

観光メモリアル動画のための 車載動画キュレーションアルゴリズムの検討

片山 洋平^{1,2} 平野 陽大^{1,2} 諏訪 博彦^{1,2} 伍 洋³ 安本 慶一^{1,2}

概要：近年、動画を観光に活用するケースが増えている。観光後に観光地の画像をキュレーションしメモリアル動画とするようなシステムは多数存在しているが、観光経路動画をキュレーションするシステムはまだ存在していない。本稿では、観光地でのレンタカー利用後の観光客に向けた、メモリアル動画のためのドライブレコーダ動画キュレーションアルゴリズムを提案する。提案手法ではクラウドソーシングにより重要度スコアを取得し、画像に写るカテゴリと地域のランドマークを利用した特徴量を用いて機械学習による重要度スコアの推定を提案した。本稿では観光経路全体の重要度スコア取得を行う本実験に向けて、沖縄で撮影されたドライブレコーダ画像を50枚利用した予備実験により提案アルゴリズムの性能評価を行った。予備実験の結果から「沖縄らしさ」が「メモリアル動画への必要度」と相関係数0.92を示し有意に強い正の相関を持つこと、提案手法から「沖縄らしさ」の重要度スコアの推定が可能であることを示唆した。

A study on driving video curation algorithm for memorial sightseeing movie generation

YOHEI KATAYAMA^{1,2} YODAI HIRANO^{1,2} HIROHIKO SUWA^{1,2} YANG WU³
KEIICHI YASUMOTO^{1,2}

1. はじめに

近年、インターネットにおける動画コンテンツは増加しており、Google社の調査では観光客の約40%以上が観光計画を立てる際に、他の観光客のメモリアル動画などを利用すると報告されている[1]。観光終了後のメモリアル動画として、スライドショー形式で観光スポットの写真を見返すなどがある。RealNetworks社の提供する「RealTimes」では、写真に対応する位置情報に基づいて画像からショートムービーを作成・提供している[2]。既存研究のように観光スポット周辺の写真を利用したメモリアル動画は多数あるが、観光経路の動画は少ない。観光において大きな時間を占める経路動画をユーザに提供することは観光全体の様子を連続的に想起するための手助けになると考える。

一方、ドライブレコーダの出荷数は増加しており、2018年度は国内で年間約367万台が出荷されている[3]。あおり運転報道の影響もあり、今後も利用者は増えると考えられるが、それらの動画は事故やアクシデント時にしか利用されておらず、動画というリッチな情報が活用されていない。本研究ではその点に着目し、取得される動画を観光経路動画としてキュレーション(収集・編纂)することで、観光客が観光全体の様子を連続的に想起するシステムを構築することを目指す。

観光経路動画の活用場面は、観光前のマーケティングやプランニング、観光後の振り返りやシェアリングなど、様々なシーンが存在する。上記のシーンはいずれも短時間で効果的にユーザに提示することが望ましいと考えられる。したがって、ドライブレコーダ動画を利用する上での課題として収集動画が長時間であり直線的な道のりや信号での停車など冗長な部分が含まれることがあげられる。

本稿では、特にレンタカーでの観光終了後の観光客に対するメモリアル動画の作成・提供のために、観光経路動画

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology
² 理化学研究所,RIKEN
³ 京都大学,Kyoto University

をキュレーションするアルゴリズムを提案する。提案アルゴリズムでは、クラウドソーシングにより重要度を取得し、画像に写るカテゴリと地域のランドマークを利用した特徴量を用いて機械学習による重要度スコアの推定を行う。また、予備実験では沖縄で撮影されたドライブレコーダ画像を利用して提案アルゴリズムの検証を行った。予備実験の結果より「沖縄らしさ」が「メモリアル動画への必要度」と相関係数 0.92 を示し強い正の相関を持つこと、提案手法から「沖縄らしさ」の重要度スコアの推定が可能であることを示唆した。

以降の章構成は以下の通りである。2 章では、提案手法に関連した既存研究を概説すると共に、本提案の位置付けを明らかにする。その後、3 章では、提案アルゴリズムについて述べ、4 章で提案アルゴリズムの有効性を示すための評価実験の内容の説明を行い、5 章で実験結果および結果の考察を行う。6 章では、5 章の結果についての議論を行う。最後に、7 章で本論文のまとめを述べる。

