

ドライブレコーダを活用した 旅の要約動画自動生成システム

尾頭 花奈 石黒 祥生 椎尾 一郎 武田 一哉

本研究では、ドライブ旅行の思い出を楽しく振り返ることを目的に、ドライブレコーダのデータを活用した旅の要約動画自動生成システムを提案する。最近のドライブレコーダは、車内外の様子を高解像度動画として記録するため、紀行ムービーの素材として活用できる。ドライブレコーダで撮影された映像データは、メモ리카ードから簡単に入手できるが、膨大なデータを編集するには多くの時間と労力が必要なため、ほとんど利用されていない。そこで、名所や会話の盛り上がりなど、旅行において重要だと考えられるシーンを判定して動画を要約し、文字テロップ・地図・合成音声の追加を行う Web システムを開発した。また、手動編集と自動編集との比較評価を行った結果、提案システムの利用に必要な手順は少なく、システムは手動編集に比べて大幅に短い時間で要約動画を生成することがわかった。さらに、アンケート調査の結果、自動生成された動画に対するユーザの満足度は高いことがわかった。本研究は、今まであまり注目されてこなかった自動車乗車後の UX 向上に貢献する。

We propose an automatic memory movie editing system using dashcam video data to summarize road trips and enjoy looking back at the road trip memories. While it is easy to obtain the high-resolution video data recorded by dashcams using Secure Digital memory cards, this data is rarely used except for safety purposes because it takes substantial time and effort to review or edit many hours of such recorded videos. The proposal system summarizes the video by selecting the important scenes in a trip, such as famous places and the excitement of conversation, and adds text messages, maps, and synthetic voices. In this paper, we describe the design of an automatic editing system implemented on the Web, the results of comparisons between manual and automatic editing, and the results of the user questionnaire. This research contributes to improving the driving experience by providing entertainment for automobile users after road trips.

1 はじめに

スマートフォンやウェアラブルカメラの登場により、個人の体験を映像として簡単に記録できるようになった[10][15]。また、SNS (Social Networking Service) の普及により、体験を他者と共有し、コミュニケーションをとる機会が増えている。自動車にも多数のカメラやセンサ類がついており、自動車乗車中の体験が細かく記録されている。しかし、これらの

センサ情報は交通安全以外の目的にはあまり利用されていない。一方で、高度安全運転支援技術や自動運転技術の進展に伴い、今後、車両内外の情報を収集するセンサはさらに増えていくと考えられる。

ここ数年、車両から得られる映像データやセンサ情報をエンターテインメントシステムに利用しようという試みがなされている[13][14][16]。しかし、これらは乗車中の UX (User Experience) 向上を目的としており、乗車後の UX 向上に焦点を当てた研究は極めて少ない。そこで本研究では、旅行の振り返りに焦点を当て、乗車中のセンサ情報を活用して降車後の UX 向上を測る。

本研究では、近年急速に普及しているドライブレコーダに着目した[19]。ドライブレコーダの主な用途は交通事故や煽り運転車との遭遇のような非常事態への対策である[17]。そのため、ドライブレコーダに

Automatic summarization system for road trip video using dashcam video data.

Kana Bito, 名古屋大学, Nagoya University.

Yoshio Ishiguro, Kazuya Takeda, 名古屋大学/株式会社ティアフォー, Nagoya University/Tier IV, Inc..

Itiro Siiro, お茶の水女子大学, Ochanomizu University. コンピュータソフトウェア, Vol.39, No.4 (2022), pp.144-157. [研究論文] 2022年2月28日受付.

記録されたデータは通常はほとんど活用されず、記録媒体の容量を超えた場合、古いデータから逐次削除される。最近のドライブレコーダはGPS (Global Positioning System) 機能を搭載したモデルや複数台の高品位カメラを搭載したモデルも多く[11]、ホームビデオやSNSで共有する動画の素材として最適である。

しかしながら、長時間の走行動画を単に再生するだけでは冗長なシーンが多いため、楽しい動画を作成するには適切な編集が必要である。ドライブレコーダから得られる大量の動画ファイルから印象的なイベントが記録された箇所を選択し、結合し、エフェクトを追加するには多くの時間と労力が必要であるため、これら全てを毎回手作業で行うのは困難である。

そこで本研究では、旅行の要約動画生成に必要なユーザの手間と労力を削減するため、ドライブレコーダのデータからドライブ時に特化したイベントを検出し、車旅行の思い出動画を自動生成するシステムを開発した。今回作成したシステムは、自動車から得られる映像データの新しい活用アイデアをサポートし、これまであまり検討されてこなかった自動車乗車後のUX向上に貢献する。

本システムのプロトタイプについては、第69回UBI研究発表会[2]とAutomotiveUI'21[1]で発表した。本稿ではこれを発展させ、一般公開用に開発したWebサービス全体の設計とその評価について述べる。

2 関連研究

2.1 ドライブレコーダの活用

ドライブレコーダのデータを交通安全に活用する試みは多数存在する。例えば、ドライブレコーダを交通事故発生予測に活用する研究や[4][23]、路上駐車のリアルタイム検知に活用する研究[18]、注意散漫なドライバーを自動検出する研究[5]などがされている。また、内蔵されたAI (Artificial Intelligence) と加速度計のデータを用いて、リスクの高い行動やインシデントを検出してドライバーに警告するシステムを備えたドライブレコーダが販売されている[21]。本研究では、これらの取り組みとは違い、ドライブレコーダのデータをエンターテインメントに活用することを目

的とする。

2.2 動画の自動要約技術

動画の自動要約に関する研究は多数されてきた。例えば特異値分解 (SVD) とクラスタリングによる映像要約技術や[8]、シーンの変化をグラフのモデル化により検出する動画要約技術[20]、ビデオフレームからの色特徴抽出とk-meansクラスタリングアルゴリズムに基づいた動画要約技術[6]などがある。これらの研究では主にシーンの変化に着目し、要約を行っている。しかし、車内映像に関しては、自動車に乗車中の人の動きは少ないため、視覚的なシーンの変化はほとんどない。また、旅行においては、全てのシーンを網羅的に切り取る必要はなく、例えば、「目的地や観光名所付近の映像は必要だが高速道路や住宅街の映像は不要である」など、走行位置や周囲の状況によってシーンの重要度が大きく異なる。そのため、本提案のためには、視覚情報以外を手がかりとした映像の要約が必要である。

