

# 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

佐藤 享憲<sup>1</sup> 成沢 淳史<sup>1</sup> 柳井 啓司<sup>1</sup>

## 1. はじめに

近年、カメラの小型化、低価格化により乗り物にカメラを取り付けて運転中の動画を撮影する人が増えてきている。このカメラを利用して動画を撮影することで道案内のための動画としても利用することができる。テーマパークなどの HP 上にあるアクセスマップだけを見るよりも、この道案内動画をあわせて見ることで実際の景色を見ることができるため、初めて行く場所でも道順がわかりやすくなると考えられる。しかし撮影された動画をすべて再生して見ることは時間と手間が掛かり現実的ではない。

そこで、本研究では車載カメラで撮影された運転中の動画の要約を行い道案内動画を作成することを目的とする。

また、シーン文字認識を行い、道路上にある案内標識に書かれている地名の文字を認識して強調することで道案内動画をさらに見やすくする。

## 2. 関連研究

岡本ら [3] はヘッドマウントカメラを装着して撮影された、一人称視点動画を横断歩道検出と自己動作分類による重要度推定を行うことで要約を行った。提案手法は岡本の手法に基づいたものになっている。提案手法では車載動画を対象として要約を行い、また、自己動作分類において計算コストが小さい手法を提案している。

車載動画を用いた研究は、Gandhi ら [1] の動画中の歩行者を認識する研究や、Ryoo ら [2] の周囲の状況から自車のイベントを認識する研究などがある。本研究では歩行者や周囲の状況とは関係なく主に自車の動作を認識している。

道案内動画を作成するツールとしては GoogleMap のストリートビューの画像を次々とつなぎ合わせることで、選択した出発地点から目的地までの動画を作成し、そのルートの景色を一連の流れとして見ることができる Google Street View Hyperlapse<sup>\*1</sup> というツールがある。このツールではストリートビュー画像を利用しているので画像にブレが生じたり、視点が固定できなかったり、任意の経路の道案内動画を作成できないといった問題点がある。

## 3. システム概要

システムの入力には車載カメラで撮影された出発地から目

的地までの経路が映っている動画と目的地付近の地名である。入力として与えられた動画を重要なシーンは通常再生、その他のシーンを高速再生することで要約を行う。動画から道案内標識の検出と自己動作分類を用いてシーン毎に重要度の推定を行う。また、検出された道案内標識に対して文字認識を行い、入力で与えられた「目的地付近の地名」が標識から認識された場合、要約動画再生中に道案内標識を拡大表示し、さらにその地名を強調することで進行方向を提示する。システムを実行した時の例は図 1 である。



図 1 システムの実行例

## 4. 手法詳細

### 4.1 案内標識検出

案内標識の検出は Local Binary Pattern (LBP) による検出器によって行う。3 フレームに 1 枚の画像を抽出して学習済みの検出器を用いて標識の検出を行う。誤検出を除去するために 30 フレーム (10 枚の画像) 連続して近傍で標識が検出された場合のみ検出されたとみなす。また、動画で見ると標識は y 軸マイナス方向に移動して見えるため、30 フレーム連続して、かつ検出結果が y 軸マイナス方向に移動している場合のみ検出されたと見なす。また、標識は画面の道路の上部に設置されているため、画像の下部から検出された結果を誤検出とみなし除去する。

65 枚の道案内標識が出現する約 54 分のテスト動画を用いて標識検出の精度実験を行った。学習画像の枚数は 461 枚の画像とそれらを反転した画像の合計 922 枚である。

再現率は 62.9%、適合率は 60.0%という結果となった。

### 4.2 文字認識

案内標識の検出を行った後、文字認識を用いて標識中の地名などの文字を認識する。入力で与えられた目的地付近

<sup>1</sup> 電気通信大学大学院 情報理工学研究科 総合情報学専攻

<sup>\*1</sup> <http://hyperlapse.tllabs.io/>

の地名が標識にあればその文字を強調して進行方向をわかりやすく提示する。

文字認識は NTT Docomo の文字認識 API <sup>\*2</sup> を使用して行う。API に画像を与えることで認識された文字、その文字の位置、カテゴリ、認識の信頼度を得ることができる。認識したい文字は地名なので、認識された文字列にアルファベットや数字が含まれている場合や認識の信頼度が 80 以下の場合には結果から除去した。

