

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, 成沢 淳史, 柳井 啓司

SS3-15

The University of Electro-Communications, Tokyo



背景 & 目的

ドライビングレコーダーの普及とカメラの高性能化道を教えるのに映像を使う方が直感的で分かりやすい



長すぎる映像を自動で要約し見やすく

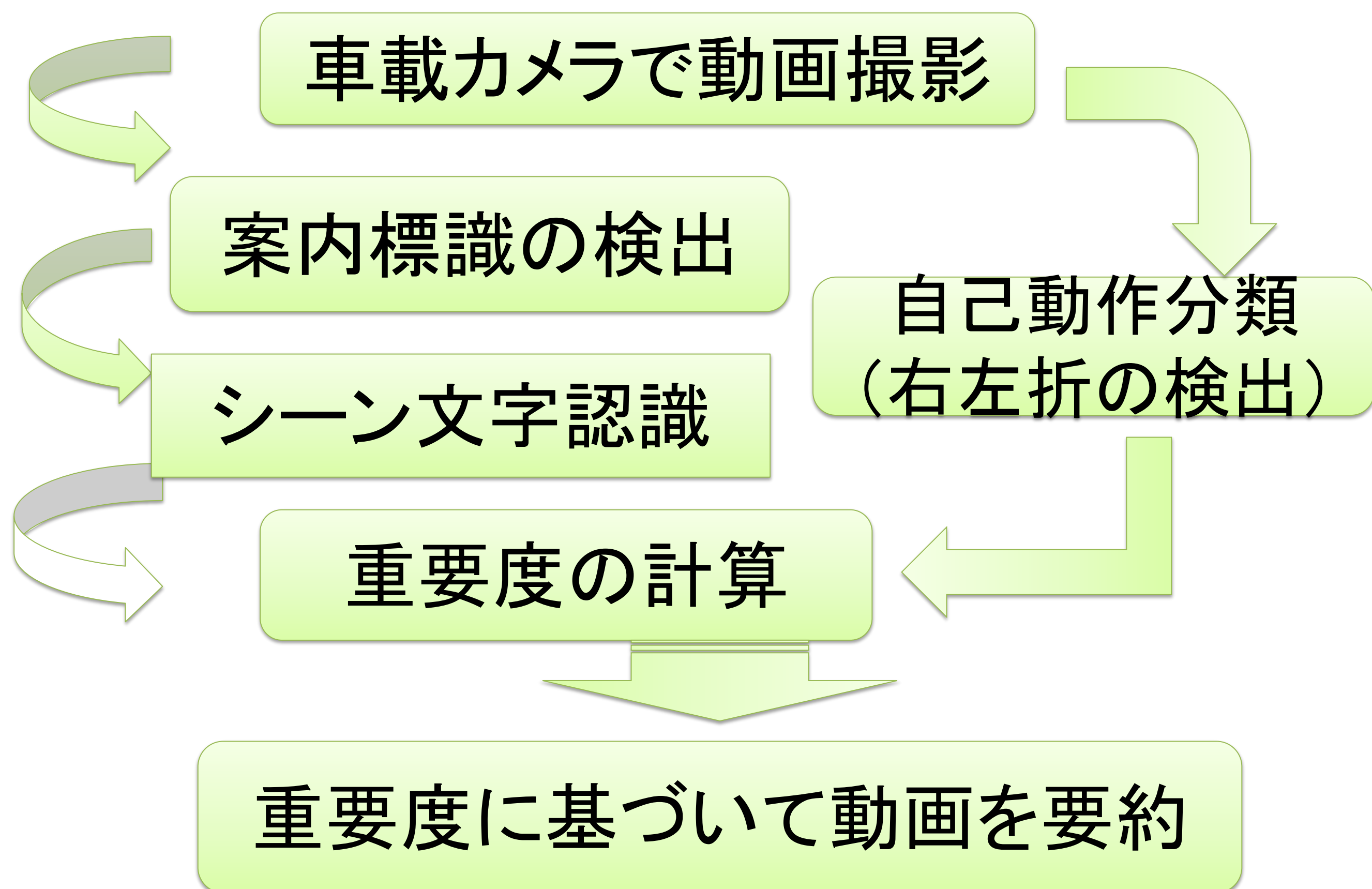
- 運転中の動画を撮影して自動要約を行い見やすい道案内動画の作成する
- シーン文字認識を組み合わせ目的地までの経路をわかりやすくする

