

# 観光動画キュレーションの実現に向けたハイライト抽出手法の検討

## A method for highlights extraction toward tour video curation

金谷 勇輝<sup>†</sup> 中村 優吾<sup>†‡</sup>

Yuki Kanaya Yugo Nakamura

諏訪 博彦<sup>†</sup> 荒川 豊<sup>†§</sup> 安本 慶一<sup>†</sup>

Hirohiko Suwa Yutaka Arakawa Keiichi Yasumoto

### 1. はじめに

2020年に東京オリンピック・パラリンピックの開催が予定されており、情報技術を利用した観光業を活性化させる取り組みが期待されている[1]。インバウンドの増加に伴う訪日外国人旅行者の増加が予想され、すべての観光客に対してストレスなく快適に観光を満喫できる環境の整備が必要不可欠である。国内外を問わず観光ルートを推薦するサイトは多数存在するが、供給されている観光コンテンツは散発的かつ静的なものが多く、観光ルートの全体像をイメージすることが困難である。さらにバリエーションが少なくユーザそれぞれの嗜好を考慮することが難しい。

一方、観光客自身が所有するスマホや小型カメラによって撮影された多様なアングルでの臨場感ある動画像、リアルな口コミなどの情報媒体 (CGM: Consumer Generated Media) がインターネットを通して共有できる時代になっており、これらを有効活用して他の観光客にとって価値あるコンテンツ生成しようとする試みもされている[2][3]。この研究は近年ニュース配信サービスなどで浸透しつつあるキュレーション (curation) という概念に基づくサービスの実現を目指している。キュレーションとは、キュレーターと呼ばれる情報の編纂者が、さまざまな情報を独自の判断で収集・整理し、新しい価値を持たせて共有することで、ユーザは整理された価値の高い情報を受け取れる仕組みである[4]。またキュレーションされたレシピ動画をユーザに提供する Kurashiru [5] のようなアプリも登場し、キュレーションという概念が浸透しつつある。

本稿では観光案内向け CGM キュレーションを題材にし、各観光スポットで撮影された多様な動画像を、消費者の嗜好データに基づいて短編動画として編纂する観光動画キュレーションシステムの実現に向けて、観光に関して特徴を持つハイライトを抽出するための手法を提案する。

### 2. 関連研究

#### 2.1 映像要約手法

映像要約とは、長時間の映像から特徴的なシーンを抽出することで、視聴時間を削減する技術である。対象となる映像はニュース番組、スポーツ中継など多岐にわたる。基本的なアプローチとして、視覚的な特徴量を用いた映像要約手法があげられ、特徴量の変化から要約箇所を抽出している[6]。また、映像に対して Google の Video Intelligence API [7] や Microsoft Azure Cognitive Services [8] を用いて動画にアノテーションし、映像要約する手法が考えられる。これらはシーンが大きく変化する部分は抽出可能だが、CGM による観光動画はそのようなシーンが生じることは少なく、目的とするシステムには不向きである。

そのほかに、音声情報を考慮した手法[9]も提案されているが、CGM にはナレーションのような音声が含まれることもないので、音声情報による抽出も不向きである。

#### 2.2 ユーザの嗜好に合わせたキュレーション内容の提供

キュレーションされた内容をユーザの嗜好に合わせて提供するアプリが多数存在する。まず SmartNews [10] や Gunosy [11] があげられる。これらはユーザが「エンタメ」や「スポーツ」等あらかじめ設定した興味あるトピックやタッチしたコンテンツの履歴に基づいて、大量のデータからユーザの興味のありそうなコンテンツを収集し、評価付けを行い取捨選択し、提供するアプリである。

また Kurashiru [5] はキュレーションされたレシピ動画をユーザに提供することでユーザは簡単に料理を学ぶことができ、その日の献立を決定するためにも利用できる。しかし、キュレーションされた観光動画を提供するシステムは未だ存在しない。藤澤ら [12] は、一般のスポーツ観戦者が撮影した複数の映像の編纂作業を機械学習により自動で行うシステムを提案している。このシステムでは、撮影された複数の映像ストリームを一箇所のストレージ空間に集約し、蓄積された映像データに対して、メタ情報の付加、TV 放送からの教師データの取得、モデルの構築を行なっている。これに対して、本研究では一般の観光客が撮影した映像のみから短編動画をキュレーションする研究を行う。

<sup>†</sup> 奈良先端科学技術大学院大学, Nara Institute of Science and Technology.

