

dash-cum: ドライブレコーダを用いた メモリアル経路動画キュレーション

dash-cum: Dashcam Video Curation for Memorial Movie Generation

片山 洋平^{1,2} * 諏訪 博彦^{1,2} 安本 慶一^{1,2}

Yohei Katayama^{1,2} Hirohiko Suwa,^{1,2} Keiichi Yasumoto^{1,2}

¹ 奈良先端科学技術大学院大学

² 理化学研究所 革新知能統合研究センター

要旨: 近年、観光後に自身の経験をメモリアル動画として保存・共有するユーザが増えている。それに合わせて、メモリアル動画の作成支援のために、観光地で撮影された写真や動画を自動でキュレーションするサービスも提供されている。しかし、これらのメモリアル動画は、主に観光地で撮影された写真・動画を対象としており、観光において大きな時間を占める移動中の様子が欠落している。観光時における移動は、大半の時間は単なる移動時間ではあるものの、景色を楽しむなど旅行における重要な一場面を含んでいる。そのため、大量の移動動画から特定の場면을効率的に抽出することで、思い出を想起する有効なキュレーションに活用できると考える。そこで本研究では、移動を含めた観光全体のメモリアル動画自動キュレーションを目指し、昨今普及が著しいドライブレコーダを用いたメモリアル経路動画キュレーション手法を提案する。提案手法では、動画から取得できる画像特徴と走行データから取得できる車両特徴を用いて、重要な場면을自動で特定する重要度想定モデルを構築する。そのために、クラウドソーシングにより3秒動画（セグメント）ごとの重要度スコアを取得し、機械学習により構築したモデルにより重要度スコアを算出する。さらに、算出された重要度スコアに基づくキュレーションアルゴリズムを提案する。評価実験では、推定モデルの評価のために、クラウドソーシングと推定結果を比較した。その結果、画像特徴のみを用いた際にMAEが最も低く6.33となった。また、キュレーションアルゴリズムの評価のために、沖縄での観光実験に基づくキュレーション動画を作成し、観光者とのグループインタビューによる定性的評価を行った。その結果、提案手法により作成されたメモリアル動画が、既存の観光地のみで撮影されたメモリアル動画に比べて、思い出の想起に有効であることを確認した。

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、観光分野においても観光前、観光後における動画の利用が盛んになっている。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者自身が撮影した写真を再生するものがある。RealNetworks社の提供する「RealTimes」のように、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・提供するサービス[2]も存在する。ただし、これらが作成するメモリアル動画では複数の観光スポット周辺の写真をまとめるものを指

しており、観光移動経路に着目した動画の作成・支援を行うサービスは少ない。理由として観光における移動は長時間に及ぶ場合が多く観光者自身での記録が難しいこと、重要なシーンの位置を手で抽出することが難しいためであると考えられる。観光において大きな時間を占める経路に関する動画をキュレーションし提供することは、観光の流れを直観的に想起・共有する手助けになると共に、観光中に現れた重要なシーンを省くことなく伝えることが可能となる。

観光における移動はバス・電車、歩行など様々あるが本研究では特にレンタカーのような車での移動を対象とする。ここで、車での観光移動におけるメモリアル動画の生成には近年、普及が広まりつつあるドライブレコーダを利用する。ドライブレコーダの国内出荷数は増加しており、2018年度は年間約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

*連絡先: 奈良先端科学技術大学院大学先端科学技術研究科
〒630-0192 奈良県生駒市高山町 8916 番地 5
E-mail: katayama.yohei.kr7@is.naist.jp

設置率は上がると考えられるが、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、取得される動画から観光メモリアル動画として自動キュレーションするアルゴリズムを構築することを目指す。観光経路のメモリアル動画の活用場面は、観光予定者に向けたマーケティングやプランニング、観光者自身の振り返りや SNS でのシェアリングなど、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視聴者に動画を提示することが望ましいと考えられる。ここで車での移動経路動画を自動キュレーションする上での課題として、代わり映えのしない道のりや信号での停車、渋滞など視聴者にとって冗長な部分が多く動画から見どころとなる想い出深いシーンを抽出すること困難であるが挙げられる。

