

黄瀬浩一

大阪府立大学 大学院工学研究科・教授

文字・文書メディアの新しい利用基盤技術の開発と  
それに基づく人間調和型情報環境の構築

## §1. 研究実施体制

### (1) 黄瀬グループ

①研究代表者:黄瀬 浩一 (大阪府立大学大学院工学研究科、教授)

#### ②研究項目

- (a) ベース認識装置開発
- (b) 大規模データベース構築
- (c) 実時間文字認識
- (d) 実時間文書画像検索
- (e) 全方位認識
- (f) 検索に基づく Reading-life log
- (g) 文字・文書メディアに対する拡張現実

### (2) 大町グループ

①主たる共同研究者:大町 真一郎 (東北大学大学院工学研究科、教授)

#### ②研究項目

- (a) ベース認識装置開発
- (b) 大規模データベース構築
- (c) 付加情報に基づく文字切り出し・認識
- (d) 全方位認識

### (3) 内田グループ

①主たる共同研究者:内田 誠一 (九州大学大学院システム情報科学研究院、教授)

#### ②研究項目

- (a) ベース認識装置開発
- (b) 大規模データベース構築
- (c) 認識に基づく Reading-life log

## § 2. 研究実施内容

(文中に番号がある場合は(3-1)に対応する)

### (1) 黄瀬グループ

#### (a) ベース認識装置開発[A-1][A-3]

昨年度に引き続き、局所特微量と最近傍探索を用いる文字認識手法を開発した。具体的には高速化のために近似最近傍探索手法を開発し、それをベース認識装置に組み込んだ。さらに、局所特微量の配置の制約を利用し、高速化を実現した。これらに加えて、抽出される特徴点が少ない字種(英数字やひらがな、カタカナなど)にこの認識方法を適用するための方法を検討した。

#### (b) 大規模データベース構築[A-5]

黄瀬グループでは大量のデータになるべく自動で正解ラベルを付与するアプローチを採っている。本年度は大規模データベースの構築で課題となるデータの収集と正解ラベル付けの両方に取り組んだ。正解ラベル付けの具体的な方法として、半教師あり学習の手法に基づく手法を提案し、歪んだ文字画像に対する実験で良好な結果を得た。データの収集については、全方位カメラを用いて情景中の画像を748,050枚撮影し、そのうち19,000枚に人手で正解ラベルを付与した。

#### (c) 実時間文字認識[A-6]

本年度の目標は字種に関係なく1fpsの認識速度を実現することである。英数字に関しては既に10fps以上を達成しているため、本年度は日本語(漢字)を認識する手法の改良に注力した。具体的には前述のベース認識装置を適用することによってこの目標を達成した。さらにこの認識装置をアイトラッカと組み合わせ、視線の先にある単語を実時間で認識し、翻訳結果を音声で提示する装置を試作し、ユーザビリティ試験により有効性を実証した。

#### (d) 実時間文書画像検索[A-2][A-4][A-7]

本年度の目標は、300万ページのデータベースに対して、100ms/query以下、検索精度90%以上を可能とすることである。研究の結果、2000万ページのデータベースに対して、49ms/query、精度99.2%を得た。これは、目標を大幅に上回る成果であると共に、現状で世界最速、最高精度の結果である。また、この技術を携帯情報端末上でも動作するようにシステム開発を行った。

#### (e) 全方位認識

全方位認識では撮影環境をコントロールできないため、様々に劣化した文字画像への対応が求められる。本年度は低解像度への対応を2種類検討した。一方はテンプレートマッチングに基づく手法であり、高解像度テンプレート画像を用いて低解像度の認識対象画像を認識する。他方は大量の事例を利用して低解像度画像から高解像度画像を作成する手法である。

#### (f) 検索に基づく Reading-life log

実時間文書画像検索に基づく Reading-life log の精度評価を行った。その結果、推定した注視点と実際の注視点の縦方向の誤差が、ユーザと文書の両方を固定しない場合でも1.5行程度に

収まることから、パラグラフ単位の位置推定には十分使えることがわかった。今後は、東京大学の佐藤洋一先生のグループとの間で知見を交換し、方式を改善する予定である。

#### (g) 文字・文書メディアに対する拡張現実

プロジェクタとカメラを接続し、紙文書に対して付加的な情報を投影することで、新しいサービスを提供するシステムを構築した。その結果、本年度の目標としていた1fpsでの動作が可能であることがわかった。実用化にはさらなる高速化に加えて、情報提示方法の工夫も必要である。