<sup>‡</sup> 日本学術振興会特別研究員 DC1, JSPS Research Fellow.

<sup>§</sup> JST さきがけ, Japan Science and Technology Agency.

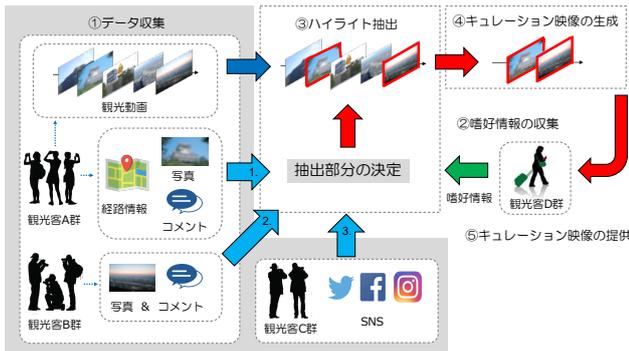


図 1: 提案システムの全体像

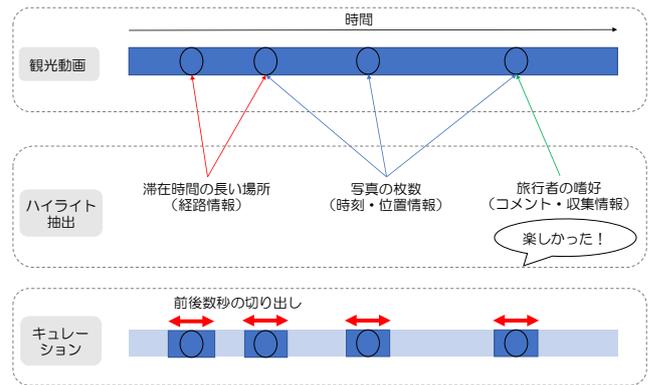


図 2: ハイライト抽出の概要

### 3. 観光動画キュレーションシステム

本章では提案システムの概要と提案システムを実現するにあたって必要となる技術をまとめる。

#### 3.1 提案システムの概要

提案システムの全体像を図 1 に示す。本システムでは、役割に応じて観光客を 4 種類に分類する。観光客 A 群は動画を撮影するとともに、写真やコメント、経路情報を収集する。観光客 B 群は観光客 A 群と共に行動し、写真とコメントのみ収集する。観光客 C 群は一般の観光客であり、SNS などに写真やコメントを投稿する。観光客 D 群は、システムの利用者であり、嗜好情報をインプットすることによりキュレーションされた観光動画を取得する。

また提案システムの処理ステップを以下に示す。

1. データ収集ステップでは、観光客 A 群が撮影した観光動画と観光地の写真、観光客 B 群が撮影した写真や観光客 C 群によって SNS に投稿された写真を収集する。
2. 嗜好情報収集ステップでは、観光客 D 群の嗜好情報を収集し、観光コースを抽出する。
3. ハイライト抽出ステップでは、観光客 A 群が撮影した写真、動画の経路情報及びコメント、観光客 B 群及び観光客 C 群から収集した写真や付随するコメントから観光動画のハイライトを抽出する。
4. キュレーション映像の生成ステップでは、ハイライト部分を組み合わせ、キュレーション映像を生成する。
5. キュレーション映像の提供ステップでは、観光客 D 群にキュレーション映像を提供する。

提案システムによって観光客 D 群は自らの嗜好情報に応じたキュレーション映像を受け取ることができる。ま

た、映像でキュレーション結果を受け取るため、他の静的な観光コンテンツに比べて観光ルートの全体像を容易にイメージできる。

#### 3.2 データ収集

観光客 A 群と観光客 B 群のデータは直接収集する。観光客 C 群によって撮影された写真や付随するコメントの収集には Twitter<sup>1)</sup> や Flickr<sup>2)</sup> などの SNS を用いる。また観光客 A 群は SilentLog<sup>3)</sup> などのアプリケーションを使用し経路情報を収集する方法が考えられる。