ドライブレコーダ中の想い出深いシーンをメモリアル動画として残すため、人手で全ての時間を確認することは現実的に難しい。したがって各シーンの重要度を自動で推定し、メモリアル動画に残す価値のあるシーンが抽出されることが望ましい。本研究では、観光メモリアル動画における重要度推定モデルを構築し、セグメント（3秒単位）動画に対して重要度を推定することで見どころ（キーセグメント）を自動キュレーションするアルゴリズムを提案する。重要度推定モデルの構築にはクラウドソーシングによりセグメントの重要度の取得を行う。筆者らの先行研究 [4] では沖縄観光を対象としたメモリアル移動経路動画の生成のためにクラウドソーシングを実施し、シーンの重要度を取得する際の適切な指標について検証した。そこでは、“メモリアル動画への重要”さという指標は、タスクの回答者にとって曖昧な質問となり高いスコアを得ることができないことが確認され、上記の指標と相関係数 0.92 と強い正の相関を示した“沖縄らしさ”という指標が適切であることを示した。本研究でも先行研究と同様に沖縄観光を対象とし、クラウドソーシングでは“沖縄らしさ”によるセグメントの重要度を取得する。ここでセグメントの重要度推定モデルの特徴として画像特徴、車両情報を利用する。画像特徴では地域のランドマークを物体検出モデルにより検出し、自動運転などのタスクのために作成されたセマンティックセグメンテーションモデルを用いて画像に占める各カテゴリの割合を算出した。車両特徴ではドライブレコーダから車速などの車両情報をする。これらの特徴を用いて回帰モデルの LightGBMRegressor を利用して推定する。また、推定されたセグメントのスコアからメモリアル動画を作成する際のセグメント選択には重要度の上位から選択する TopDown 法に加え多様性を持った選択を行う手法として KTS 法、Similarity 法を検討する。

提案キュレーション手法では重要度推定モデルで高いスコアを得たセグメントからメモリアル動画として

抽出するため、推定モデルの性能が非常に重要となる.. よって評価実験では実際に沖縄のレンタカー会社により取得されたドライブレコーダの動画とクラウドソーシングにより得られたスコアを利用してモデルの推定誤差の評価を行い、評価指標である MAE において 6.33 という結果が得られた。また、推定されたスコアを用いたキュレーション手法の有効性の検討を行うため、観光経験者によるセルフディスカッションによる考察結果を後述する。

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。3 章では、本研究で作成されるメモリアル動画の要件について述べ、4 章でメモリアル動画を作成する一連の自動キュレーションのフローについて、5 章ではキュレーション手法の内部で利用する重要度推定モデルの詳細について述べる。6 章では重要度推定モデルの評価実験について、7 章推定モデルの結果についての議論を及びキュレーション結果に関する有効性の検討を行う。最後に、8 章で本論文のまとめを述べる。

2 関連研究

観光におけるキュレーションの既存研究の多くは、観光地を対象としたものが多い。一方で複数の観光地を訪問する際には経路を移動する時間が観光の時間の多くを占める。本章では経路動画を対象としたキュレーションの研究を紹介する。経路動画のキュレーションの目的は主に 2 つに大別され、1 つは今後経路を移動するユーザに向けた推薦のためのキュレーションであり、他方は移動を終えたユーザへのフィードバックを対象としたものである。

2.1 推薦のための移動動画キュレーション

観光前のユーザに向けた研究では、交差点や風景が急激に変わるシーンが重要であるとしてキュレーションが行われている。また、運転や歩行経路を正しく伝えることを目的としているためキュレーション動画ではカット処理を施さず、部分的に動画の再生倍速を変える手法が取られている。直感的な歩行経路を提示するための、1 人称映像をキュレーションする研究 [5] ではヒストグラム差分から風景の切り替わりを検出し、VFR(可変フレームレート)方式でキュレーションを行っている。また、キュレーション動画により自動車での運転経路をユーザに提示する研究 [6] では、ドライブレコーダ映像中から動き特徴であるオプティカルフローの抽出と機械学習による道路案内標識を検出を組み合わせることで動画のキュレーションを行っている。これらの研究によりキュレーションされる部分は観光経路を正し