2. 関連研究

本章では、本研究と関係する既存研究としてドライブレコーダ動画を活用した研究とシーンの重要度を利用した動画要約の研究について概説し、本研究の位置づけを明確にする。

2.1 ドライブレコーダ動画の活用研究

既存のドライブレコーダ動画の主な活用事例は、運転技術や運転マナーのフィードバックに関するものである [4][5]。これらの研究では運転者や車体の運転情報を用いて車外の風景を取り扱うものは少ない。Takenaka ら [5] の研究ではドライブレコーダ動画から抽出される運転動作に対応した音楽を合わせて要約動画を作成している。これは抑揚のあるキュレーションによって、視聴者を退屈にさせない要約手法が必要であることを示唆している。佐藤ら [6] は、動画中から動き特徴や標識を検出して観光のための道案内要約動画を自動生成している。この手法では今後観光をするユーザを対象に動画生成を行っているため、メモリアル動画の生成には適さない。

2.2 画像の綺麗さを評価する研究

Talebi ら [7] の研究では、深層学習により画像の美しさについて定量的な評価を行うモデルを提案している。学習データには画像とアマチュア写真家による 10 段階の評価を利用している。この研究では写真家が考える構図などの綺麗さを表現しているものの、運転風景での綺麗さの定量化はされない。したがってメモリアル動画を作成する上で重要となる綺麗な景色を定量化するタスクにおいては適さない。

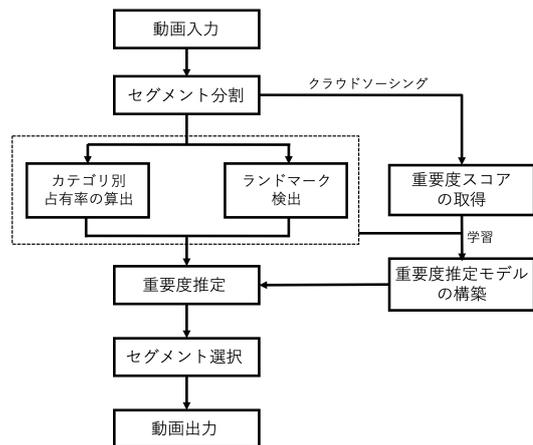


図 1 提案手法の概要図

2.3 重要度推定を用いた動画要約の研究

CISCO 社の調査では 2017 年の世界のデータ通信量のうち 75% が動画であることが報告されている [8]。また同調査では、2022 年には動画の通信量が 82% まで増加することが予測されている。これに伴い動画要約の研究が各分野で盛んに行われている。Song ら [9] は YouTube 上の多様なカテゴリの動画に対して 2 秒ごとに重要度を推定し、そこから要約動画を作成している。汎用的な要約動画の作成には適しているが、メモリアル動画を作成するというタスクにおいては重要なシーンが落とされる可能性がある。金谷ら [10] は、観光歩行時の 1 人称カメラから重要度の推定を含むキュレーションを行い、観光経路動画を作成しユーザに提示する手法を提案している。ここで、歩行と走行車両では通るルートや視点が異なるため、この手法を直接メモリアル動画作成に応用することは難しい。

ドライブレコーダ動画からメモリアル動画を作成する上で、視聴者の興味にあわせて重要な場面を提示することが求められるが、各個人に合わせて人手でキュレーションすることは現実的ではない。よって、本稿ではメモリアル動画における重要なシーンの抽出アルゴリズムを提案する。

3. 提案手法

本研究で提案するアルゴリズムの概要を図 1 示す。入力として車両のエンジンが始動してから、切られるまでのドライブレコーダ動画を使用し、出力としてメモリアル動画を生成する。

3.1 セグメント分割

提案手法では、車載動画をセグメントと呼ぶ同じ意味を持つフレーム群に分割し、各セグメントの重要度を推定しキュレーションを行う。セグメントの長さの単位として経験的に 3 秒を採用する。これは車の時速 60km と仮定した際に、50m 先のオブジェクトがフレームアウトするまでの時間である。重要なセグメントとしては、観光地から車が発進・到着時のセグメント、観光地を表現するランド

