

平成 18 年度 情報工学コース卒業研究報告要旨

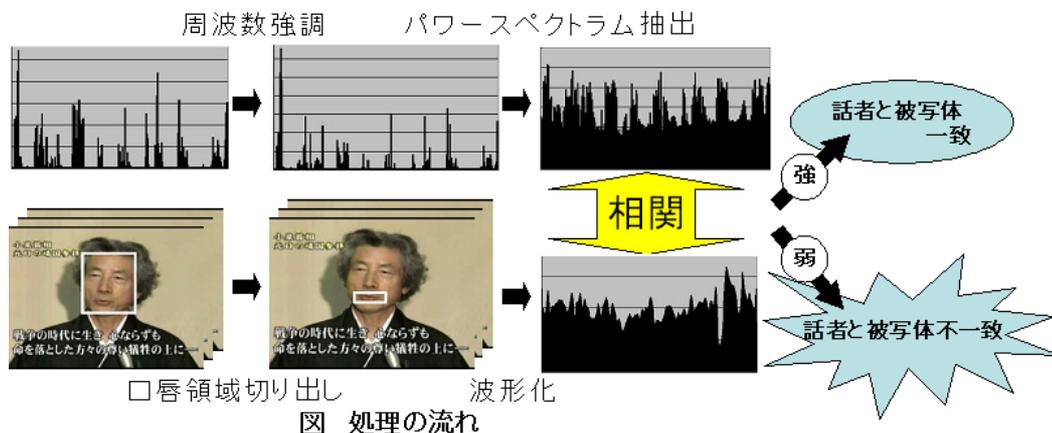
村瀬研究室 研究室	氏 名	小林尊志
卒業研究題目	ニュース映像における被写体と話者の不一致検出	

日本は今、コンテンツ大国への道を歩みつつある。それとともに付加価値を高めるコンテンツ処理技術の需要が高まっている。そこで今回は放送映像の処理技術に注目する。

日本でTV放送が始まって60余年、ようやくアーカイブとしての放送映像の蓄積が本格化してきた。しかし統一的な解析手法はまだ研究段階であり、アーカイブからの映像の検索や要約など放送映像の再利用は非常に困難である。例えば特定の人物が出現する映像区間を検索する場合、膨大な量の映像を閲覧して該当区間を探し出す必要がある。計算機を用いて自動で発言シーンを抽出する手法として、映像から画像処理により顔領域の検出を行い登場人物の発言シーンを抽出する手法がIdeら [1] によって提案されている。しかし顔領域が検出された映像区間についても、ニュース映像においては被写体とは別の人物、多くの場合アナウンサによるナレーションが、音声として記録されることがある。

本研究では、顔領域を含むニュース映像区間について画像情報と音声情報を組み合わせた登場人物の発言シーン検出手法を提案する。画像情報と音声情報の相関を考慮することにより、映像区間における話者と被写体の一致・不一致を分類することができる。まず顔領域を含む映像区間を画像と音声に分離する。音声は人間の声の周波数帯域を強調し、高速フーリエ変換を適用してパワースペクトラム波形を抽出する。画像については登場人物の口唇部分を矩形領域として切り出し、矩形領域の形状特徴の変化を波形として抽出する。次に音声情報による波形と画像情報による波形の相関を計算し、話者と被写体が一致しているクラスと不一致であるクラスに分離する。最後に顔領域の含まれる映像区間群から被写体と話者が一致していない映像区間を除外し、登場人物の発言シーンを検出する。

NHKのニュース7において2004年1月に放送された映像群から話者と登場人物が一致している映像区間50セット、話者と登場人物が不一致の映像区間50セット、合計100セットについて実験を行ったところ、一部について相関が見られ不一致区間を検出できた。なお、顔画像中における登場人物の口唇領域の切り出しについては、本実験では人手による切り出しを行った。



[1] Ichiro Ide, Naoki Sekioka, Tomokazu Takahashi, Hiroshi Murase, "Assembling personal speech collections by monologue scene detection from a news video archive", Proc. Eighth ACM SIGMM Intl. Workshop on Multimedia Information Retrieval, pp.223-229(Oct.2006)