く伝えるという目的はあるものの、想い出深い動画をユーザに提供するという目的には合致しない。

2.2 振り返りのための移動動画キュレーション

運転手の運転技術向上のために、自動車で運転後の経路動画をキュレーションした研究 [7] では、ドライブレコーダの映像から運転動作を抽出し、動作に対応した音楽を合わせてキュレーション動画を作成している。これはユーザの運転技術や運転マナー向上を目的としたフィードバック動画となっており、本研究の対象とするユーザの想い出の振り返りとなる経路動画キュレーションとは目的が異なる。

3 メモリアル運転経路動画の要件

本研究ではドライブレコーダから取得された移動経路動画を用いたメモリアル動画のキュレーションを行う。ユースケースには沖縄でレンタカーを借りて1日観光した際のメモリアル動画を想定する。メモリアル動画は視聴対象者や目的によって求められる内容が異なるため、本研究が作成するメモリアル運転経路動画の要件を以下に示す。

要件 1 観光の流れを想起・共有できる動画であること

要件 2 観光地域を表現する風景を残すこと

要件 3 観光地で撮影された写真と組み合わせられること

観光時の真に想い出深いシーンはその観光を経験した者のみが知りうる。それらのシーンは運転時の景観の良し悪しだけでなく、観光とは関係のない車内の会話の内容からも定まる。これらのシーンは個人で想い出を振り返る上で重要となる。一方で、近年では観光のメモリアル動画を個人で楽しむだけでなく、SNSなどで他人に共有する楽しみ方 [8] や観光地のマーケティングとしてのメモリアル動画の活用もなされている。本研究では、要件 1 にあるようにメモリアル動画の視聴対象者を旅行者に限定せず、旅行者がどのような流れで観光を行ったかを想起・共有できる動画を生成する。

要件 2 に関して、メモリアル動画において地域の特徴を表す風景が重要であるとして、横浜みなとみらい区を対象として予め定義されたランドマークのデータベースと画像認識の結果から想い出動画の自動生成を行っている高橋らの研究 [9] がある。本研究でも同様に観光地域の特徴的な風景がメモリアル動画で必要であるとする。高橋らの研究では海や空といった非オブジェ

クトの重要性について考慮できていない。本研究が検討する沖縄ではこのようなランドマークに限らない非オブジェクトの景観について考慮する必要がある。

最後の要件として、観光地で撮影された写真と組み合わせられるメモリアル運転経路動画であることとする。メモリアル動画において観光地の写真は重要である。したがって、本研究で提案するメモリアル運転経路動画では観光地間の移動を1つのセッションとして、各セッションごとに移動経路のメモリアル動画を生成する。

3.1 ドライブレコーダを用いた経路動画収集

自動車での移動経路のメモリアル動画を作成する際に、本研究ではドライブレコーダ動画を活用する。ドライブレコーダから得られるデータは事故発生時の対応に使われることが主で、その他の活用は一般的ではない。一方でドライブレコーダは自動車から配給される電力により常時撮影が可能であり、移動経路の動画を収集する上で適している。本研究では観光客が利用したレンタカーから取得される動画を用いてメモリアル動画の作成を行う。また、ドライブレコーダには車外だけでなく車内の動画も取得することができる機種も市販されているが、車内のコンテキストにおける想い出の重要度は盛り上がりや会話内容などに依存しており、シーンの重要さ以外の指標が必要となる。よって本研究では車外の動画を用いたメモリアル動画の作成をスコープとする。

3.2 クラウドソーシングを用いた重要度スコア取得

運転経路におけるシーンの重要度は実際の観光終了後に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移動動画を観光者が全て評価するのは現実的に難しい。そこで本研究ではドライブレコーダ動画を3秒単位のセグメントに分割し、セグメントごとにクラウドソーシングを用いて重要度を取得する。セグメントの詳細は4.1節で示す。経路動画において道案内をする上で重要なシーンと、メモリアル動画を作成する上で重要なシーンは異なる。本研究では特に沖縄観光時のドライブレコーダ映像からメモリアル動画を作成する上でのシーンの重要度を評価するため、以下の3指標をクラウドソーシングから取得する。