2.3 身近な製品を利用した体験要約

視覚情報以外を手がかりとした体験要約手法として、身近な製品を使用し、生活シーンの切り取りや要約、記憶支援を実現する研究がある。ComicDiaryは、携帯情報端末 (PDA) やキオスク端末を利用し、個人のプロフィールや行動履歴、他のユーザーとの交流記録などから、学術会議参加や博物館見学の日記を漫画形式で自動生成する[22]。Video-Recording Your Lifeでは、ウェアラブルカメラであるGoProと、スマートフォンの加速度センサを利用し、日常生活を記録した映像から利用者が興味を引くと思われるシーンを自動抽出する[3]。マイクロソフトのプロジェクト“MyLifeBits”で使用されたウェアラブルカメラSenseCamには多数のセンサが組み込まれており、身につけているだけで光量の著しい変化やカメラの前の体温などを自動で検出して写真を撮影する[7][12]。身近な製品を利用することは、多くの人に使うってもらう1つのきっかけになると考えられる。本研究では、近年急速に普及しており、手軽に購入可能なドライブレコーダの活用に着目し、要約の対象を自

動車乗車中とした。

また、世に広く普及している旅行の自動記録システムとして、Google マップのタイムライン機能がある。このサービスは、移動経路や滞在場所を地図上にプロットし、滞在時間などの情報と合わせた旅行の記録を自動でまとめる。また、Google フォトに保存されている写真を合わせて表示することもできる [9]。本研究でも、移動経路、滞在場所、時刻などの情報や、ユーザが旅行中に撮影した写真を自動で動画に組み込むように設計した。

3 ドライブ動画自動要約 Web システム

提案する自動要約システムは、ドライブレコーダから得られた動画ファイル、GPS のログファイル、ユーザが旅先で撮影した写真をもとにハイライトシーンを抽出し、サウンドエフェクトやトランジション効果を加えながら旅の要約動画を自動生成する。また、自動要約システムを広く使用してもらうために、Web システムを構築した。本 Web システムの概要を図 1 に示す。提案システムは <https://bitokana.com> から誰でも利用可能である (図 2)。対応ブラウザは Safari, Chrome, Firefox であり、PC から閲覧することを前提としている。ユーザは本サイトに必要事項を入力し、ドライブレコーダのデータをアップロードするだけで、思い出動画を手に入れることができる。Web サイトの構築のためにさくら VPS を使用し、SSL 証明書には Let's Encrypt を用いた。VPS の CPU は仮想 4Core、メモリは 4GB、SSD は 400GB、OS は CentOS を選択した。本システムは Python の Web フレームワークである FastAPI を利用している。以下、2022 年 2 月時点の最新バージョンについて説明する。

3.1 対応機種

ドライブレコーダは機種によってカメラの撮影範囲、保存されるデータの拡張子、1 ファイルあたりの撮影時間などが異なるため、使用する機種によって自動編集システムの処理が異なる。本稿では、水平方向に 360° 撮影可能であり、GPS 測位機能が搭

載されている Yupiteru 社 Q20-P^{†1} を使用した際の処理について述べる。このドライブレコーダは動画データと位置情報を、それぞれ MP4 ファイルおよび NMEA (National Marine Electronics Association) ファイルとして SD カードに記録する。いずれも約 1 分間のデータを 1 ファイルとして (以下、1 分録画ファイル)、複数のファイルが生成される。NMEA ファイルには、GPS により検出された位置情報が 1 秒毎に書き込まれる。MP4 ファイルの映像の記録形式は魚眼と 2 分割 (上半分が車外、下半分が車内の様子) が選べるが、本システムでは 2 分割の形式で記録していることを前提とする。なお、現段階では Yupiteru 社 Q20, Q20-P 以外の機種では検証していないため未対応であるが、今後は他の機種にも対応する予定である。

3.2 システムの利用方法

ユーザは Web サイト (<https://bitokana.com>) を通じて、システムに以下の 5 つの情報を送る必要がある。

(1) ドライブレコーダに内蔵された microSD カード内に生成されるフォルダ: Q20-P ドライブレコーダは、常時録画フォルダの中に動画ファイルと GPS ファイルを保存する。また、必要に応じて G センサー記録フォルダと、ワンタッチ記録フォルダを作成する。これらのフォルダを、それぞれフォルダごと、本システムにアップロードする必要がある。

(2) 旅先で撮影した写真: 旅行期間中に撮影した写真を要約動画に組み込みたい場合、写真を本システムにアップロードすることができる。ただし、要約動画に組み込むことができる写真は、Exif (Exchangeable image file format) 情報があるものに限る。

(3) 完成した動画をユーザに送信するためのメールアドレス: 本システムが生成する要約動画はデータサイズが大きいため、メールに動画をそのまま添付することはできない。そこで、本システムは動画の共有に Google ドライブのリンクを利用する。そのため、本システムの利用には Google アカウントが必要であ

^{†1} <https://www.yupiteru.co.jp/products/drive-recorder/q-20/>

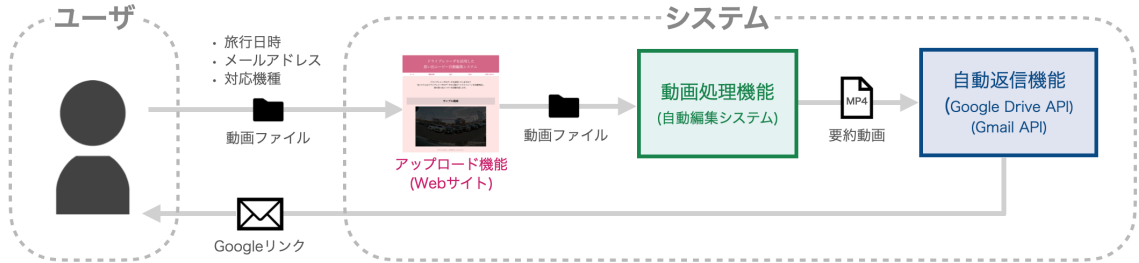


図 1 構築した Web システムの概要図。

り、G メールアドレスの入力を必須とする。

(4) **ドライブレコーダの機種**：前述のように、ドライブレコーダは機種によって映像の記録形式や保存データのディレクトリ構造が異なる。そこで、ドライブレコーダの機種をユーザに選択してもらうように設計した。