文字認識の精度実験を行った。道案内標識を手動で切り出し、docomo の文字認識 API で標識に書かれている地名を認識した。

テスト画像 212 枚中に出現した単語総数は 731 単語で、認識できた単語は 486 単語、認識精度は 66.48% という結果となった。文字認識を行った例を図 2 に示す。



図 2 文字認識の例

#### 4.3 自己行動分類

道案内動画では右折や左折をするシーンが重要となるため、右左折の検出を行う。これらの行動を分類するために LucasKanade アルゴリズムに基づいたオプティカルフローの計算を行う。3 フレーム毎に画像を抽出し、前後の画像からオプティカルフローを計算する。画像全体からオプティカルフローの計算を行い、30 フレーム (10 枚の画像連続) で式 1 が成り立つときは右折、式 2 が成り立つときは左折と認識する。なお、 $N_r$  は x 軸正方向のベクトルの数、 $N_l$  は x 軸負方向のベクトルの数である。

$$N_r - N_l < -100 \quad (1)$$

$$N_r - N_l > 100 \quad (2)$$

右折、左折の認識実験を行った。動画は約 19 分で、右折 47 回、左折 29 回の計 76 回の右左折が含まれている。再現率は 96.05% で適合率は 83.91% という結果になった。

#### 4.4 重要度推定と動画の制御

重要度を 0 と 1 の 2 値で定める。道案内標識が検出された前後 5 秒間のシーン、右左折が検出された前後 5 秒間のシーン、出発地と目的地が映っているシーンを重要とみなして重要度を 1 とする。それ以外のシーンの重要度は 0 とする。出発地は動画の初め 5 秒、目的地は動画の終わり 5

秒に映っていると仮定して、動画の初めと終わりの 5 秒間の重要度を 1 とする。

動画の要約は重要度が 1 となるシーンを通常再生し、重要度が 0 となるシーンを  $N$  倍速再生することで行う。

### 5. ユーザ評価実験

3 本の車載動画を提案手法、自己動作分類のみを用いた手法、均等に早送りする手法の 3 つの手法によって道案内動画を作成し、計 9 本の動画を 10 人のユーザに見てもらいその見やすさを評価してもらった。3 つの観点から 5 段階 (5 が最高) で評価してもらい、平均を取った結果と標準偏差をまとめたものが表 1 である。

表 1 ユーザによる評価実験の結果。5 段階評価の平均値と標準偏差を示す。

	提案手法	自己動作分類のみ	均等に早送り
出発地と目的地がわかる	4.1±0.9	4.0±1.0	3.5±1.0
曲がる交差点がわかる	4.3±0.8	3.6±1.0	2.8±1.1
道中の目印がわかる	4.1±0.8	2.6±1.0	2.5±1.3

また、どの手法で作成された道案内動画がわかりやすかったかを投票してもらった。投票してもらった結果、3 本の動画すべてにおいて提案手法が良かったと答えたユーザが 9 人、自己動作分類のみを用いた手法が 1 人、均等に早送りする手法が 0 人という結果となった。

得られた評価を用いて t 検定を行った。提案手法と他の 2 つの手法の評価を比較すると有意差があったことから、提案手法によって作成された道案内動画が最もわかりやすい動画であることが示された。

### 6. まとめ

車載動画のみを用いて、道案内標識の検出と自己動作分類による重要度推定を行うことで動画の要約を行い、自動で道案内動画を作成する手法を提案した。また、シーン文字認識を行うことで作成された道案内動画をより理解しやすいものにした。

今後の課題としては、標識の検出だけではなく、コンビニなどの看板の検出を行うことがある。また、信号の横にある交差点名が書かれている標識に対して文字認識を行うことでその交差点名を認識し、GPS 情報の利用なしに、動画中の現在地を推定することも今後の課題である。

#### 参考文献

- [1] Gandhi, T. and Trivedi, M. M.: Pedestrian protection systems : Issues, survey, and challenges., *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 3, pp. 413–430 (2007).
- [2] Ryoo, M. S., Lee, J. Y., Joung, J. H., Choi, S. and Yu, A. W.: Personal Driving Diary: Constructing a Video Archive of Everyday Driving Events, *Proc. of Workshop on Applications of Computer Vision(WACV)* (2011).
- [3] 岡本昌也, 柳井啓司: 道案内動画の作成のためのウェアラブルカメラ映像の自動要約, 電子情報通信学会論文誌. D, Vol. 97, No. 8, pp. 1237–1249 (2014).

<sup>\*2</sup> [https://dev.smt.docomo.ne.jp/?p=docs.api.page&api\\_docs\\_id=9](https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_docs_id=9)