- 沖縄らしさ (Okinawan)
- 綺麗さ (Beautiful)
- メモリアル動画への必要度 (Memorial)

4 提案キュレーション手法

本研究で提案するキュレーションアルゴリズムの概要を図1に示す。提案アルゴリズムでは入力動画をセグメントと呼ばれる3秒ごとの動画に分割し、各セグメントの重要度を推定する。続いて推定された重要度からメモリアル動画を作成する。入力には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、出力として対応するメモリアル動画を生成する。またドライブレコーダが取得した車速や加速度などの車両情報をキュレーションに利用する。

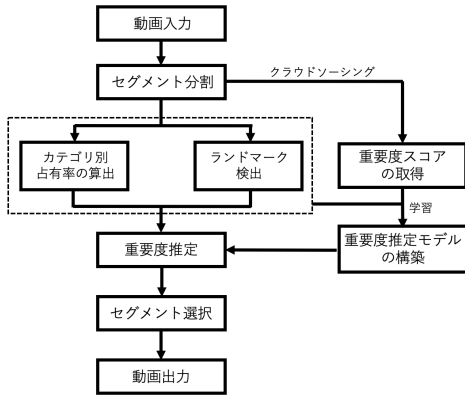


図1: 提案手法の概要図

4.1 セグメント分割

本来、クラウドソーシングの回答者は観光中のドライブレコーダ動画を全て確認した後に、その中から重要なシーンを判定するのが好ましい。一方で1日の観光全てのドライブレコーダを全て目視で確認することは現実的でない。そこで提案手法では、ドライブレコーダ動画をセグメントと呼ばれる短時間の動画に分割し、各セグメントの重要度を推定しキュレーションを行う。多くの人々にとって視覚情報は3秒間で安定期に入ると言われている[10]。したがって本研究でもセグメントの長さの単位に3秒を採用している。また、3秒は車の時速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り車外の風景の変化を確認するのに十分な時間であると考えられる。ここで、重要なセグメントに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。

4.2 重要度推定モデル

本節では重要度推定モデルの概要を記述する。詳細については5章で述べる。キュレーションを行う上で動画中からメモリアル動画に残す重要度の高いセグメント(キーセグメント)を推定することは重要である。本研究ではランドマークの検出とセマンティックセグメンテーションを組み合わせることでセグメント内での景色の特徴を抽出し、重要度の推定を行うモデルを構築する。

各セグメントの重要度について考察した片山ら[4]の研究では、1セグメントあたり20人のクラウドソーシングの回答者から5段階で得た”メモリアル動画での必要度”という指標のみではクラウドソーシングからキーセグメントを抽出することが難しいことを示した。また、”メモリアル動画への必要度”に代わる指標として”沖縄らしさ”が相関が高く代替指標として適していることを示した。クラウドソーシングでは景色の間でスコアに大きな分散が現れないため、推定モデルをそのまま学習することは困難である。Querciaらの研究[11]では、クラウドソーシングで得た評価値の3次関数により景観のスコアを定義し、モデルによるスコア推定を行った。本研究でも同様にクラウドソーシングで得た5段階の沖縄らしさ s_i について式1により定義するokinawan scoreを算出し、重要度推定モデルでこのスコアを学習・推定する。

$$\text{okinawan score} = \left(\frac{\sum_{i=1}^n s_i}{n} \right)^3 \quad (1)$$

4.3 重要度を用いたセグメント選択

推定されたセグメントごとの重要度からどのようにキーセグメントを選択するかについて述べる。本研究ではTopDown法、KTS法、Similarity法の3手法により選択するセグメントを比較する。メモリアル動画はエンジンのオンからオフまでを1つの観光地間の移動とみなし、この移動間の経路動画からメモリアル動画を作成するものとする。また、車の出発、到着時は観光の流れをメモリアル動画に残す上で重要であるとして、以下のセグメント選択アルゴリズム全てに共通してこれらのセグメントはキーセグメントとする。