(5) **旅行期間**：通常、ドライブレコーダから取り出した microSD カードの中には、要約したい旅行期間以外のデータが含まれている。そこで、アップロードしたデータ全てを要約するのか、それとも一部を要約動画の素材として使用するのかが選択できるようにし、要約動画に使用したい動画データの期間を設定できるようにした。

以上の項目を入力し、送信ボタンを押すと、データがサーバに送信される。データの送信が完了すると、ユーザに受け付け完了のメールが自動送信される。

3.3 動画処理

サーバにアップロードされたファイルは、以下の手順で要約動画に変換される。

1. システムがデータ管理しやすいように、フォルダ名をユーザのメールアドレスとするフォルダを作成する。そして、配下に NormalRec ディレクトリと photo ディレクトリを作成する。
2. アップロードした全ての動画データを、作成した NormalRec ディレクトリに、全ての写真を photo ディレクトリに格納する。
3. アップロードされたデータのうち、一部の期間のデータのみを要約動画の素材として利用する場合、NMEA ファイルから取得した日時の情報を用いて、ユーザが指定した時間外の動画データと

GPS データを削除する。

4. 上記の手順で作成されたフォルダを自動編集システムが読み込み、動画編集処理を行う。

自動編集システムは、Python3 によって記述され、コマンドラインツールである FFmpeg を利用して動画処理を行う。自動編集システムの詳細な設計については 4 章で述べる。

3.4 自動返信機能

作成された要約動画は以下の手順で Google ドライブのリンクとしてユーザに共有される。まず、自動編集システムが作成した動画を Google Drive API を用いて Google ドライブにアップロードし、動画の共有リンクを取得する。この時、ユーザがリンク先の動画を閲覧できるように、ユーザが入力した G メールアドレスを利用してユーザに動画の共有権を付与する。そして、取得した共有リンクをメールの本文に貼り付け、Gmail API の自動送信機能により、ユーザのメールアドレス宛に自動送信する。

以上の処理により、ユーザは要約動画を取得する。また、全ての処理を終えたら、アップロードされた全データを自動で消去し、サーバ上に個人情報を残さないようにした。

4 自動編集システム

本システムは以下の 7 つの処理を全て自動で行う。

4.1 動画のサイズの調整

Q20-P ドライブレコーダから得られた動画データは、横縦が 2048×1536 画素、28fps である。1 フレームは上下に 2 分割され、それぞれに車内外の映像が



図2 作成した Web サイトを通してデータをアップロードする様子。データ送信前の画面(左)、アップロード中の画面(中央)、データ送信後の画面(右)。



図3 自動編集システムによる明るさの調整。調整前(左)と調整後(右)。

記録されている。車外の映像を使用するときは上半分中央部の 1360×765 画素を、車内の映像を使用するときは下半分中央部の 1360×765 画素を切り出し、 1280×720 にリサイズして使用する。なお、本システムが最終的に出力するムービーの画素数も 1280×720 画素である。

4.2 動画の明るさの調整

ドライブレコーダの主な目的は車外の様子の撮影であるため、 360° 全天球カメラで撮影された映像は、窓から見える景色の光量に合わせた露出となっている。そのため、昼間でも車内の映像は暗く、夜間に至っては暗くてほとんど何も見えない。そこで、車内の映像に対し、ガンマ補正、彩度調整、コントラスト調整を行う(図3)。

4.3 ハイライトカットの自動判定と切り取り

本システムは、ドライブ中に注目すべき事象が発生した際に撮影されるカット(以下、ハイライトカッ

ト)を自動抽出する。現在実装している抽出対象は、以下に示す6種類である。

- (1) 出発時：車両の ACC (アクセサリ) 電源が ON になったタイミングを出発時と判定し、そのときの車外の映像を7秒、車内の映像を10秒切り取る。車外の映像には NMEA ファイルから取得した時刻の情報を用いて、文字テロップと合成音声の挿入を行う。なお、ドライブレコーダは ACC 電源の供給に連動して自動的に記録を開始/終了するため、運転中だけが記録対象である。本研究ではこれを利用し、録画時間が10分以上開いた場合、休憩地点もしくは目的地に到着したと判断する。
- (2) 到着時：ACC 電源が OFF になったタイミングを到着時とし、車内の映像を10秒、車外の映像を7秒切り取る。車外の映像には、NMEA ファイルから取得した時刻と座標の情報、場所情報 API を用いて、到着時の時刻と住所を取得し、文字テロップと合成音声の挿入を行う(図4)。
- (3) 県境を通過したとき：NMEA ファイルから取得した座標情報と場所情報 API を用いて都道府県名を取得する。取得した都道府県名の変化から県境を判定し、県を跨いだ瞬間の車外の映像を3秒切り取る。また、都道府県名を文字テロップの挿入と合成音声の挿入により示す(図4)。
- (4) 名所付近を通過したとき：NMEA ファイルから取得した座標情報と場所情報 API を用いて周辺の

規模施設や地名などの地域情報を取得する。そして、取得した情報から名所付近を通過したことを判定し、そのときの車外の映像を3秒切り取る。また、地域情報を文字テロップの挿入と合成音声の挿入により示す。これらの情報は、場所情報 API の Score (取得した地域情報の確信度) が 99.9 以上の場合、すなわち、名所のすぐ近くを通過するときのみ表示する (図 4)。

(5) 高速道路の出入り口：NMEA ファイルから速度情報を取得し、1 分間の平均速度を計算する。本研究では、1 分間の平均速度が 60km/h 以上なら高速道路、60km/h 未満なら一般道と判断する。また、ETC (Electronic Toll Collection System) ゲートでは 20km/h 以下での徐行が求められていることから、高速道路と一般道の切り替わる箇所のうち、10km/h ~ 20km/h の箇所を高速道路の出入り口であるとし、その部分を 4 秒切り取る。