道順を知る上での重要なポイント

- 右左折の場所
- 案内標識のような目印



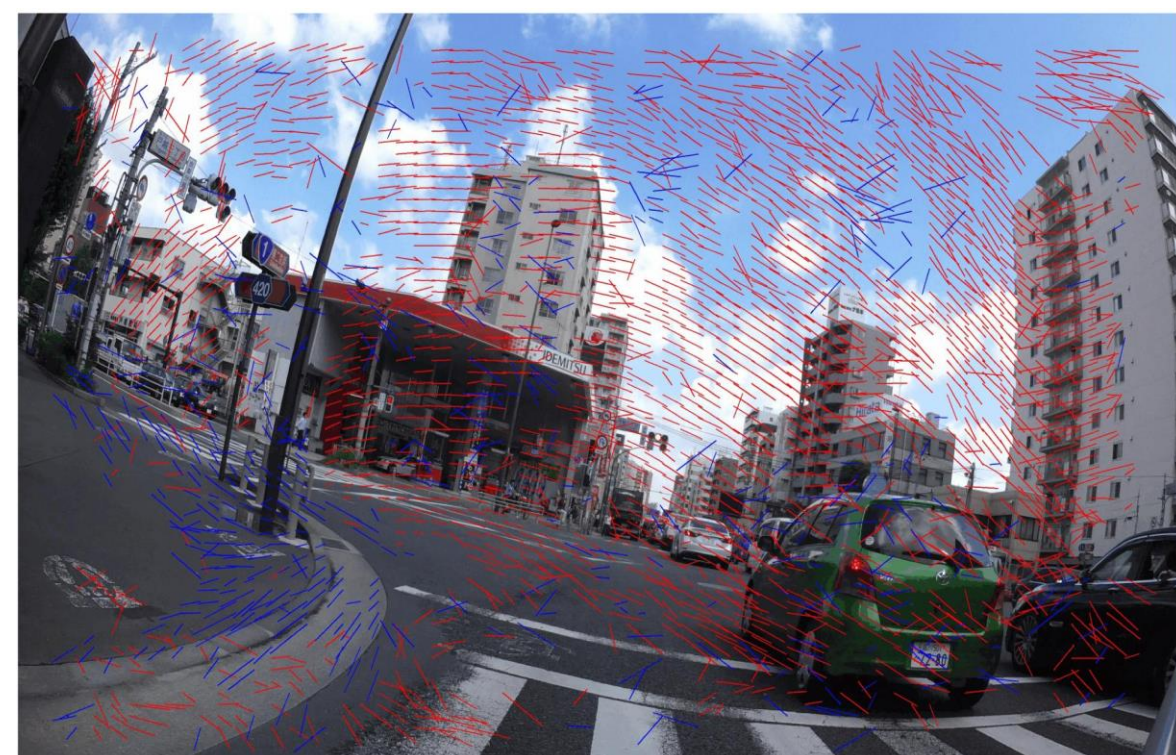
システムの流れ



自己動作分類

- オプティカルフローによる右左折分類

負方向のベクトルの数 N_r
 正方向のベクトルの数 N_l
 $N_r - N_l > 100$ と $N_r - N_l < -100$
 ベクトル総数で進行方向を決定



