごとの尤度を算出する。そこで、Visual Concept を最も基本的な画像内容特徴とする。しかし、Visual Concept で表現できる語彙は少ないため、Word2Vec [4] を用いて、各 Visual Concept の名称と共起確率が高い語句を用いて語彙を拡張する。これによって得られる拡張された語彙の尤度を並べたものを特徴量とする。この特徴抽出の様子を図 2 に示す。

### 2.2.2 画像の比較と選択

先ほど述べた特徴抽出を、キーワードのみを用いて検索して得た画像と、故人の人名とキーワードを用いて検索して得た画像群の両方に適用する。

まずはキーワードのみをクエリとする検索をして、得られた全ての画像から画像内容特徴を抽出する。この画像内容特徴を、そのキーワードに関する一般的な画像内容特徴とする。次に人物名とキーワードの両方をクエリとして画像検索を行い、得られた画像から画像内容特徴を抽出する。これを先ほど抽出した一般的な画像内容特徴と比較し、最も一致する画像を選択する。具体的には、キーワードに関する一般的な画像内容特徴と、人物名とキーワードから得た画像の画像内容特徴の内積値が最大の画像をキーワードに最も一致する画像として選択する。

## 3. 実験

提案手法の有効性を検証するため、15 人の著名人に対

表 1 実験結果

選択手法	比較手法	提案手法
平均選択率	0.243	0.379

し、その人に関連あるキーワードをよく表す画像を選択する実験を行った。被験者は人物名とキーワードで検索した検索結果の上位 4 枚から、最もキーワードを表していると考えられる画像を 1 枚選択する。その選択された画像と、提案手法により選択された画像の一致度を調べた。比較手法は、著名人の人物名とそれぞれに関係のあるキーワードをクエリとして画像検索を行った結果の第 1 位の画像を利用したものである。評価は手法の選択率で行い、総設問数のうち各手法が選択された設問数の割合とする。

表 1 に実験結果である各手法の選択率を載せる。実験結果より、提案手法は比較手法より 13.6% の制度向上が見られた。また、提案手法が有効だったキーワードは、「アカデミー賞」や「ユニセフ親善大使」など、よく知られているものであった。これらのキーワードは、「トロフィー」や「ドレス」などといった一般的な特徴が抽出しやすいものであったために精度が高かったと考えられる。一方、提案手法が有効でなかったキーワードはあまり認知されていないキーワードであった。これより、被験者はキーワードの内容がわからなかったために正しく画像の評価が行えなかったと考えられる。

## 4. むすび

本デモ発表では Web 上のコンテンツを用いた計報映像の自動生成手法について述べた。提案手法では画像を選択する際に Visual Concept を用いたが、語彙にないものは検出できない。今後は語彙にない Visual Concept に対しても、より適した画像を選択する手法へと拡張したい。

## 5. 謝辞

本研究の一部は、科学研究費補助金による。

## 参考文献

- [1] C. Bizer, “DBpedia—A crystallization point for the Web of data,” Web Semantics: Science, Services and Agents on the World Wide Web, 7(3), pp.154–165, Sept. 2009.
- [2] Y.-G. Jiang, C.-W. Ngo, and J. Yang. “Towards optimal bag-of-features for object categorization and semantic video retrieval,” in Proc. 6th ACM Int. Conf. on Image and Video Retrieval, pp.494–501, July 2007.
- [3] S.K. Divvala, A. Farhadi, and C. Guestrin, “Learning everything about anything: Webly-supervised visual concept learning,” in Proc. 2014 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.3270–3277, June 2014.
- [4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” International Conference on Learning Representations, Sept. 2013.