(6) 会話が盛り上がっているとき：MP4 ファイルから取得した音声データにバンドパスフィルターを適用し、日本人の音声周波数帯域とされる 300~3400Hz の周波数帯域を抽出する。そして、抽出した音声データの振幅が一定の閾値を超えた箇所を会話が盛り上がっているシーンと判定し、そのときの車内の映像を 6 秒切り取る。また、Python の SpeechRecognition ライブラリを利用して会話に字幕を表示する (図 4)。

4.4 旅先で撮影した写真の追加

スマートフォンやデジタルカメラで撮影された写真には、撮影日時や位置情報を含む Exif 情報が記録されている。そこで、旅先で撮影した写真に Exif 情報が埋め込まれていた場合、そこから撮影日時を取得し、作成動画の適切な位置に表示させる。画像の表示時間は 1.2 秒に設定し、縦長の写真の場合は左右に余白をつけることで 16 : 9 の比率に合わせる。

4.5 地図の生成

車旅行では、移動経路や滞在場所の地理情報も重要である。そこで、ハイライトカットの対象にならなかった 1 分録画ファイルに対しては、地図を伴った表示を行う。このために、Python の Folium ライブラリを用いて取得座標を地図上に可視化する処理を

行う。出発してから目的地に到着するまでに通過した道を赤色の線で示し、目的地に到着したときには赤色のマーカーをプロットする。そして、再出発時には、これまでに通過した道を赤色の線からピンク色の線に変更し、再出発してから次の目的地に到着するまでに通過した道を新たに赤色の線で示す。この地図を車外映像と組み合わせるために、2 種類の表現方法を採用した。まず、それぞれの 1 分録画ファイルの 1 フレーム目を切り出し、1 つ前の 1 分録画ファイル第 1 フレームの対応画素の RGB 値の差分の絶対値の総和を計算する。この総和を、設定した閾値と比較して以下の処理を行う。

差分が閾値以上だった場合：周囲の景色の変化が大きいと判断する。地図上に、移動経路、車外の映像から切り取った静止画、車のイラストを載せた静止画を、1 分録画に対して 1 枚作成しこれを 0.3 秒程度表示する。車外映像領域は画像 1 枚あたり約 1 秒間表示する (図 5。以下、地図主体表示)。

差分が閾値未満だった場合：景色の変化が少ないと判断する。1 分録画ファイルから切り出した車外映像を 100 倍速にして、移動経路を表示した地図をアルファブレンディングする (図 5。以下、100 倍速動画)。

4.6 処理の優先順位の調整

切り取り箇所の重複回避：ここまでの手順により作成された動画ファイルを時系列順に繋げることで、思い出動画を作成することができる。しかし、このままでは同じシーンが、複数のハイライトカットの対象となり、重複して切り取られてしまう可能性がある。そこで、ハイライトカットに優先順位を設定し、同一箇所が複数回切り取られることがないように調整する。**地図表示手法の調整：**作成したハイライトカットや地図表示をそのまま繋げると、画面が過剰に切り替わり、見辛い動画になってしまう。そこで、以下の処理を行い、シーンの過剰な変化を抑える。

1. 地図主体表示が 100 倍速動画に挟まれている場合、地図主体表示の代わりに 100 倍速動画を表示する。
2. 100 倍速動画が 3 連続未満である場合、それら全てを地図主体表示に変更する。



図4 自動編集システムが切り出すハイライトカット。左から順に到着時、県境を通過したとき、名所付近を通過したとき、会話が盛り上がっているとき。

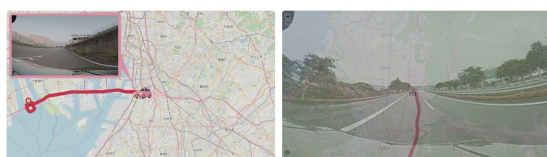


図5 ハイライトシーンがなかった1分録画ファイルに対する2種類の処理。景色の変化が大きいときの処理(左)と景色の変化が小さいときの処理(右)。

3. 地図主体表示が続かない場合、その地図情報は使用しない。

4.7 動画の連結とエフェクトの追加

以上の処理によって生成された素材に対して、以下の視覚音響効果とファイルの結合処理を行うことで要約動画が完成する。

吹き出しの拡大・縮小：地図主体表示とその他の動画ファイルの切り替えをスムーズに行うため、吹き出しのエフェクトを使用する(図6)。

フェードイン・フェードアウト：出発時にはフェードインの処理を、到着時にはフェードアウトの処理を加えることで一旦画面を暗くし、目的地に到着したことを示す。

放射状に広がるワイプエフェクト：通常速度の動画と100倍速動画の切り替えをスムーズに行うために、放射状のワイプエフェクトを加える(図7)。

ジングル：県境と名所付近を通過したとき、吹き出しのエフェクトによるトランジション、放射状に広がるワイプエフェクトには、それぞれ効果音をつける。100倍速動画にはアップテンポのジングル、地図主体表示にはミディアムテンポのジングルをつける。

動画ファイルの結合：以上の処理で作成された動画は全て、切り取り箇所該当時間に対応した名前がつけられ、完成動画用フォルダに保存される。たとえば、



図6 吹き出しの拡大エフェクトによるトランジション。



図7 放射状に広がるワイプエフェクト。

80ファイル目の30秒地点から4秒間の映像を切り取った場合、“00000080_30.mp4”という名前がつけられる。そこで、全ての処理完了後に、完成動画用フォルダにある処理済みの動画を名前順にソートし、全てのファイルを連結して1つの動画を生成する。

5 評価

5.1 評価実験1：オンラインアンケート調査

Webシステムの利用と、生成された要約動画に対する満足度や改善点を調査するために、本Webシステムを利用したユーザに対して、Googleフォームを用いた匿名アンケートを行った。アンケートリンクは、生成した要約動画のGoogleドライブのリンクと共に、メールでユーザに送信した。なお、アンケート対象はドライブ旅行の参加者全員とし、アンケートメールを受信したユーザには、同乗者にもアンケートリンクを共有するようメールで呼びかけた。

本調査は2021年10月～2021年11月の2カ月間実施し、計14個の動画が生成され、13件の匿名アンケートが集まった。アンケートは40代の男性からの回答が1件、40代の女性からの回答が12件であっ