### (2) 大町グループ

#### (a) ベース認識装置開発[B-1][B-2]

本年度は、統計的手法による高精度文字認識アルゴリズムおよび、変形した文字をより柔軟に認識するためのグラフマッチングに基づいた手法について検討した。

#### (b) 大規模データベース構築

フォントの自動合成による大規模データベース構築に取り組んでいる。手動でフォントをいくつか作成し、そのフォントと同じ特徴を持つフォントを自動的に合成するための基本的な方針について検討し、(c)の結果を用いて大規模データベースを作るためのシステムを構築した。

#### (c) 付加情報に基づく文字切り出し・認識

付加情報に基づく文字切り出しと認識を実現するための基礎技術となる、文字フォントの自動構築法について検討した。文字の局所的な特徴を抽出する手法についても検討を行ない、文字パターンの大局的情報と局所的情報を分離し、局所的な情報を別の文字に付加することで、特定の局所構造を持った文字フォントの構築を可能とする手法を提案した。

#### (d) 全方位認識

全方位カメラからの環境中の文字の検出について基礎的なアルゴリズムの検討を行った。全方位カメラを用いる場合、取得される画像が大きいことから、高速処理が可能なエッジを利用して文字領域を抽出し、色を用いて文字検証することで高精度化を達成する手法を提案した。さらに、全方位カメラとパーソナルコンピュータを用いたプロトタイプの実験機を構築した。

### (3) 内田グループ

#### (a) ベース認識装置開発

環境内の多様な文字を高精度に検出し、認識するためのベース認識手法について、「部品に基づく文字検出・認識」および「環境コンテキストを用いた文字検出」の2点について検討した。

「部品に基づく文字検出・認識」について、具体的には、局所部分の安定性(具体的には特徴空間中での局所部分の分布特性) [C-1]、多数決統合の高度化 [C-2]、辞書なし文字認識 (Document image decoding) [C-3]を試みた。さらに、実際に局所部分が情景中の文字検出に使えるか否かに関する実験 [C-4]も行い、単純な局所部分でも文字・非文字の区別利用可能であることを実証した。また、局所部分を用いることで、情景画像中文字の傾き角度の高精度推定が可能であることを示した [C-5]。以上、H23 年度中の各種検討により、局所部分が文字の検出・認

識・幾何補正に利用しうることが十分に示された。

「環境コンテキストを用いた文字検出」について、この基本的な考え方の妥当性を示した[C-5]後、コンテキストとして利用する範囲の最適化、文字の顕著性(Saliency)のコンテキストとしての利用を図った。特に後者について、当研究室で準備したデータベース(後述)を用いた実験を行った結果、やはり情景内の文字には一定の顕著性が認められ、従って顕著性をコンテキストとして利用することが文字検出に有用であることを実証した[C-6]。

#### (b) 大規模データベース構築

Ground-truth 付きシーン文字画像データベースを作成した。まずラベル付け用のソフトウェアを開発した後、世界中で標準的に用いられている ICDAR2003 Robust Reading Database および Google Streetview Database 中の画像について、Ground-truth を付与した。さらにインターネットからダウンロードした画像を対象とした、オリジナルな大規模情景内文字データベースの作成も進行中である。H24 年 3 月現在で約 1,500 枚の画像について処理が完了している。

#### (c) Reading-life log

Reading-life log の実現を目的として、H22 年度から、市販 OCR を購入して、環境内文字の検出・認識システムを構築している。本システムでは複数の仮説の下での複数の認識結果を多数決統合することで、最終的な認識率を得る。検出 F 値 62%、検出後の認識率 98%という高い精度を実証した。また、認識ベース Reading-life log のプロトタイプ実装を行った。注視領域付近について市販 OCR を適用することで、認識結果を得る。その成果については、国内研究会において、昨年 11 月に発表した。

### §3. 成果発表等

#### (3-1) 原著論文発表

##### ●論文詳細情報

[A-1] 岩村 雅一, 堀松 晃, 丹羽 亮, 黄瀬 浩一, 内田 誠一, 大町 真一郎, 段階的な枝刈りによるアフィン不変な文字認識, 電気学会論文誌(D), vol.131, no.7, pp.873-879 (2011-7).

