

## 最適化および統計的性質に基づく情景内文字認識に関する研究

武部, 浩明

<https://doi.org/10.15017/1654906>

---

出版情報：九州大学, 2015, 博士（工学）, 課程博士  
バージョン：  
権利関係：全文ファイル公表済

氏 名 : 武部 浩明

論 文 名 : 最適化および統計的性質に基づく情景内文字認識に関する研究

区 分 : 甲

## 論 文 内 容 の 要 旨

本論文は、情景内文字認識に関する複数の試みをまとめたものである。情景内の文字は、多様な撮影環境や文字形状により、一般的なビジネス文書中の文字に比べて非常に困難な認識対象になる。それどころか、認識の前段階である文字の抽出ですら容易ではない。そこで本論文では、高精度に文字を抽出する手法と異なる2つのアプローチの高精度な個別文字認識手法を提案している。さらに、文字認識を用いることにより、講義を撮影した動画と講義に用いたスライドを高精度に同定する文字認識応用システムについて提案している。

第3章では、複雑かつ多様な環境において撮影された情景画像から、高精度に文字を抽出する手法について述べた。本手法では、画像の部分領域に対する文字/非文字の識別器として文字認識を用いながら、グラフカットによる最適な2次元セグメンテーションを行うことで、高精度に文字抽出を行う。具体的には、2値化のしきい値を段階的に変えて得られる個々の連結成分を文字仮説とした多重仮説をグラフとして生成し、グラフカットによる2次元最適化の枠組みで、文字仮説の近傍を考慮しながら最適な仮説を選択する。ICDAR2003のRobust Reading Datasetsを用いた定量的評価の結果、比較基準が若干異なるものの、他の手法を大きく凌駕することに成功した。再現率と適合率の平均値であるF値で比較すると、従来手法は0.63~0.7の範囲であるのに対して、本手法は0.8を達成した。

第4章では、従来の個別文字認識と同じアプローチの延長で、文字特徴量の正規分布を動的に推定することにより、高精度に文字を認識する手法について述べた。本手法では、処理対象である文字パターン集合の変動傾向を差分分布という統計的性質によって捉え、その情報を従来の文字認識手法のひとつである擬似ベイズ識別関数に組み込んだ形で文字認識を行う。形状が通常とは大きく異なるフォントに対する評価実験では、従来手法の精度87.8%に対して、本手法は91.4%と4.6%の精度向上が見られた。さらに、劣化文書に対する実験では、従来手法の精度95.9%に対して、本手法は96.7%と0.8%の精度向上が見られた。

第5章では、従来の個別文字認識と異なるアプローチで、学習データとして大量のフォントを集めて最近傍探索で認識する手法について述べた。ここでは、高次元の特徴空間を低次元空間に写像することにより、大規模データを圧縮してから最近傍探索を行う手法のひとつであるAGH(Anchor Graph Hashing)について考察を進めた。AGHでは、アンカーをK-means法によって選択しており、学習データの分布によっては最近傍探索の精度が低下する問題があった。そこで、アンカーの選択を類似度空間における低次元部分空間の選択という側面で捉えることにより、類似度空間における主成分分析によって最適なアンカーを選択する手法を提案した。文字認識実験によって従来手法との比較を行ったところ、特徴量の分布が比較的正規分布に近い手書き文字のデータベースに対しては、従来のK-means法による手法が最も良かったが、正規分布とは大きく異なると考えられる、多

様な文字フォントのデータベースに対しては提案手法が最も良い精度を持つことを確認できた。

第6章では、情景内文字認識の一応用として、講義動画とスライドを高精度に同定する手法について述べた。講義を撮影した動画に対する文字認識は精度が低く、文字認識エラーが多発する。そこで、多数決原理に基づく統計的な性質を利用することにより、文字認識エラーを許容しながら高精度に動画とスライドを同定する手法を提案した。具体的には、動画のフレームを文字認識した結果と、スライド中の文字について、文字配置に基づく類似度を算出して照合を行う。実際の講義動画に対して本手法を適用し、コンテンツ作成の作業コストを測定したところ、すべて手作業で行う場合と比較して、約70%のコストを削減できることがわかった。