表 1 乗車時間と手動編集への意欲を問う質問、およびその回答。

乗車時間	左で回答した時間分の ドライブレコーダのデータを手動で編集し 要約動画を生成したいと思いませんか？
30分未満	どちらかと言えばやりたい
30分未満	どちらかと言えばやりたくない
30分未満	どちらかと言えばやりたくない
1時間以上 2時間未満	どちらかと言えばやりたくない
1時間以上 2時間未満	やりたくない
2時間以上 3時間未満	どちらかと言えばやりたい
2時間以上 3時間未満	どちらかと言えばやりたくない
約 5 時間	やりたい
約 5 時間	どちらかと言えばやりたくない
約 6 時間	やりたくない
約 8 時間	やりたくない
約 10 時間	やりたくない

た。また、13件中6件がシステムに動画をアップロードしたアップロード者の回答、残りの7件が同乗者の回答であった。

以下、アンケート結果の詳細について述べる。なお、無回答の場合はデータの集計から除外している。

5.1.1 手動編集への意欲

ユーザの手動編集に対する意欲を調査するために、アップロードしたデータの乗車時間と、その時間分のデータを手動で編集したいと思うかどうかを問う設問を用意した(表1)。

手動編集への意欲について、6段階(非常にやりたい～非常にやりたくない)で評価したところ、やりたいが1件(8.3%)、どちらかといえばやりたいが2件(16.7%)、どちらかと言えばやりたくないが5件(41.7%)、やりたくないが4件(33.3%)であり、手動編集をすることにに対して消極的な意見が多いことが確認できた。

5.1.2 自動生成された動画の満足度

システムが自動生成した動画の満足度を調査した。「生成された要約動画の満足度を教えてください」という問いに対して、6段階(非常に満足～非常に不満)で評価したところ、非常に満足が2件(15.4%)、満足が6件(46.2%)、やや満足が3件(23.1%)、やや不満が2件(15.4%)という結果が得られた。「やや不満」と回答した2件のうち1件は、「運転ルートはある程度思い出せたが、運転ルートの振り返りにはそ

こまで意味がない。会話の内容の切り取りが欲しい」と回答しており、車内の様子や会話の要約を望む意見であった。もう1件は「ドライブレコーダの設定を間違えたのか、出来上がりの映像が見にくくなってしまった」と回答しており、これは本システムが2分割形式での記録にしか対応していないところ、魚眼形式で記録したデータをアップロードしたことが原因であると考えられる。

また、「今後、本システムを利用し、ドライブ旅行の要約動画を自動生成したいと思いますか？」という問いに対して、7項目(非常にそう思う、そう思う、少しそう思う、あまりそう思わない、そう思わない、非常にそう思わない、改善されたら使用したい)で評価したところ、非常にそう思うが8件(61.5%)、そう思うが4件(30.8%)、少しそう思うが1件(7.7%)という結果が得られた。

以上の結果より、システムが自動生成した動画についての満足度は高く、多くの被験者が今後も本Webシステムを利用したいと思っていることが確認できた。

5.1.3 自動生成された動画に対する意見

自動生成された動画を評価するために、生成された要約動画について自由記述式で意見を求めた。以下に、4つの設問(「生成された要約動画を見てどう思いましたか?自由に感想を教えてください」「生成された要約動画について、良かった点や評価できるポイントがあれば教えてください」「生成された要約動画について、欠点や改善すべき点があれば教えてください」「その他、感想やご要望があればお願いします」)に対して得られた回答をまとめる。

自動生成される動画の利点として、日時や通過場所を示すテロップ・音声は自動で入る点、短時間で要約されたデータが得られる点、走行ルートが地図で分かりやすく表示される点が挙げられた。また、「楽しく笑っているシーンが選択されていた」「旅の思い出がコンパクトにまとまっており、後から要約動画を見れば思い出せる」「要約されているので整理しやすい」といった意見も得られた。これらは生成された要約動画の満足度を高めた1つの要因であると考えられる。さらに、「ただの走行が楽しい思い出になり、感激した」「色々思い出が蘇り、楽しかった」「見返し

て盛り上がる」といった感想も得られ、本システムにより自動生成した要約動画が、自動車乗車後の UX 向上に貢献したと考える。

一方、改善点として、車内の会話の音量が小さいとき、車内の様子の切り取りが少なくなってしまう点、手動修正ができない点、後部座席の人の様子が映像に映らない点、音声認識により自動生成する字幕の精度が悪い点、同じ名所が複数回切り取られることがある点が挙げられた。

5.1.4 Web システムに対する意見

アップロード者に対して、本 Web サイト (<https://bitokana.com>) の利点や欠点を問う設問を用意した。本システムの評価できる点としては、「工程がシンプルで分かりやすい」「必要な手順が少なく、自動化率が高い」といった意見が得られた。一方、欠点としては、「アップロードに時間がかかり、途中不安だった」といった意見が得られた。

本システムの利用にあたってユーザーが必要な作業は、必要事項の記入と送信するフォルダの選択のみであり、送信ボタン押下後はユーザーの作業は一切不要である。しかし、ドライブレコーダのデータは、1 時間半程度の乗車時間に対して約 10GB と非常に大きいため、データをアップロードするには多くの時間がかかる。

5.2 評価実験 2：手動編集と自動編集の比較

著者の 1 人が母親と 2 人で実施した 5 回のドライブ旅行のデータ (表 2) を本システムに適用した。そのうち、3 回の旅行に対して (#1~#3)、自動編集と手動編集の違いを検証するため、双方とも動画編集初心者である、母親 (P1) と兄弟 (P2) に手動で動画編集をしてもらった。

P1：50 代女性、旅行に参加

P2：20 代男性、旅行に参加

手動編集には iMovie を利用した。本実験では、要約動画に使用したいハイライトシーンを選択後、文字やシーンの切り替え効果などのエフェクトを追加した。また、手動編集終了後、被験者に自動生成された要約動画を視聴してもらい、口頭でインタビューを行った。

表 2 評価実験 2 で使用したドライブレコーダのデータの詳細。

#	旅行期間	乗車時間	ファイルサイズ
1	5h 19m	約 1h 38m	12.4GB
2	8h 24m	約 1h 48m	15.6GB
3	8h 55m	約 2h 39m	20.1GB
4	1d 12h 23m	約 10h 39m	83.3GB
5	1d 11h 41m	約 14h 45m	106.6GB