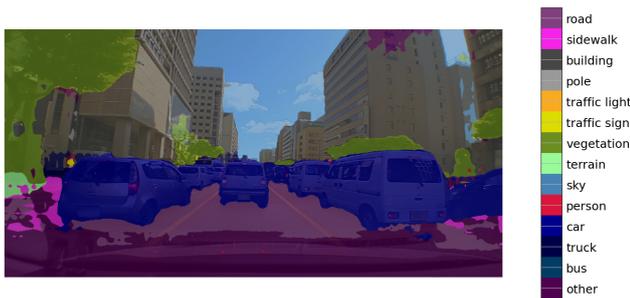


図 2 DeepLabv3 による検出結果の例

マークが含まれるセグメント，綺麗と感じる風景が含まれるセグメントなどが想定される。

3.2 カテゴリ別占有率の抽出

画像中のカテゴリを画素単位でラベル付けするタスクをセマンティックセグメンテーション (Semantic Segmentation) と呼ぶ。予備実験の結果から，画像中の空の広さや道路の混雑度が重要度に影響しているとのコメントを得た。よって提案手法ではセマンティックセグメンテーションで高い精度を出している DeepLabv3[11] を利用し，ドライブレコーダ画像中に含まれるカテゴリの検出を行う。学習データには街の画像に対応するセグメンテーション画像を含むオープンデータセットである CityScapes[12] と BDD100k[13] を用いた。検出結果を図 2 に示す。画像は車，道路，空などを含む 20 カテゴリに画素単位で分類される。画像中に各カテゴリの存在する画素数を用いてカテゴリ別の画像内占有率を算出する。

3.3 ランドマーク検出

ランドマークとはその土地を表現するオブジェクトである。予備実験で受け付けたコメントの結果からヤシの木などのその土地のランドマークが重要度に影響しているとの指摘を得た。よって提案手法では物体検出で高い精度を出している，YOLOv3[14] を利用し，ヤシの木の検出を行う。学習データには同時期に撮影された沖縄でのドライブレコーダ画像と Web からのデータクロウリングにより取得された画像を利用する。検出結果を 3 に示す。

3.4 重要度スコアの取得

セグメント重要度に関するアノテーションは観光終了後の観光者から取得するのが好ましい。しかし長時間におよぶ動画を観光者が全て評価するのは現実的に難しい。Song ら [9] は YouTube 上の動画に対し，クラウドソーシングを利用して重要度のアノテーションを付与している。そこで本研究でもクラウドソーシングを利用し，セグメントごと

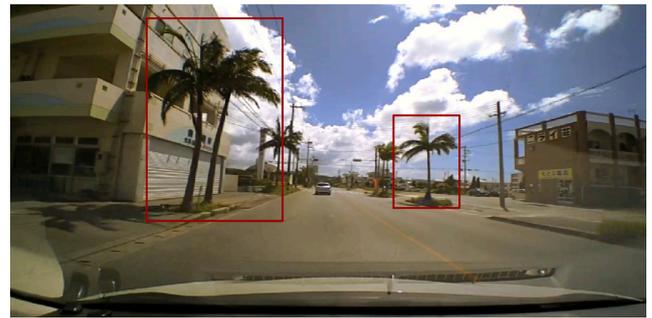


図 3 YOLOv3 による検出結果の例

の重要度の定量化を図る。観光経路動画の重要度を評価する際に，その土地への観光経験の有無は影響すると考えられる。よってクラウドソーシングのタスクを実施する際には，事前にスクリーニング検査として対象地域への観光経験の有無を調べた。観光経験のある対象をホワイトリストに追加し，タスクを実施する。タスクは，評価を各セグメントごとに 20 人が回答するように設定する。予備実験では沖縄の風景の評価を得た。回答者にはレンタカー利用者としての立場を想定して以下の三項目について「5. とてもそう思う」から「1. 全くそう思わない」の 5 段階で重要度について回答する。レベルは Song らがクラウドソーシングを実施した際の指標に合わせている。

- 沖縄らしい風景
- 綺麗な風景
- メモリアル動画に必要な風景

図 4 は回答者から見たタスクの画面である。回答者はリンクをクリックすることでセグメントの画面に一時的に遷移する。

3.5 重要度推定モデルの構築

3.4 節で取得された重要度スコアを利用して，重要度推定モデルの構築を行う。モデルの学習には特徴量として，セマンティックセグメンテーションから得られた 20 カテゴリについて各カテゴリが画像を占める割合とランドマーク検出から得られたヤシの木の有無についての 21 次元の特徴量を用いる。推定には回帰分析のモデルとして重回帰分析と Support Vector Regression (SVR) の 2 手法を提案する。

