



2008 年度下期未踏 IT 人材発掘・育成事業 採択案件評価書

1. 担当PM

田中 二郎 PM (筑波大学大学院 システム情報工学研究科 教授)

2. 採択者氏名

チーフクリエイター: 太田 悠平(東京工業大学大学院 総合理工学研究科
物理情報システム専攻修士 1 年)
コクリエイター: 馬越 健治(早稲田大学大学院 基幹理工学研究科
情報理工学専攻修士 2 年)

3. プロジェクト管理組織

株式会社オープンテクノロジーズ

4. 委託金支払額

4,500,000 円

5. テーマ名

音声認識併用による手書き文章処理アルゴリズム

6. 関連Webサイト

なし

7. テーマ概要

手書きデータを、検索やデータマイニング等、電子データのように扱いたい。音声認識と OCR を組み合わせれば、手書きデータを電子データのように扱える可能性があ

る。音声認識と OCR の誤差ベクトルは全く異なるため、認識精度が爆発的に向上する可能性があるからである。

本プロジェクトでは、基盤アルゴリズムとして、手書き OCR と音声認識の結合評価結果である最適筆記を用いた、高精度かつ誤記にロバストな手書き文字列認識法を提案する。結合に必要な計算のほぼ全てが最適筆記探索問題となることを示し、その汎用性を、提案アルゴリズムによって実演する。

提案する最適筆記探索法では、最終結果のみの結合ではなく、メモリ制約による枝狩りが生じる前に結合することで、精度と速度を同時に向上させる。リアルタイム処理実現のため、探索途中結果のマージ、再利用を行う Live Bands による筆記マッチングアルゴリズムを構築し、OCR 尤度を音声認識アルゴリズムにフィードバックし、最適筆記を定める。

最適筆記探索アルゴリズムの探索幅を狭めた逐次的探索によって、高精度な筆記書き起こしシステムを実現する。さらに、探索幅を広くし、認識尤度の高い局所最適解を拾い上げることで、OCR 結果中のキーワードの検索を可能とする。

筆記書き起こしシステム、キーワード検索システムによって構築されたアルゴリズムの応用として、素早く、正確に手書き OCR 結果を補正できる GUI を作成し、本提案の有効性を確認する。

8. 採択理由

音声認識と OCR を組み合わせ、手書きデータを電子データのように扱うことを提案している。提案する最適筆記探索法では、最終結果のみの結合ではなく、探索途中結果のマージや再利用を行う Live Bands による筆記マッチングアルゴリズムを構築し、OCR 尤度を音声認識アルゴリズムにフィードバックし最適筆記を定めている。筆記書き起こしシステムの応用として、素早く正確に手書き OCR 結果を補正できる GUI の作成システムを挙げている。興味深い提案である。

9. 開発目標

音声認識と OCR の認識結果を効率的に組み合わせるアルゴリズム、LiveBands アルゴリズムを構築する。さらに提案アルゴリズムのアプリケーションとして、OCR の誤認識を含む単語を読み上げることで誤認識箇所を自動で検索、修正できるアプリケーションを開発する。

本アルゴリズム及びシステムにより、従来ユーザビリティの低さよりあまり重要視されていなかった紙ベースメディアを最大限に有効利用できる環境を作る。

10. 進捗概要

OCR 結果と音声認識結果を効率的に統合するアルゴリズムの構築、及び OCR の誤認識を含む単語を読み上げるだけで、誤認識箇所を自動的に検索、修正するアプリケーションの開発、という目標は達成できた。さらに本プロジェクトでは、文字行の認識、切り取り、別に用意した一行のフリーピッチ文字認識用 OCR 帳票への貼り付け、をすることで、枠がある文字に対してだけでなく、より一般的な枠の無い文字に対しても本アルゴリズムの適用を可能とした。これにより、事前に指定した帳票に従った書類でなくとも本アルゴリズムが適用可能となり、本アルゴリズムの適用範囲が大きく広がった。

11. 成果

従来では、OCR の誤認識はキータイプにより手動で修正していた。しかし、本システムでは誤認識文字を含む単語を読み上げるだけで誤認識の自動修正が可能となった。例えば同じ単語が複数誤認識をしていた場合、従来ではひとつひとつキータイプで直す必要があったため時間がかかってしまっていたが、単語を一度読み上げるだけで複数の同じ文字の単語を同時に修正可能なため、修正にかかる時間の大幅な削減ができる。

さらにキーボードレスで OCR の誤認識を修正可能なため、キータイピングが遅い、手が不自由等の理由によりそもそもキータイプができない、といった人でも容易に OCR 結果の修正が可能となり、OCR の更なる普及につながる可能性がある。

12. プロジェクト評価

音声認識と OCR を組み合わせるという提案である。提案する最適筆記探索法では、最終結果のみの結合ではなく、探索途中結果のマージや再利用を行う Live Bands による筆記マッチングアルゴリズムを構築し、OCR 尤度を音声認識アルゴリズムにフィードバックし最適筆記を定めている。筆記書き起こしシステムの応用として、素早く正確に手書き OCR 結果を補正できる GUI の作成システムを挙げている。興味深い提案であったが、試作システムの完成度をさらに高めることが実用化のためには必要であると思われる。

13. 今後の課題

本手法による OCR 結果修正の精度は音声認識の精度に依存するため、音声認識の精度をさらに向上させる必要がある。具体的には、音声認識のモデルを特定の話者のモデルに近づける話者適応技術の導入が考えられる。さらに、OCR の認識結果の文字候補群に正解単語が入っていない場合、修正が困難になるという問題点もある。前後の単語から該当箇所の単語に入るべき文字を類推するなどのテキストの辞書処理の導入によって解決できる可能性がある。フリーピッチ文字認識及び LiveBands アルゴリズムの動作速度の向上も今後の課題である。