5.2.1 時間効率

自動編集システムの処理に要した時間と (表 3)、P1 が手動編集するのにかかった時間 (表 4)、P2 が手動編集するのにかかった時間 (表 5) を比較する。

要約動画の生成に要する時間は、P1 では乗車時間の約 1.8~2.8 倍、P2 では約 1.4~2.1 倍であったが、自動生成システムでは、約 0.67~0.71 倍であり、自動編集に要する時間は手動編集にかかる時間に比べて大幅に短かった。

5.2.2 内容の重複度

システムにより自動生成された動画と、被験者 (P1, P2) により手動編集された動画の内容と構成の類似度を測定する。両者の違いや共通点を調査することを目的に、動画の重複度の算出、生成した動画の長さの比較とハイライトシーン 1 つあたりの平均の長さの比較、エフェクト数の比較を行った。なお、本実験において、動画の重複度が高いことは、比較した 2 つの動画の内容 (選択したハイライトシーン) が類似していることを意味する。

本研究では、重複度の指標に、適合率と再現率の調和平均である F 値を用い、一致したシーンの数を調査した。

本研究における 1 シーンの定義は以下の通りである。手動編集された動画については、要約動画生成後のインタビュー調査時に、編集者が 1 シーンであると発言した区間を 1 シーンと定義した。そのため、例えば、沈黙部分のカットにより、映像が時間的に連続していなくても、1 シーンとみなしている場合がある。システムによる自動編集については、1 回のハイライトカット処理で選択された映像を 1 シーンとした。シーンの一致判定については、2 つの要約動画間で少

表 3 自動編集システムの利用にかかる時間とシステムが自動生成した動画の詳細

#	要約動画生成に要する時間 (アップロード時間+動画処理時間)	自動生成した 動画の長さ	ハイライト シーンの数	エフェクト の数
1	1h 10m (36m + 34m)	4m 25s	32	56
2	1h 19m (46m + 33m)	3m 17s	17	39
3	1h 47m (1h 00m + 47m)	4m 43s	24	58
4	7h 31m (4h 12m + 3h 19m)	15m 22s	84	205
5	10h 06m (5h 17m + 4h 49m)	19m 35s	116	330

表 4 P1 が手動編集にかけた時間と手動生成した動画の詳細

#	要約動画生成に要した時間 (シーンの選択 + エフェクトの追加)	手動生成した 動画の長さ	ハイライト シーンの数	エフェクト の数
1	4h 37m 03s (2h 28m 25s + 2h 08m 38s)	16m 27s	23	45
2	3h 51m 11s (2h 23m 32s + 1h 27m 39s)	8m 10s	11	23
3	4h 47m 35s (3h 42m 24s + 1h 05m 11s)	15m 03s	21	41

表 5 P2 が手動編集にかけた時間と手動生成した動画の詳細

#	要約動画生成に要した時間 (シーンの選択 + エフェクトの追加)	手動生成した 動画の長さ	ハイライト シーンの数	エフェクトの数
1	3h 08m 10s (2h 38m 59s + 29m 11s)	8m 31s	20	28
2	3h 45m 17s (2h 50m 08s + 55m 09s)	5m 52s	20	22
3	3h 48m 25s (3h 27m 34s + 20m 51s)	9m 02s	34	38

しでも重なりがある場合、2つのシーンは一致しているとみなした。以上の方法に基づいて測定した結果、自動編集により生成した動画と P1 による手動編集により生成した動画のシーンの一致率は約 14~50%、自動編集と P2 による手動編集の間の一致率は約 27~33%、P1 による手動編集と P2 による手動編集の間の一致率は約 44~45% であった。このように、全ての比較において一致率が約 50% 以下と低かった。

また、生成した動画の長さについては、3回のドライブデータ全てにおいて、P1 > P2 > システムの順で長かった。これは、選択されたハイライトシーン 1 つあたりの平均の長さの違いに起因する。自動運転システムは動画のリズムを重視したため、1シーンあたりの長さの平均は約 7~9 秒だった。一方、手動編集では、会話の内容を意識していたため、P2 では約 16~26 秒、P1 に関しては約 43~45 秒と長かった。

エフェクトの数については、3回の旅行全てにおいてシステム > P1 > P2 であった。

6 議論

これまでの評価結果を踏まえ、自動編集システムの利点と課題、今後の展望について議論する。

6.1 自動編集システムの強みと手動編集の強み

オンラインアンケート調査の結果、多くのユーザが自動生成された動画に満足していることがわかった(5.1.2 節参照)。しかし、5.2.2 節で示した通り、自動生成された動画と手動生成された動画のシーンの一致率は低かった。これは、自動編集と手動編集で、要約の目的や抽出可能なシーンが異なるためであると考えられる。

現在発売されている多くのドライブレコーダは車

外のみを撮影し、車内の様子は記録できない。そこで、本システムは旅行ルートの要約と景色の切り取りを目的に設計しており、手動編集では難しい、周辺地域情報に基づいた要約が可能である。一方、手動編集では、システムには難しい、会話の内容に基づく要約が得意である。

また、自動編集により生成した動画と手動編集により生成した動画の一致率の向上を目指す場合、例えば、手動編集ではほとんど選ばれなかった、名所付近を通過したシーンを抽出しないように設計することが考えられる。しかし、手動編集と自動編集の比較評価では、P1、P2ともに、ランドマーク付近のシーンを選択することは好ましいとコメントしていた。この理由として、「車両周辺の地域情報が分かるので良い」「名所付近を通過したシーンが入っているだけで、間が少なく楽しい動画になる」「(名所付近の通過を示す)文字テロップが良い」といったことが挙げられた。

このように、システムが選択したシーンの中には、要約動画に必要であるものの、手動で抽出することが困難であるシーンも多く含まれる。そのため、シーンの一致率を高めることが、必ずしもより良い要約動画の生成につながるとは限らない。以上の結果を踏まえ、自動編集と手動編集のそれぞれの利点と適した使用用途を以下のようにまとめる。