3.6 重要度推定

セグメントの重要度の推定には特徴量として，セマンティックセグメンテーションから得られた 20 カテゴリについて各カテゴリが画像を占める割合とランドマーク検出から得られたヤシの木の有無についての 21 次元の特徴量を入力とし，前節で構築した重要度推定モデルにより重要

設定した設問ID: 1

あなたは沖縄でレンタカーを借りて観光をしました。旅行から帰ってきたあなたは、思い出動画(以降ではメモリアル動画と呼びます)を作成しようとしています。観光中の移動も旅行の一部だと考えたあなたは、レンタカーのドライブレコーダーからシーンをいくつか切り取り、見比べています。画像を確認し、直感的に以下のアンケートにお答えください。

画像はこちらです。 [チェックするページを見る](#)

画像は「沖縄らしい風景」だと思いますか？

5.とてもそう思う 4.そう思う
 3.どちらともいえない 2.そう思わない
 1.全くそう思わない

画像は「綺麗な風景」だと思いますか？

5.とてもそう思う 4.そう思う
 3.どちらともいえない 2.そう思わない
 1.全くそう思わない

画像は「沖縄旅行のメモリアル動画に必要」だと思いますか？

5.とてもそう思う 4.そう思う
 3.どちらともいえない 2.そう思わない
 1.全くそう思わない

ご協力ありがとうございました。

[確定して次へ](#)

図 4 クラウドソーシングでのタスク画面

度を推定する。

3.7 セグメント選択

推定された重要度に基づいてメモリアル動画として必要なセグメントを選択する。この際に車の出発到着セグメントとは現在地の理解を促すため必要となる。残りのセグメントについて、メモリアル動画の再生時間の要件を満たすように、重要度の高いセグメントから順番に選択する。

4. 予備実験

メモリアル動画を作成する際の重要度スコアの取得には、動画単位でのアノテーションやメモリアル動画を表現するだけのサンプル数を用意した本実験が必要である。本稿では本実験に備えた予備実験を実施した。予備実験では画像からの重要度スコアの取得を少量のデータを利用して、提案アルゴリズムの重要度推定までを実施した。以下に実験目的と実験方法を記す。

4.1 実験目的

予備実験の目的の1つとして、観光経路での風景において重要度に有意な差が出るかを検証する。また、2つ目の目的として、提案アルゴリズムにより正しく重要度スコアの判定が行えるかを調査する。以上の2点を予備実験の目的とする。

4.2 実験方法

予備実験ではセグメントを1枚の画像と見立て、50枚

表 1 回答者の観光経験

沖縄への観光経験	回答者数
行ったことがない	489
現在、滞在している	6
1年以内	60
2年以内	44
3年以内	34
4年以内	19
5年以内	32
6年～10年以内	106
それ以上前	210

の画像について各20件のアノテーションの取得をYahoo!クラウドソーシングを利用して行った。画像は2019年の8月頃に沖縄を運転した車に備え付けられたドライブレコーダーから撮影されたものを利用した。取得されたアノテーションに対して、3.5節で述べた重要度推定手法を用いてそれぞれの重要度について算出した。モデル作成には機械学習ライブラリのscikit-learn[15]を利用した。また、データセットの規模が小さいため、50枚の画像に対してLeave-one-out交差検証を実施した。

5. 実験結果

実験結果について、クラウドソーシングでのスクリーニング検査の結果、回答者のコメント、得られた重要度スコアを順に記す。続いて得られた重要度スコアから「沖縄らしさ」を推定した回帰分析の結果を記す。

5.1 スクリーニング検査

クラウドソーシングのスクリーニング検査として1000人を対象として沖縄への観光経験の有無を尋ねた。「沖縄に行ったことがない」と答えた回答者を除いた、489人をホワイトリストに追加し重要度を回答するタスクを割り当てた。回答者の分布を図1に示す。

5.2 クラウドソーシング回答者のコメント

予備実験の際にユーザには画像に対してなぜその重要度を付したのかについて任意でコメントを受け付けた。そこで得られたコメントをいくつか紹介する。

- ポジティブなコメント
 - ヤシの木が沖縄らしさを感じた
 - 広い空で爽快感がある
 - 海が見える
- ネガティブなコメント
 - どこにでもある普通の道路
 - 雑多な感じがする
- ドライブレコーダに関するコメント
 - フロントガラスの反射が気になる
 - 暗い