[A-2] Kazutaka Takeda, Koichi Kise, Masakazu Iwamura, “Memory Reduction for Real-Time Document Image Retrieval with a 20 Million Pages Database”, Proceedings of the Fourth International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2011), pp.59-64 (2011-9).

[A-3] Masakazu Iwamura, Takuya Kobayashi, and Koichi Kise, “Recognition of Multiple Characters in a Scene Image Using Arrangement of Local Features”, Proc. 11th International Conference on Document Analysis and Recognition (ICDAR 2011), pp.1409-1413 (2011-9).

[A-4] Kazutaka Takeda, Koichi Kise, Masakazu Iwamura, “Real-Time Document Image Retrieval for a 10 Million Pages Database with a Memory Efficient and Stability Improved LLAH”, Proceedings of the 2011 International Conference on Document Analysis and Recognition, pp.1054-1058 (2011-9).

[A-5] Masaki Tsukada, Masakazu Iwamura, Koichi Kise, “Expanding Recognizable Distorted Characters Using Self-Corrective Recognition”, Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS2012), pp.327-332 (2012-3).

[A-6] Takuya Kobayashi, Takumi Toyama, Faisal Shafait, Masakazu Iwamura, Koichi Kise, Andreas Dengel, “Recognizing Words in Scenes with a Head-Mounted Eye-Tracker”, Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS2012), pp.333-338 (2012-3).

[A-7] Kazutaka Takeda, Koichi Kise, Masakazu Iwamura, “Real-Time Document Image Retrieval on a Smartphone”, Proceedings of the 10th IAPR International Workshop on Document Analysis Systems (DAS2012) (2012-3).

[B-1] Tomo Miyazaki and Shinichiro Omachi, “Representative Graph Generation for Graph-Based Character Recognition,” The Journal of the Institute of Image Electronics Engineers of Japan, vol.40, no.3, pp.439-447, May 2011

[B-2] Masako Omachi, Shinichiro Omachi, Hiroto Aso, and Tsuneo Saito, “Pattern recognition using boundary data of component distributions,” Computers & Industrial Engineering, vol.60, no.3, pp.466-472, April 2011 (DOI: 10.1016/j.cie.2010.08.007)

[C-1] Song Wang, Seiichi Uchida and Marcus Liwicki, “Look Inside the World of Parts of Handwritten Characters”, Proceedings of The 11th International Conference on Document Analysis and Recognition (ICDAR 2011), pp.784-788, Sept. 2011

[C-2] Song Wang, Seiichi Uchida and Marcus Liwicki, “Comparative Study of Part-Based Handwritten Character Recognition Methods”, Proceedings of The 11th International Conference on Document Analysis and Recognition (ICDAR 2011), pp.814-818, Sept. 2011

[C-3] Wang Song, Marcus Liwicki, and Seiichi Uchida, “Toward Part-based Document Image Decoding” Proceedings of The 10th IAPR International Workshop on Document Analysis Systems (DAS2012), pp.266-270, March 2012

[C-4] Seiichi Uchida, Yuki Shigeyoshi, Yasuhiro Kunishige and Yaokai Feng, “A Keypoint-Based Approach Toward Scenery Character Detection”, Proceedings of The 11th International Conference on Document Analysis and Recognition (ICDAR 2011),

pp.819-823, Sept 2011

[C-5] Soma Shiraishi, Yaokai Feng and Seiichi Uchida, “A Part-Based Skew Estimation Method”,

Proceedings of The 10th IAPR International Workshop on Document Analysis Systems (DAS2012), pp.185-189, March 2012

[C-6] Yasuhiro Kunishige, Yaokai Feng and Seiichi Uchida, “Scenery Character Detection with Environmental Context”, Proceedings of The 11th International Conference on Document Analysis and Recognition (ICDAR 2011), pp.1049-1053, Sept 2011

[C-7] Asif Shahab, Faisal Shafait, Andreas Dengel and Seiichi Uchida, “How Salient is Scene Text?”, Proceedings of The 10th IAPR International Workshop on Document Analysis Systems (DAS2012), pp.317-321, March 2012

### **(3-2) 知財出願**

- ① 平成 23 年度特許出願件数(国内 2 件)
- ② CREST 研究期間累積件数(国内 3 件)