TopDown 法

TopDown法では全セグメントを降順にソートし、スコアの上位セグメントからキーセグメントを選択する。これは重要度が高い動画を実際にメモリアル動画に残すことが可能である一方で、スコアが高いセグメント

が時系列で隣接する際に特定の時間帯のセグメントのみしか選択ができない問題が存在する。

KTS 法

KTS 法ではスコアが高いセグメントは時系列上で隣接するという考えに基づいて選択を行う。従来から動画要約に用いられてきた Kernel Temporal Segmentation (KTS) と呼ばれる手法 [12] を用いて動画を時系列で分割を行う。KTS ではセグメントごとの特徴量の変化からドライブレコーダ動画で大きく風景が変わった区切り目を推定する。推定された区間ごとにキーセグメントを選択することにより、隣接するセグメントが選択されない。

Similarity 法

Similarity 法ではスコアが高いセグメント同士の景観の類似度は高い傾向にあるという考えに基づいて選択を行う。スコア上位からセグメントの選択を行い、すでに選択されたキーセグメントと \cos 類似度が閾値以上の時に対象セグメントを選択しないようにする。これにより内容の多様性があるメモリアル動画を作成する。

出力動画では、全てのセグメントを残しつつキーセグメントのみの動画再生速度を落とす可変フレームレート方式と、キーフレームのみをクロスフェードを用いて再生するカット方式のそれぞれで作成する。可変フレームレート方式では経験則によりキーセグメント部分の再生速度を 2 倍速に、それ以外の再生速度を 16 倍速となるように生成する。

5 重要度推定モデルの構築

本章ではメモリアル運転経路動画におけるセグメントごとの重要度を推定するためのモデルについて述べる。はじめにモデル構成について図 2 に示す。提案手法のモデルでは 3 秒動画であるセグメント (Segments) ごとに主に 2 種類の特徴を抽出する。1 つめはフレーム単位の特徴 (Frames) である。ここでは地域のランドマークや、セグメントに含まれるカテゴリを取得する。2 つめは車両情報 (Car Info) である。ここでは車速や加速度といった運転に関する情報を用いる。これらの特徴量から帰帰モデルを生成し、重要度の推定を行う。

5.1 画像からの特徴抽出

本研究では 3 秒のセグメントから画像解析のため 1 秒あたり 1 枚のフレーム画像を抽出する。本来はセグ

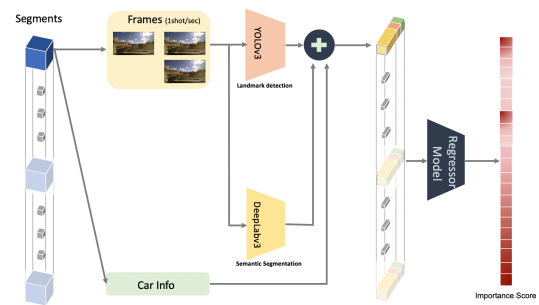


図 2: 重要度推定モデルのアーキテクチャ

メント内の全てのフレーム画像から特徴を抽出することが好ましいものの、隣接するフレームは情報がほとんど重複すること、解析における時間を減らすことを考慮して本研究では 1 セグメントあたり 3 枚の画像をサンプリングし、これらの画像から特徴を抽出する。それぞれの画像から抽出された特徴について統計量 (平均、最小、最大) を算出し特徴量とする。

5.1.1 地域のランドマーク検出

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。先行研究 [4] で受け付けたコメントの結果から、沖縄の場合にヤシの木が地域のランドマークとして重要度に影響しているとの回答を得た。そこで提案手法では物体検出で高い精度を出している YOLOv3 (darknet53) [13] を利用し、ヤシの木の検出を行う。YOLOv3 は畳み込み層を含む 53 層から構成される。本研究では YOLOv3 の出力層でヤシの木のバウンディングボックスの位置、サイズ、確信度を出力するようにファインチューニングを行う。学習時には実際にドライブレコーダで確認された 80 枚の画像の 90% を学習用に、10% を検証用に利用する。