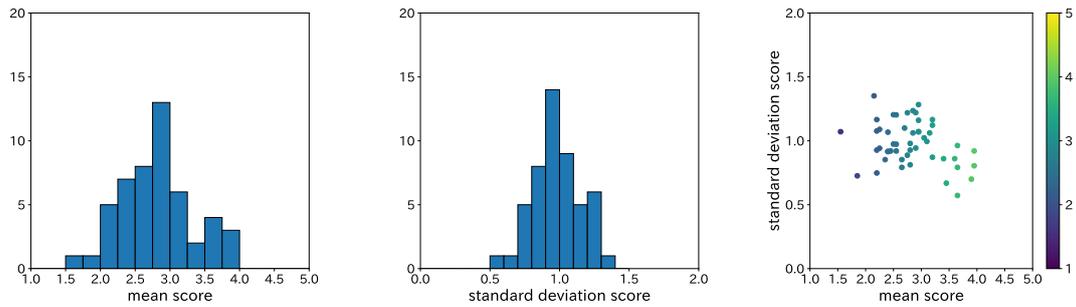


図5 「沖縄らしさ」のヒストグラム。左: 平均値のヒストグラム, 中央:標準偏差のヒストグラム, 右: 平均値と標準偏差のヒストグラムの結合。

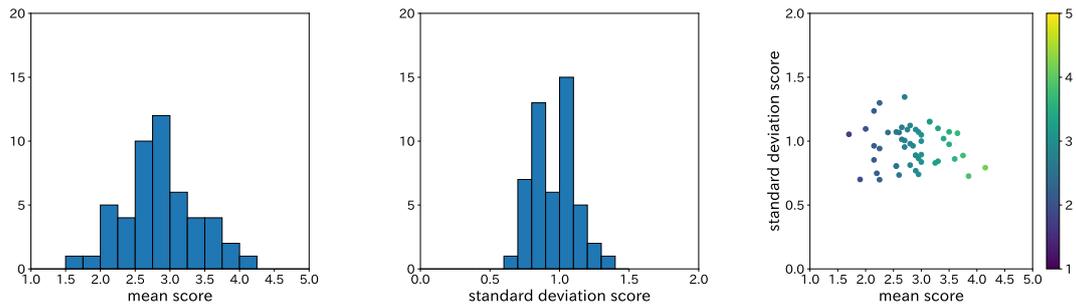


図6 「綺麗さ」のヒストグラム。左: 平均値のヒストグラム, 中央:標準偏差のヒストグラム, 右: 平均値と標準偏差のヒストグラムの結合。

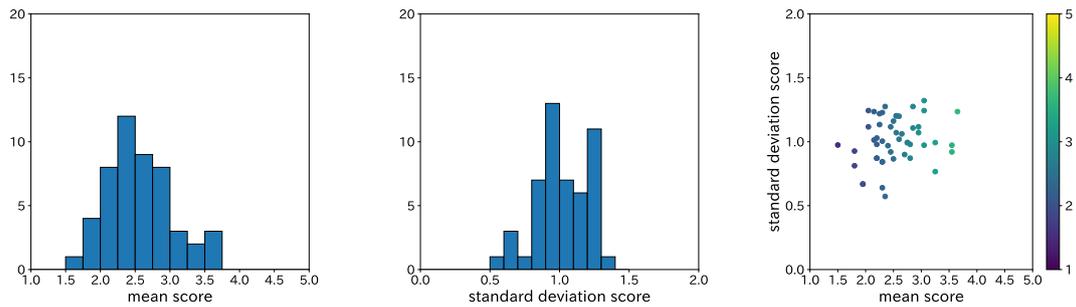


図7 「メモリアル動画への必要度」のヒストグラム。左: 平均値のヒストグラム, 中央:標準偏差のヒストグラム, 右: 平均値と標準偏差のヒストグラムの結合。



(a) No32, score:1.50 (± 0.97)



(b) No7, score: 3.65 (± 1.23)

図8 予備実験で用いた画像の例。 μ と σ は「メモリアル動画への必要度」について平均値と標準偏差を表している。(a) 最低スコア ($\mu = 1.50, \sigma = 0.97$), (b) 最高スコア ($\mu = 3.65, \sigma = 1.23$)

沖縄では京都の寺社仏閣や東京の東京タワーのように道路からでも見られる際立ったランドマークが少ない。その代わりポジティブな意見でも見られるような「ヤシの木」や「海」といったものが地域を表現するランドマークとして共通認識ができていることが得られた。またネガティブ