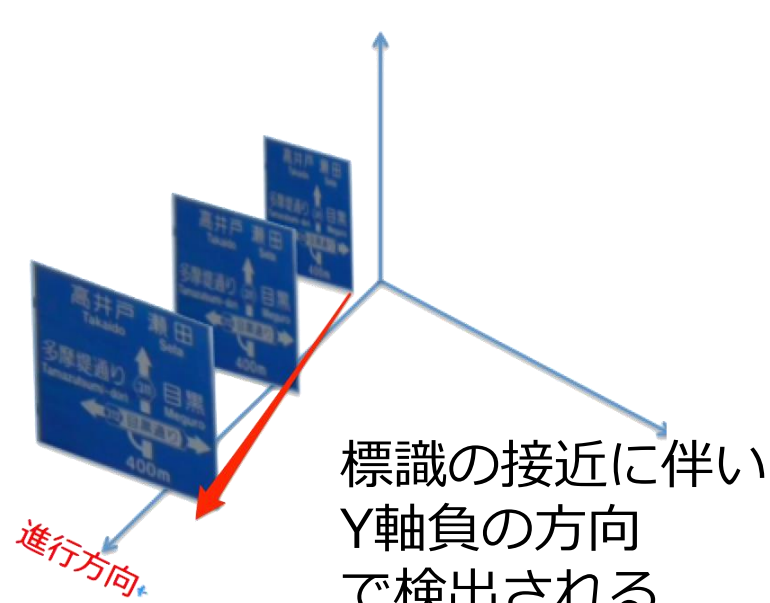
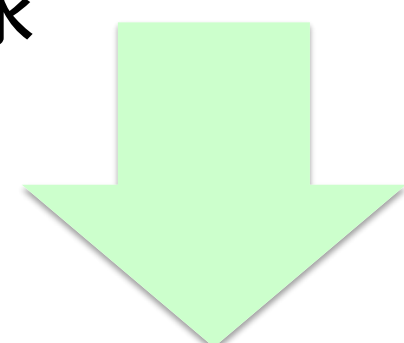
- テスト動画

右折47回,左折29回の計76回の右左折再現率:96.05%,適合率:83.91%

道案内標識の検出

- Local Binary Pattern 画像を使った検出器
- 学習画像は反転したものを含め424枚
- 3フレーム毎に検出を行いフォルスポジティブを除くため30Fの状態を記録

標識の出現位置と接近に伴う平行投影の変化を考慮



シーン文字認識

解像度により文字がつぶれるため4K以上の高解像度なカメラを使用



HD



4K

- Docomo文字認識APIを利用
- テスト画像212枚
- 出現単語数:731の地名 認識精度66.5%

重要度の計算、要約

- 重要度は1までの実数値で再生速度の早さとなる
- 出発地、目的地のシーン又は右左折、標識が検出されたシーンは重要度1
- 右左折、標識が検出された前後5秒間を通常再生
- 重要度が0のシーンはN倍速

ユーザー評価実験

- 3通りの方法で動画を要約しユーザー評価実験を行う
- 重要なシーンは等速で再生しその他は12倍速で要約する

提案手法	自己動作分類のみ	均等に早送り
<ul style="list-style-type: none"> • 道案内標識が検出されたシーン • 右左折のシーン • 始めと終わりの5秒間 	<ul style="list-style-type: none"> • 動画の始めと終わりの5秒間 • 右左折のシーン <p>を通常再生</p>	<ul style="list-style-type: none"> • 動画全体を均等に早送りする • 提案手法と自己動作分類と同じ長さになるよう全体を早送り
<ul style="list-style-type: none"> • 案内標識の地名を認識し赤枠で囲み拡大表示を行う 		

各方法による3つの観点からのユーザー評価実験(5段階評価)

	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



いくつかの動画を要約しユーザーから1-5点の評価を得た
 その結果提案手法が最もよい評価を得た

まとめ

- 車載動画の自動要約を行い道案内動画を作成するシステムを提案
- シーン文字認識を用いて進行方向を提示
- ユーザ評価実験では好評を得た
- 今後、標識だけでなくコンビニなどの看板も検出対象にしたい
- ランドマークなども重要度の推定に活かしたい