画像に映るヤシの木の大きさは重要度に影響すると仮定し、検出されたバウンディングボックスについて確信度が 0.5 以上のものを対象に以下のようにビンニング (ビン分割) 処理を行う。ここで画像サイズに対して 5% 以下を Small, 5%~10% を Medium, 10% 以上を Large とする。ビンニングにより算出されたそれぞれのスケールでの検出数を特徴として利用する。

5.1.2 風景のセマンティックセグメンテーション

セグメントの重要度において画像にどのような風景が映るかは特に重要である。そこで本研究では画像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) と呼ばれる処理を行うことでドライブレコーダが捉えている風景を推

定する．提案手法では路上風景の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[14] に WideResNet38 を組み合わせたモデルを利用する．モデルは Rotabulo ら [15] によって構築され，オープンデータセットである Mapillary Vistas Dataset[16] で学習されたものを利用する．

画像は車，道路，空，海などを含む 65 カテゴリに画素単位で分類される．画像における各カテゴリの存在する画素数からカテゴリ別の画像内占有率を算出する．本研究で用いた全ドライブレコーダ映像におけるそれぞれのカテゴリの占有率のひげ鬚図を??に示す．ここで snow などのいくつかのカテゴリでは沖縄のドライブレコーダ映像に登場していないことが読み取れる．これらのカテゴリは沖縄を対象とした場合にほとんど現れず回帰モデル構築の際にノイズとなりうる．したがって全データのうち 95 パーセンタイル部分のカテゴリ占有率が 0 となった 26 カテゴリは特徴として使用せず，49 カテゴリの占有率を利用する．

5.2 運転車両に関する特徴抽出

先行研究で受け付けたコメントから車速が遅く渋滞しているセグメントは重く退屈な印象を与えるとされた．同様にスピードが出ているセグメントからは爽快な印象を与えるとされた．これらから車の速度や運転に関する情報はセグメントの重要度にも影響すると考えられる．本研究ではドライブレコーダから車両に関する情報として 3 軸加速度および車速を取得する．特徴にはセグメント内のこれらの値について平均値を算出した値を利用する．

5.3 回帰モデルによる重要度推定

画像から得られた特徴，および運転車両に関する特徴から沖縄らしさの重要度推定モデルの構築を行う．推定にはベースラインの回帰分析モデルとして重回帰分析を，高精度な回帰モデルとして LightBGMRegressor を利用した推定を行う．

6 重要度推定モデルの評価

提案キュレーション手法では，重要度推定モデルにより高いスコアを得たセグメントがキーセグメントとしてメモリアル動画に残る．したがって重要度推定モデルの性能が重要となる．本章では沖縄での観光実験で取得されたドライブレコーダ映像とクラウドソーシングにより得られたスコアを用いて推定モデルを構築し，性能を評価する．評価指標として沖縄らしさにつ

いての推定値から MAE (Mean Absolute error) を算出する．以下に実験環境と実験結果を記す．

6.1 実験環境

実験では 2019, 2020 年の 7, 8, 11 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から撮影された映像を利用した．ドライブレコーダには株式会社デンソーテンの G500Lite を利用する．この機種では車内カメラを同時録画させる場合と，車外カメラのみで録画する場合に fps が異なる仕様となる．セグメントに対する重要度の取得には Yahoo!クラウドソーシングを利用して行う．ここで取得されたスコアから算出される *okinawan score* 対して，4 章で述べた重要度推定モデルを用いて重要度を推定する．また，データセットの規模が小さいため，エンジンのオンからオフまでを 1 セッションとして Leave-one-session-out 交差検証を実施した．

6.2 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．表 1 に提案手法で述べた重要度推定モデルを用いて *okinawan score* についての推定を行った際の MAE を示す．また，図 3 に推定モデルでの推定値とクラウドソーシングのスコアを軸にとった散布図を示す．MAE では画像特徴のみを用いた際が最も MAE が低い値となった．この要因として実際の車両の速度や加速度はスコアに影響しないことが考えられる．