なコメントとして高速道路のような眺めは開けているがどこにでも見られるような風景や、建物に囲まれているような景色を指摘する意見がみられた。その他のコメントとしてドライブレコーダの性能に起因する意見も挙げられた。

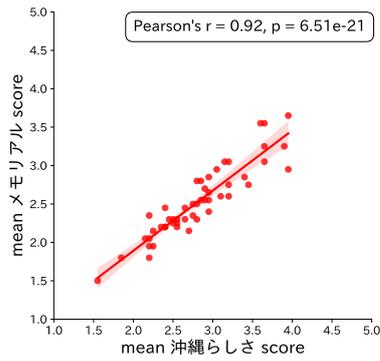


図 9 各スコアの分布と相関関係

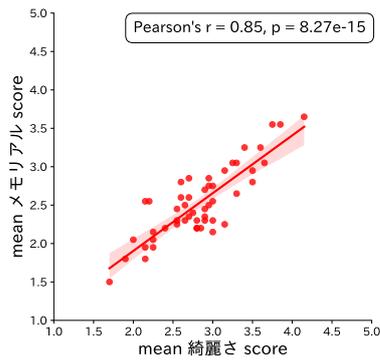


図 10 各スコアの分布と相関関係

5.3 重要度スコア

各画像に対して得られた 20 件の 3 項目への回答について、平均値と標準偏差を算出したものを図 5, 図 6, 図 7 に示す。「沖縄らしさ」や「綺麗さ」の重要度は最頻値は 2.5 ~ 3.0 にあるもので、正規分布に近い分布をしている。「綺麗さ」では 4.0 以上の重要度が得られる画像も存在した。一方で、「メモリアル動画への必要度」についてのヒストグラムでは最頻値が 2.0 ~ 2.5 に存在しており、分布が他の 2 項目と比較して左に偏っている。このことから「メモリアル動画に必要」と回答者が確信を持って回答できる画像は少ないことが推測できる。標準偏差についてはすべての項目において 1.0 に集中した分布になっている。これは、質問内容の曖昧性に起因していると考えられる。

図 8 は実験で用いた 50 枚の中で「メモリアル動画への必要度」の項目で最低、最高スコアを示した画像の例である。最低スコアを示した (a) はトンネル内部を運転する風景で、最高スコアを示した (b) の画像は、沖縄の観光地としても有名な古宇利島へ繋がる橋を渡る風景であり、左右には海がみられる。これらの画像からもクラウドソーシングで直感的な評価が取得できていると考えられる。

5.4 項目間での重要度の相関

算出された重要度について「メモリアル動画への必要度」との相関を算出したものを図 9, 図 10 に示す。相関はピアソンの積率相関係数 (r) と有意確率 (p) を用いた。赤色の直線は単純な線形回帰直線を周囲の薄赤色領域は 95% 信頼

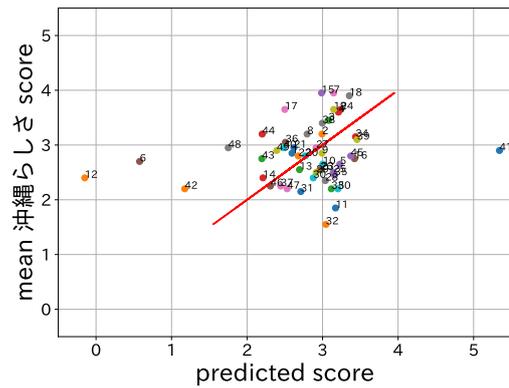


図 11 重回帰分析の結果

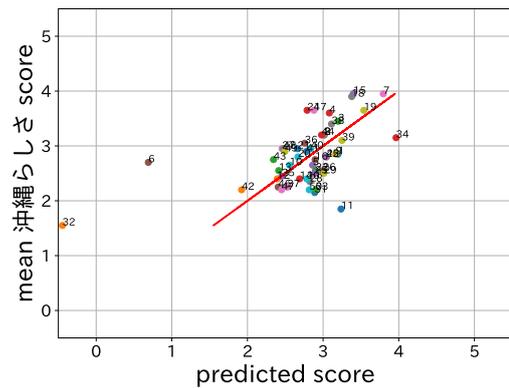


図 12 SVR の結果

表 2 「沖縄らしさ」推定結果

method	検定手法	決定係数
重回帰分析	leave-one-out	-1.48
SVR	leave-one-out	-0.20

区間を表す。

それぞれの項目間の相関係数より、正の強い相関が確認された。またどちらも有意水準 ($0.05 > p$) を満たしているため、有意に相関があることが認められた。ここで最も高い数値として、「沖縄らしさ」と「メモリアル動画への必要度」の項目間の相関係数は 0.92 であった。このことは「沖縄らしい」という指標を用いてその土地のメモリアル動画を作成することが有効であることを示す。また他の地域においても同様に「その土地らしさ」を利用することでメモリアル動画が作成できることを示唆した。

