

深層学習による読唇手法の実用化に向けた モデリング改善に関する研究

研究の概要と特徴

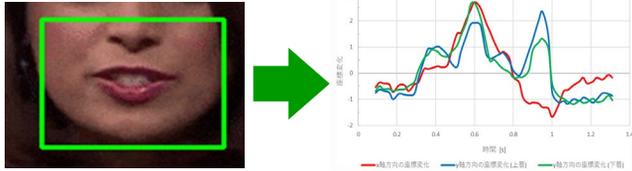
音声認識に代わる言語認識技術として、口唇部の動きを分類する読唇システムを提案
音声認識を利用する際の発声に対する抵抗感や運転時の車内外の騒音による影響を解消

研究の内容

運転時に使用する単語の発話時の口唇部特徴点の動きを取得し、深層学習を用いて分類
口唇部の動きの時系列情報を取得し、学習データとすることで精度の向上を実現

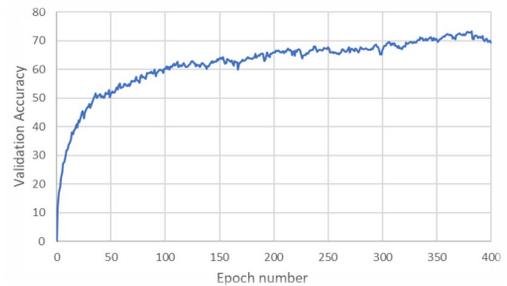
口唇部の取得

運転時に使用される単語を発話した際の口唇部特徴点の座標変化から時系列情報を取得



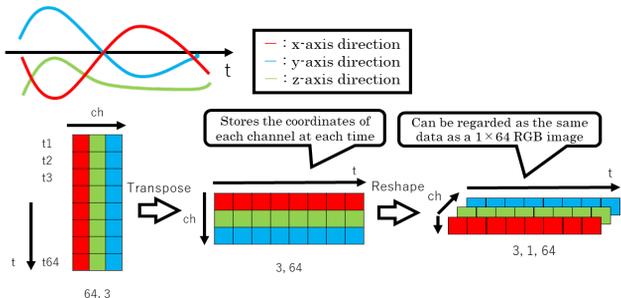
学習データの分類

学習が増加するに従い、検証用データに対する学習モデルの精度が向上



時系列情報の構成

取得した口唇部特徴点の時系列変化の情報を畳み込みニューラルネットワークを用いて分類するために、学習データの配列構造を変換



単語ごとの分類精度は平均70%程度
用途に応じて単語を分類することでより高い分類精度の実現を目指す

単語	正答率 (%)	単語	正答率 (%)	単語	正答率 (%)
ライト	70	ラジオ	59	前	80
スモール	47	音楽	84	後ろ	82
ハイ	93	スキップ	72	リクライニング	68
パッシング	80	音量	64	ドア	47
フォグランプ	76	マップ	51	窓	76
ルームランプ	69	ルート	72	開く	84
ワイパー	59	目的地	76	閉じる	74
フロント	59	スタート	80	ロック	76
リア	59	ストップ	69	アンロック	72
右	64	共通語	オン		
左	89		オフ		
ハザード	70		上げる		
クラクション	64		下げる		

研究の効果並びに優位性

口の動きの情報から車内装置に指示を与えることができるため、音声認識技術に代わる言語認識システムの開発が可能

技術応用分野・企業との連携要望

運転支援システムや自動運転技術に関連する企業との提携を希望