#### 3.3 嗜好情報の収集

観光客 D 群は興味関心のある内容をシステムにインプットする。例えば、神社が好きで観光客もいれば食に重きを置いて観光を行いたい観光客もいる。そのため、各観光客の嗜好を収集し、嗜好に合わせた観光コースを抽出する。

#### 3.4 ハイライト抽出

ハイライト抽出の概要を図 2 に示す。

ハイライトとして抽出すべき程度をここでは「興味度」として定義する。まず、抽出された観光コースに合わせた観光動画を収集された動画データから抽出する。その中から、興味度の高い場面を抽出し、前後数秒を切り出しハイライトとして抽出する。抽出する方法として、以下の三つを考える。

1. 観光客 A 群が撮影した写真を用いた方法
2. 観光客 B 群が撮影した写真を用いた方法
3. 観光客 C 群が撮影した写真を用いた方法

動画撮影者（観光客 A 群）が撮影した写真は、特に動画撮影者の興味が高い場面と考える。そのため、1 を用

<sup>1)</sup> <https://twitter.com>

<sup>2)</sup> <https://www.flickr.com>

<sup>3)</sup> <https://silentlog.com/ja/>

いることで動画撮影者の興味の高い映像が抽出できると考える。また、動画撮影者の滞在時間が長い場所は興味度が高い部分としてハイライト抽出に利用できる。

また、観光客 B 群が撮影した写真は、同行者の興味が高い場面と考える。そのため 2 を用いることで同行者の興味が高い映像が抽出できると考える。同行者の興味を加えることで、観光客 A 群が撮り逃した場面を補完することができる。

観光客 C 群が撮影した写真は、一般の観光客の興味が高い場面と考える。そのため、3 を用いることで一般の観光客の興味が高い映像が抽出できると考える。SNS への写真のアップロード数が多い場所は、誰もが行きたいと考える場所であり、ハイライトとして抽出すべき場所と考える。

さらに、あらかじめ収集された観光客 D 群の嗜好情報を元に、関連する映像を抜き出し、ハイライトとして抽出する。加えて、写真投稿者のコメントを用いることで、興味度の判定や嗜好とのマッチングを行う。

### 3.5 キュレーション映像の作成

ハイライトとして抽出された部分を単純に組み合わせると、観光ルートのイメージではなく、観光スポット付近のみを写した映像が出力されることが予想される。そこでハイライトとして抽出されなかった部分はタイムラプスのように映像化し、抽出された部分はスロー再生することで、映像を受け取る観光客 D 群は観光ルートの全体像をイメージしやすくなる。

また観光客 D 群はハイライト部分を前後何秒切り出すかを設定できる。例えば、観光客 D 群が観光地に訪れる前に詳細な情報を得たい場合には、前後の切り出し時間を長くする。対照的に、すでに観光地に到着しており、すぐに観光ルートのイメージを得たい場合には前後の切り出し時間を短くする。このシステムは、観光客の嗜好を考慮したキュレーションの実現につながる。

## 4. ユーザスタディ

ここでは団体客による岐阜県への観光を対象とし、その間の観光動画と参加者による写真を用いてハイライト抽出手法の検討を行った。本章ではその結果をまとめる。

### 4.1 ハイライト抽出

31 人の参加者によって、59 本の動画と 179 枚の写真が集まった。このうち岐阜城の散策ルートを撮影した動画 1 本を対象としてキュレーションを行う。

動画は 2017 年 6 月 11 日の 10 時 31 分から 10 時 40 分の 9 分間によって撮影された。その時間内に参加者によって撮影され投稿された写真は 9 枚であった。それぞれが撮影された時間軸を合わせてハイライト抽出を行った。撮影された写真の枚数を分単位でまとめたグラ