5.5 重要度の推定

図 7 の結果から「メモリアル動画への必要度」を回帰分析により推測することは、重要度の分布が全体的に低く偏っていることから難しいと考えられる。よって、回帰分析では強い相関関係が見られる「沖縄らしさ」の推定を行う。SVR ではハイパーパラメータチューニングは実施せずにデフォルト値を利用した。

表 2 に提案手法で述べた 21 次元の特徴量を用いて「沖縄らしさ」についての重要度の推定を行った際の決定係数

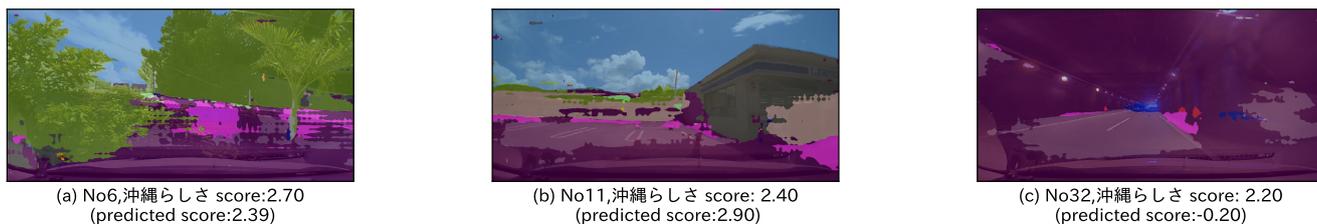


図 13 SVR 分析で特に誤差が大きかった画像の例

を、図 11, 図 12 にそれぞれの分析手法での推定値の散布図を示す。どちらの手法においても極端に誤差の大きいサンプルがみられるため、決定係数は低下したとみられる。散布図では、重回帰分析が回帰直線から離れたデータが多数見られるのに対して、SVR では比較的多くのサンプルについて回帰直線で表現できている。

6. 考察

本章では、予備実験の結果から得られた課題と改善方法について考察する。

6.1 クラウドソーシングでのホワイトリスト

今回の予備実験では、日本国内の沖縄への観光経験者をホワイトリストに追加し、クラウドソーシングのタスクを依頼した。したがって、重要度の判定基準などは日本人に依存している。しかし、近年レンタカーの利用者にはインバウンドの観光客も増加している。今後はメモリアル動画を作る対象の属性(性別や出身国)などを考慮してクラウドソーシングを行い、よりパーソナライズされたモデルを作ることが望まれる。

6.2 セマンティックセグメンテーションの検出カテゴリ

今回 DeepLabv3 で検出を行った 20 カテゴリには海のラベルは入っていない。しかし、沖縄の風景の重要度に海が影響することは予備実験のコメントから得られている。よって学習データを拡張して海のラベルを得る必要がある。ここでの関連研究として、土井らの研究 [16] では電車の車窓に対してセマンティックセグメンテーションを行っている。本研究と同様に CityScapes データセットを利用している。この研究では、さらに ADE20k データセット [17] で追加のセマンティックセグメンテーションを行い、水域の取得を試みている。今後はこの手法を用いて海領域の取得を目指す。

6.3 重要度推定において誤差が大きいサンプル

図 12 内で特に誤差の大きかった 6, 11, 32 番の画像について DeepLabv3 により出力されたセグメンテーション画像を図 13 に示す。(a), (b) は駐車中の画像である。これらは駐車中のため他の画像と比べた際に明らかに景色が

異なっている。したがって画像中に写るカテゴリから正しく推定できなかったと考えられる。これはドライブレコーダから車速を利用することで解消できると考える。

(c) はトンネルの中のシーンである。今回学習の際にトンネルを 1 つのカテゴリとして学習しておらず、検出の際に画像の大部分が Other カテゴリとして検出されている。Other には橋の脚など学習サンプルが多くないものが含まれている。また、今回の 50 枚の学習データの中でトンネル内部の画像はこの 1 枚であった。これにより (c) のようなカテゴリの占有率が特徴的な画像について正しく検出されなかったと推測できる。この問題への対処法として 1 つは学習データ量の拡充である。また別の方法として、画像中のカラーヒストグラムで閾値判定することも可能である。