自動編集システムの利点は、手間と時間がかからない点、動画編集の技術が不要である点、地図情報や周辺情報を提示できるため、旅行に参加していない人に対しても、旅の概要を分かりやすく伝えられる点である。このような動画は、日々のハイライトを共有するSNSでの使用にも適していると考えられる。また、P2は「少ない労力で見栄えの良い動画を作成できる。テンポも良く、SNSでの共有に向いている」と評価した。したがって、自動編集システムは、特に長期間の旅行時に有効であり、移動経路に着目した旅行動画の要約やSNSでの共有に適していると考えられる。

一方、手動編集の利点は、会話の内容や車内の状況、常識、同乗者間の関係を理解した上でシーンの選択ができる点、各々の好みに合った動画を自由に生成できる点である。そのため、手動編集は、家族や親戚の間で共有するホームビデオとしての用途や、会話の

内容に焦点を当てた思い出動画の作成に向いていると考えられる。

6.2 ユーザによる半手動編集の検討

5.1.1節で示した通り、オンラインアンケートの調査の結果、多くのユーザがドライブ旅行のデータの手動編集について、「どちらかといえばやりたくない」、もしくは「やりたくない」と回答した。一方、アンケートの自由記述欄には、テロップの手動修正やスタンプの追加などの簡単な手動編集機能の追加を求める意見があった。そのため、サイズや明るさの調整、注目シーンの抽出など、費用対効果の低い処理を自動化し、文字の入力やBGM (Background Music) の選択、生成する動画の長さの選択などの、簡単な手動編集機能を追加することは、より満足度の高い動画編集システムの実装に繋がると考えられる。

また、5.2.1章で示したように、手動編集では、シーンの選択だけでも膨大な時間を必要とする。しかし、人によって面白いと思う会話や思い出に残したいシーンが大きく異なるため、全自動でユーザの好みに沿ったシーンを正確に抽出することは難しい。実際に、P1、P2共に、主に会話の内容を基にシーンの選択を行っていたが、5.2.2説で示した通り、両者が作成した動画の一致率は低かった。これは、人によって動画編集時に重視する点が異なることが原因であると考えられる。動画編集の際に意識したことについて、P1は「面白い発言や会話を入れること」、P2は「旅の流れを把握すること」と回答した。以上を踏まえ、例えば、「システムが、会話の内容(旅行の話、家族の話など)や車内の様子(飲食のシーン、乗降者のシーンなど)をもとにシーンを分割・分類してハイライトの候補を提示し、ユーザは提案されたシーンを残すか否かだけを選択する」といったような、半自動化処理を実現することで、各々の好みに沿った動画を、少ない時間と手間で作成できるようになると考える。

6.3 自動編集システム作成の際に考慮すべき事項

フォローアップ研究として、以下の点を考慮する必要がある。

今回のアンケート調査では「楽しく笑っているシー

ンが選択されていた点によかった」という意見が複数得られたため、会話が盛り上がったシーンの選択を音量閾値で決定することはある程度有効であると考えられる。しかし、この方法では、口論や環境音が大きいシーンなどの好ましくないシーンが選択される可能性がある。また、会話の音量が小さい場合、車内の映像が要約動画に組み込まれる回数が著しく低下する。今後は、表情の分類や笑い声の検出など、乗客の感情を認識する機能の導入が必要であると考えられる。

名所付近を通過したシーンの抽出について、本研究では GPS データと周辺地域情報を返す API のみから判定していたため、旅行の思い出動画としては不要なシーンが切り取られることや、何度も同じランドマークに API が反応してしまうことがあった。ヘッドトラッキングやアイトラッキングを導入することで、どのランドマークが重要かを認識し、明確なランドマークがないシーン（海や山の景色など）をカットすることが可能になると考える。

プライバシー保護の観点からは、旅の要約動画を SNS にアップロードする場合に備え、他人の顔や車のナンバープレートにモザイク処理を施すことが望ましい。

6.4 提案する Web システムの制約

ドライブレコーダで撮影可能な範囲は限られている。そのため、車内映像に関しては運転席と助手席はよく見えるが、後部座席の様子撮影は難しい。しかし、オンラインアンケート調査の結果、「家族の旅の思い出が作れるよう、後部座席も映る車内の 360 度映像を撮影し、それも盛り込めるとよい」という意見や、「GPS 付きの他のカメラ（ドライブレコーダではないもの）でも同じように編集できる機能」の追加を求める意見があった。今後は、ドライブレコーダ以外の機器で撮影した映像データも要約動画の素材として使用できるようにしたい。

なお、車内の重要なシーンの記録を目的に、プロトタイプ作成時には、リードスイッチによるドアの開閉の検出、マットスイッチによる離着席の検出、押しボタンスイッチによる車内外の任意のシーンの切り取りを実装していた。しかし、これらの機能は追加の装

置が必要であるため、広く一般に使用してもらうことを想定した本 Web システムでは実装しなかった。これらの追加装置による車内映像の切り取りの詳細については [1] を参照されたい。

7 おわりに

本研究では、ドライブレコーダのデータを活用し、旅のハイライトを自動検出して思い出動画を自動生成する“ドライブレコーダを活用した旅の要約動画自動生成システム”の提案と実装を行った。ユーザは Web サイトにデータをアップロードするだけで、ドライブ旅行の思い出を紀行ムービーとしてまとめることができる。また、手動編集と自動編集の比較評価結果、および本 Web システムの利用者に対する匿名のアンケートの結果を報告した。今後は、これらの結果を踏まえ、より満足度の高いシステムの実装に向けて、新しい機能を随時追加していく予定である。

近年、自動運転の技術が成熟し、急速に普及しつつある。自動運転車では、高解像度の運転データ、乗客の生体情報、現状よりも高性能な位置情報を活用できるようになる。これらを活用することで、現在のドライブレコーダを使った手法に比べてはるかに高いユーザエクスペリエンスを提供できると考えている。