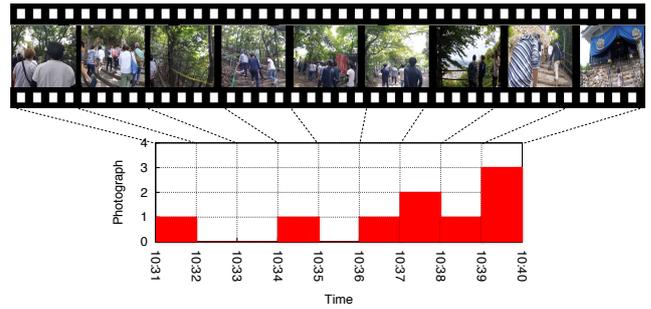


図 3: 各時間における写真枚数とビデオフレーム



図 4: キュレーション結果の抜粋

フとそれぞれの時間に対応するビデオフレームの一部を図 3 に示す。

図 3 より写真の枚数が多い 10 時 37 分と 10 時 39 分の動画がハイライトとして抽出された。

### 4.2 動画キュレーション

抽出されたハイライトに基づき、10 時 37 分 0 秒と 10 時 39 分 0 秒から前後 20 秒ずつ切り出しキュレーションを行なった。その結果、図 4 のように写真と動画に写る画像の類似度が高い映像部分を抜き出すことができた。

### 4.3 考察

ここでハイライト抽出に用いた写真は 3.3 節の「2. 観光客 B 群（動画撮影者と一緒に観光した人）が撮影した写真」に該当する。写真枚数に従ってハイライト抽出すると最終目的地となる岐阜城付近の映像とそのルート中の見晴らしが良い場所が抽出された。抽出された映像には多くの参加者が写っており、参加者の滞在時間も長かった。よって、参加者の興味が高い場面を抽出できたと考えられる。

なお、経路情報やコメントについては収集されていなかったため、今回のハイライト抽出には反映されていない。今後、これらを反映させたハイライト抽出の検討を行う必要がある。

## 5. まとめ

本稿では観光動画キュレーションシステムの提案とその実現に向けた観光動画のハイライト抽出手法を提案した。またユーザスタディより提案手法によって興味度の高いハイライト抽出が可能であることを示した。今後は、コメントや経路情報を考慮したハイライト抽出を検討するとともに、観光動画キュレーションシステムの構築、キュレーション映像に対する評価実験を行う予定である。

## 参考文献

- [1] 国土交通省観光庁. <http://www.mlit.go.jp/common/001126601.pdf>.
- [2] 中村優吾, 諏訪博彦, 荒川豊, 山口弘純, 安本慶一ほか. 観光案内向け cgm キュレーションのためのローカル IoT プラットフォームの提案. マルチメディア, 分散協調とモバイルシンポジウム 2016 論文集, Vol. 2016, pp. 1108–1115, 2016.
- [3] 日高真人, 松田裕貴, 河中祥吾, 中村優吾, 藤本まなど, 荒川豊, 安本慶一ほか. 実時間観光コンテンツ提供に向けた観光情報収集・キュレーションシステムの提案. 研究報告高度交通システムとスマートコミュニティ (ITS), Vol. 2017, No. 15, pp. 1–8, 2017.
- [4] 安本慶一, 山口弘純. 多数のデータストリームを実時間で融合・編纂し利活用するための次世代「情報流」技術. 情報処理, Vol. 55, No. 11, p. 2014, 2014.
- [5] Kurashiru. <https://www.kurashiru.com>.
- [6] Bin Zhao and Eric P Xing. Quasi real-time summarization for consumer videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2513–2520, 2014.
- [7] Google cloud video intelligence. <https://cloud.google.com/video-intelligence/?hl=ja>.
- [8] Microsoft azure cognitive services. <https://azure.microsoft.com/ja-jp/services/cognitive-services/>.
- [9] Georgios Evangelopoulos, Athanasia Zlatintsi, Alexandros Potamianos, Petros Maragos, Konstantinos Rapantzikos, Georgios Skoumas, and Yannis Avrithis. Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention. *IEEE Transactions on Multimedia*, Vol. 15, No. 7, pp. 1553–1568, 2013.
- [10] Smartnews. <https://www.smartnews.com/ja/>.
- [11] Gunosy. <https://gunosy.co.jp>.
- [12] Kazuki Fujisawa, Yuko Hirabe, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. Automatic live sport video streams curation system from user generated media. *International Journal of Multimedia Data Engineering & Management*, Vol. 7, No. 2, pp. 36–52, 2016.