これらの 3 画像を除いた際の SVR の決定係数は 0.36 であった。したがって今後は速度情報の追加や、データ数の拡張が必要であると考えられる。

6.4 重要セグメントの選択

今回の予備実験では重要度の推定までを行った。提案手法でも述べているように、メモリアル動画として作成される動画では重要度の高いセグメントからの選択を検討している。本実験ではメモリアル動画を視聴したユーザ評価を元にセグメントの選択手法についても改善をする必要があると考えられる。

7. 結論

本稿では、観光地でのレンタカー利用後の観光客に向けたメモリアル動画のためのドライブレコーダ動画キュレーションアルゴリズムを提案した。提案手法では画像処理手法であるセマンティックセグメンテーションや地域のランドマーク検出を利用したセグメントからの特徴量の抽出方法や機械学習による重要度の推定を提案した。次に、50 枚の小規模な学習データセットを用いた予備実験を行った。この予備実験では過去の沖縄への観光経験者を対象としてホワイトリストに追加し、クラウドソーシングにより画像のアノテーションタスクを依頼した。実験結果として、「メモリアル動画への必要度」について、「沖縄らしさ」が有意に強い相関関係を示すこと、推定モデルが検出誤差の大

きいサンプルを除くことで決定係数 0.36 を示すことを確認した。また、検出誤差の大きいサンプルの原因について考察した。今後は、データセットを拡張させて、動画からの重要度の取得をクラウドソーシングのタスクとして実行する。また、今後の課題としてはセマンティックセグメンテーションの検出ラベルの追加、セグメント選択手法の改善、作成されたメモリアル動画について視聴者の主観評価実験を行う。

謝辞 本研究を実施するにあたり、株式会社デンソー様よりドライブレコーダ動画の提供をいただいた。ここに記し感謝申し上げる。

参考文献

- [1] Ipsos MediaCT. “the 2014 traveler’s road to decision-google travelstudy”. Google, 2014.
- [2] Real Networks. Real times, 2020. <https://jp.real.com/realtimes/>.
- [3] JEITA・JDRC(一般社団法人ドライブレコーダー協議会). ドライブレコーダー国内出荷実績, 2018.
- [4] Kazuhito Takenaka, Takashi Bando, Shogo Nagasaka, and Tadahiro Taniguchi. Drive video summarization based on double articulation structure of driving behavior. In *Proceedings of the 20th ACM international conference on Multimedia*, pp. 1169–1172, 2012.
- [5] Kazuhito Takenaka, Takashi Bando, and Takayuki Miyahara. Drive video recording device and method, drive video recording system, and summarized moving image creating device, May 24 2016. US Patent 9,350,952.
- [6] 佐藤享憲. シーン文字認識と自己動作分類を用いた車載動画の要約. In *MIRU: The Meeting on Image Recognition and Understanding*, 2015.
- [7] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing*, Vol. 27, No. 8, pp. 3998–4011, 2018.
- [8] CISCO. “cisco visual networking index global ip traffic forecast,2017-2022”. CISCO, 2017.
- [9] Yale Song, Jordi Vallmitjana, Amanda Stent, and Alejandro Jaimes. Tvsum: Summarizing web videos using titles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5179–5187, 2015.
- [10] Yuki Kanaya, Shogo Kawanaka, Hirohiko Suwa, Yutaka Arakawa, and Keiichi Yasumoto. Automatic route video summarization based on image analysis for intuitive touristic experience. *Sensors and Materials*, Vol. 32, No. 2, pp. 599–610, 2020.
- [11] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [12] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223, 2016.
- [13] Fisher Yu, Wenqi Xian, Yingying Chen, Fangchen Liu, Mike Liao, Vashisht Madhavan, and Trevor Darrell. Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, 2018.
- [14] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [15] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [16] 土井亜香里 al. セマンティック・セグメンテーションにより抽出した景観構成要素に基づく視点のクラスタリング—大阪環状線の車窓シークエンス景観を対象として—. Master’s thesis, 大阪市立大学大学院, 2020.
- [17] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 633–641, 2017.