参考文献

- [1] Bito, K., Sii, I., Ishiguro, Y., and Takeda, K.: Automatic Generation of Road Trip Summary Video for Reminiscence and Entertainment using Dashcam Video, in *13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '21, New York, NY, USA, ACM, 2021.
- [2] 尾頭花奈, 椎尾一郎: ドライブレコーダを活用した思い出ムービー自動編集システム, 研究報告ユビキタスコンピューティングシステム (UBI), 2021-UBI-69(19), 2021, pp. 1-8.
- [3] Buschek, D., Spitzer, M., and Alt, F.: Video-Recording Your Life: User Perception and Experiences, in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, New York, NY, USA, ACM, 2015, pp. 2223-2228.
- [4] Chan, F.-H., Chen, Y.-T., Xiang, Y., and Sun, M.: Anticipating Accidents in Dashcam Videos, *Computer Vision – ACCV 2016*, Lai, S.-H., Lepetit, V., Nishino, K., and Sato, Y.(eds.), ACCV 2016,

- Cham, Springer, 2017, pp. 10114:136–153.
- [5] Cronje, J., and Engelbrecht, A. P.: Training Convolutional Neural Networks with Class Based Data Augmentation for Detecting Distracted Drivers, in *Proceedings of the 9th International Conference on Computer and Automation Engineering*, ICCAE '17, New York, NY, USA, ACM, 2017, pp.126–130.
- [6] de Avila, S. E. F., Lopes, A. P. B., da Luz, A., and de Albuquerque Araújo, A.: VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method, *Pattern Recognition Letters*, (2011), pp. 32(1):56–68.
- [7] Gemell, J., Bell, G., Lueder, R., Drucker, S., and Wong, C.: MyLifeBits: Fulfilling the Memex Vision, in *Proceedings of the Tenth ACM International Conference on Multimedia*, MULTIMEDIA '02, New York, NY, USA, ACM, 2002, pp. 235–238.
- [8] Gong, Y., and Liu, X.: Video summarization using singular value decomposition, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2000, IEEE, 2000, pp. 2:174–180.
- [9] Google: Google Maps Timeline, <https://www.google.com/maps/timeline>. (accessed 2021-09-18).
- [10] GoPro: HER10 Black, <https://gopro.com/en/us/shop/cameras>, 2021. (accessed 2021-09-18).
- [11] Grand View Research: Dashboard Camera Market Size, Share & Trends Analysis Report By Technology (Basic, Advanced, Smart), By Product, By Video Quality, By Application, By Distribution Channel, By Region, And Segment Forecasts, 2020 - 2027, 2020.
- [12] Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., and Wood, K.: SenseCam: A Retrospective Memory Aid, in *Proceedings of the 8th International Conference on Ubiquitous Computing*, UbiComp '06, Berlin, Heidelberg, Springer-Verlag, 2006, pp. 177–193.
- [13] Holoride: Holoride, <https://www.holoride.com>, 2019. (accessed 2021-09-18).
- [14] Honda: Honda Dream Drive, <https://www.honda.co.jp/CES/2017/detail/007/>, 2017. (accessed 2021-09-18).
- [15] Insta360: Insta360 Go 2, <https://www.insta360.com>, 2021. (accessed 2021-09-18).
- [16] 石黒祥生, 山田献二郎: 自動運転車両内インフォテインメントのための VR ゲーム, エンタテインメントコンピューティングシンポジウム 2017 論文集, 2017, pp. 210–216.
- [17] 国土交通省: 自動車用の映像記録型ドライブレコーダー装置について, <https://www.mlit.go.jp/monitor/R1-kadai01/24.pdf>, 2020. (accessed 2022-02-28).
- [18] Matsuda, A., Matsui, T., Matsuda, Y., Suwa, H., and Yasumoto, K.: A Method for Detecting Street Parking Using Dashboard Camera Videos, *Sensors and Materials*, (2021), pp. 33(1):17–34.
- [19] Mordor Intelligence: DASHBOARD CAMERA MARKET - GROWTH, TRENDS, COVID-19 IMPACT, AND FORECASTS (2021 - 2026), <https://www.mordorintelligence.com/industry-reports/dashboard-camera-market>. (accessed 2021-09-18).
- [20] Ngo, C.-W., Ma, Y.-F., and Zhang, H.-J.: Automatic video summarization by graph modeling, in *Proceedings Ninth IEEE International Conference on Computer Vision*, ICCV 2003, IEEE, 2003, pp. 1:104–109.
- [21] Samsara: AI Dash Cams, <https://www.samsara.com/products/safety/dash-cam>. (accessed 2021-09-18).
- [22] Sumi, Y., Sakamoto, R., Nakao, K., and Mase, K.: ComicDiary: Representing Individual Experiences in a Comics Style, in *Proceedings of the 4th International Conference on Ubiquitous Computing*, UbiComp '02, Berlin, Heidelberg, Springer-Verlag, 2002, pp. 16–32.
- [23] Takimoto, Y., Tanaka, Y., Kurashima, T., Yamamoto, S., Okawa, M., and Toda, H.: Predicting Traffic Accidents with Event Recorder Data, in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility*, PredictGIS '19, New York, NY, USA, ACM, 2019, pp. 11–14.



尾頭花奈

2021年3月お茶の水女子大学理学部情報科学科卒業。同年より名古屋大学情報学研究科博士前期課程在籍。自動車のエンターテインメントに関する研究に従事。



石黒祥生

2012年東京大学学際情報学府博士課程修了。博士(学際情報学)。米国 Walt Disney Imagineering 研究員を経て、現在、名古屋大学未来社会創造機構特任准教授および株式会社ティアフォーシニアリサーチャ。Mixed reality, augmented reality, human-computer integration 等に関する研究に従事。

**椎尾 一郎**

1979年3月名古屋大学理学部物理学
科卒業。1984年3月東京工業大学大
学院総合理工学研究科博士課程修了。

同年4月、日本アイ・ビー・エム株
式会社東京基礎研究所。1997年4月玉川大学工学部
助教授をへて2002年4月教授。2001年4月～2002
年3月ジョージア工科大学客員研究員。2005年4月
よりお茶の水女子大学理学部情報科学科教授、2022
年4月名誉教授。情報処理学会フェロー。工学博士。

**武田 一哉**

1985年名古屋大学大学院工学研究科
博士前期課程修了。博士(工学)。国
際電信電話株式会社、ATR自動翻訳
電話研究所、マサチューセッツ工科

大学滞在研究員、名古屋大学助教授を経て、現在、名
古屋大学大学院情報学研究科教授。行動信号処理に
関する研究に従事。