散布図では，10 から 30 のスコアあたりが最もサンプル数が多くなる．これは移動経路の大半のセグメントは沖縄らしさを感じるような重要なセグメントが少ないことを示している．メモリアル動画でキーセグメントとして選択される可能性が高い，推定値が高いセグメントに関してもクラウドソーシングのスコアと大きく離れていないことが読み取れる．

表 1: *okinawan score* の推定結果

特徴	MAE	MAPE
車両特徴のみ	9.23	89.86%
画像特徴のみ	6.33	50.97%
画像・車両特徴	6.58	55.40%

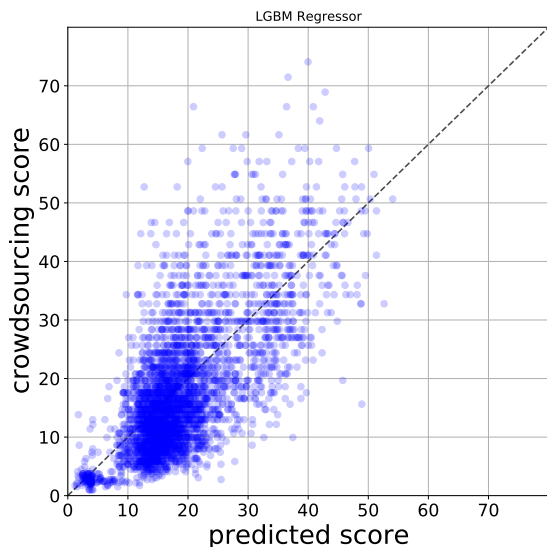


図 3: LightBGMRegressor による推定結果の分布

7 考察

本章では、6 章の評価実験の結果に関する考察および、実際に推定されたスコアを用いて提案キュレーション手法から作成される動画のメモリアル動画としての有効性を評価するため実施したセルフディスカッションについて記述する。

7.1 重要度推定モデルに関する考察

重要度推定モデルの性能についての考察を記述する。図 4 に評価実験で高い *okinawan score* を推定した画像を示す。海沿いや海に挟まれた橋を走行している風景や、椰子の木が写り込んでいる風景などが挙げられている。これらの画像はクラウドソーシングでも高いスコアを示しており、メモリアル動画を作成するうえでもキーセグメントとなりうると考えられる。一方でクラウドソーシングの結果とモデルの推定誤差が大きなサンプルもいくつか存在した。まず、クラウドソーシングでスコアが高かったものの、モデルの推定値が低かったセグメントではシーサーの石造のようなデータセット全体で見てもそのサンプルでしか現れないような特徴的なオブジェクトが写っているセグメントや、停車中に沖縄の地名が書かれた看板が写り込んだセグメントが存在した。石造などの特徴的なランドマークに関してはランドマーク検出時のデータセットを拡張することで対応可能であると考えられる。看板に関しては本研究が対象とする運転経路メモリアル動画の必要要件を満たさないため、クラウドソーシング実施に文字情報を考慮しないスコア付けを回答者に求める必要性が明らかとなった。また、クラウドソーシングの

スコアは低いものの推定値では高いスコアを得たセグメントではヤシの木が中央分離帯に植えられている道路などが含まれた。これらのセグメントに関して本研究では沖縄の特徴表す風景でメモリアル動画としては重要であると要件で述べている。したがってクラウドソーシングのスコアの確かさについて再度検証する必要があると考えられる。



図 4: 重要度推定モデルで高いスコアを示した画像

7.2 キュレーション結果の有効性に関する考察

重要度推定モデルにより推定された重要度を用いて作成されたメモリアル動画の有効性に関して考察したセルフディスカッションの結果を記述する。ディスカッションは筆者を含めた 3 名で実施し 1 名は運転者、残り 2 名は同乗者である。はじめにメモリアル動画として TopDown, KTS, Similarity それぞれの手法により選択されたキーセグメントおよび観光地の写真を含んだ動画に関して評価を得た。結果 TopDown 法では時間的に近く景観も似た風景がメモリアル動画として選択されていて冗長な内容であるとの意見を得た。一方で KTS, Similarity 法ではメモリアル動画に含まれる内容が移り変わるため見ていて飽きが生じにくいとの意見を得た。これらから単純に推定スコアの高いセグメントを選択するだけでなく、内容の多様性を含ませることは改めて重要であると考えられる。

また、今回のディスカッションでは観光地の写真だけをスライドショー形式で再生するメモリアル動画、観光地の写真に加え可変フレームレート方式で全てのセグメントを用いてキーセグメント部分のみ動画再生倍速を変更したメモリアル動画、観光地の写真に加えカット方式でキーセグメントのみをフェードで再生するメモリアル動画それぞれを観賞し評価を得た。観光地の写真のみのメモリアル動画と比較して移動経路を含んだ 2 種類のメモリアル動画では、観光の流れを直感的に思い出すという意味で非常に有効であるとの意見を得た。一方で可変フレームレート方式では全てのセグ

メントを利用するため動画の再生時間が20分を超えるものとなり、短時間で集中して動画を観賞するには適さないとの意見を得た。カット方式のメモリアル動画では3分程度の短時間のメモリアル動画となったためSNSでの共有や家族間、友人間で観光のメモリアル動画を振り返る上で最適でもあるとの意見を得た。

また、今回は沖縄らしく重要となりうる景色がメモリアル運転経路動画として適しているとしてキュレーションを行った。セルフディスカッションでは本手法により選択されたキーセグメントはメモリアル動画として十分受け入れられるものであるとされた、一方でメモリアル動画として良い動画という基準は個人によって異なることもわかった。今後の展望として、より個人にパーソナライズされたメモリアル動画を作成できるような動画編集支援システムとしての応用の可能性があると考えられる。

8 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。提案手法ではドライブレコーダから動画を取得し、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画像特徴および車両特徴を利用した重要度推定モデルを構築した。また、推定された重要度からメモリアル動画に残すキーセグメントの選択手法としてTopDown法、KTS法、Similarity法の3手法を提案した。全ての選択手法において推定される重要度が強く影響するため、重要度推定モデルの性能評価実験を実施した。その結果、画像特徴および車両特徴を用いたスコア推定では画像特徴のみを用いた際にMAEが最も低く6.33となった。実際に評価実験で構築した推定モデルを用いてキュレーションされたメモリアル動画の有効性を検討するディスカッションでは、キーセグメントとして選択されたセグメントはメモリアル動画として受け入れられる景色であったとし、観光地写真だけを再生するメモリアル動画よりも明らかに観光の流れを想起する助けとなったとの回答を得た。今後の課題としては、データセットの拡張による重要度推定モデルの改善、より多くの視聴者を対象としたメモリアル動画の主観評価実験が挙げられる。

謝辞

本研究を実施するにあたり、株式会社デンソーテン様よりドライブレコーダのデータ提供をいただきました。ここに記し感謝申し上げます。

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] 諏訪博彦 伍洋 安本慶一片山洋平. 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. マルチメディア, 分散, 協調とモバイル (DICOMO 2020) シンポジウム, 2020.
- [5] 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.
- [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] 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.
- [8] JTB Publishing. たびのび, 2018. <https://tabinobi.jp/>.
- [9] 高橋祐介, 佐藤貴美, 平田恭二. 車載カメラ映像中のオブジェクト認識方式と認識結果に基づく自動要約への応用. 電子情報通信学会論文誌 A, Vol. 88, No. 2, pp. 110–121, 2005.
- [10] Timothy F Brady, Talia Konkle, George A Alvarez, and Aude Oliva. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, Vol. 105, No. 38, pp. 14325–14329, 2008.
- [11] Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pp. 116–125, 2014.

- [12] Danila Potapov, Matthijs Douze, Zaid Harchaoui, and Cordelia Schmid. Category-specific video summarization. In *European conference on computer vision*, pp. 540–555. Springer, 2014.
- [13] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [14] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [15] Samuel Rota Bulò, Lorenzo Porzi, and Peter Kotschieder. In-place activated batchnorm for memory-optimized training of dnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [16] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kotschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4990–4999, 2017.