

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. 研究実施の概要

文字・文書は我々の生活や知的活動と密接な関係にある。本研究の目的は、文字や文書を自由環境で瞬時に認識・検索する技術を用いて、人間と調和する情報環境を作り出すことである。

平成22年度は、そのための第一歩として、大規模データベース構築のための準備を行うとともに、本研究の基盤技術となる(a) ベース認識装置、(b) 実時間文字認識、(c) 実時間文書画像検索、(d) 付加情報に基づく文字切り出し・認識、(e) Reading Lifelog について研究に着手した。その結果、次の成果を得た。

(a)ベース認識装置：局所特徴とその組み合わせに基づく認識、Massive 文字認識、多重仮説処理などを考案した。

(b)実時間文字認識：英数字を対象とした実時間文字認識システムにスペルチェッカーを統合し、単語認識の精度を向上した。

(c)実時間文書画像処理：1千万ページDBに対して、精度99.4%、処理時間35msの高速検索を実現した。

(d)付加情報に基づく文字切り出し・認識：付加情報の基礎となるフォントからの自動構造抽出とそれに基づく合成法を考案した。

(e) Reading Lifelog：アイトラッカと文書画像検索を連結して動作させることに成功した。また、同技術の基盤となる、情景内画像からの文字検出について、局所特徴と顕著性に注目した手法、環境コンテキストを用いた手法など、複数のアプローチを試み、定量的評価を行った。

## §2. 研究実施体制

### (1) 黄瀬グループ

① 研究分担グループ長：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

・ベース認識装置開発

- ・実時間文字認識
- ・実時間文書画像検索
- ・Reading Lifelog

(2) 大町グループ

① 研究分担グループ長: 大町 真一郎 (東北大学大学院工学研究科、教授)

② 研究項目

- ・ベース認識装置開発
- ・付加情報に基づく文字切り出し・認識

(3) 内田グループ

① 研究分担グループ長: 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

- ・ベース認識装置開発
- ・Reading Lifelog

### §3. 研究実施内容

#### (1) 黄瀬グループ

##### (a) ベース認識装置開発

局所特徴量と最近傍探索を用いる情景画像中の文字認識を試みた[A-1]。局所特徴量を用いる従来の文字認識手法は情景画像中に含まれる複数の文字を効率よく扱えなかったが、提案手法ではこの問題を解決した。この方法は特に漢字に有効であり、実験の結果再現率約 97%、適合率約 98%を達成した。今後はこの手法をベースにして、英数字やひらがな、カタカナにも有効な手法の開発し、扱うデータを大規模化する予定である。

##### (b) 実時間文字認識

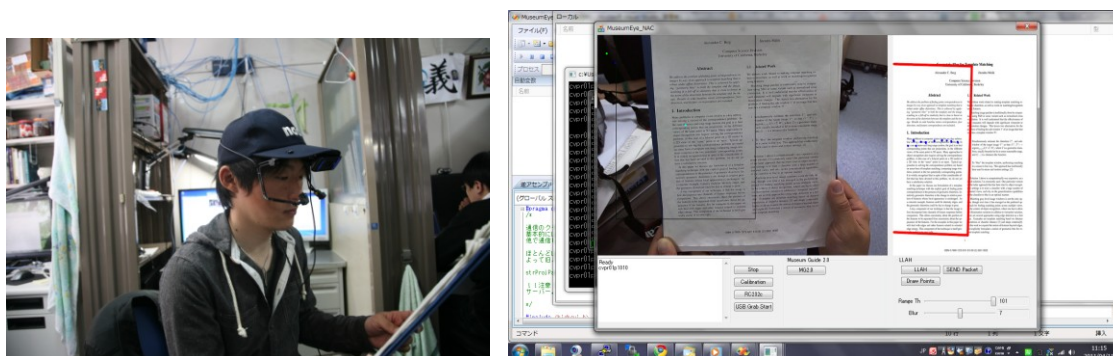
英数字を対象とした実時間文字認識システムを開発し、ノートパソコン上で約 10fps で動作する事を確認した。そして、単語認識の精度を向上するために、このシステムにスペルチェッカーを統合した。スペルチェッカーの統合は特に斜めから撮影した単語画像に有効であり、紙面に対して 20 度程度の角度から撮影した単語認識精度が約 40%から約 98%に向上する単語もあった。この技術は平成 23 年度から実施予定の全方位認識にも使用できると考えている。

##### (c) 実時間文書画像検索

研究代表者らが開発した LLAH と呼ぶアルゴリズムを改良し、多言語への対応[A-2]を行うとともに、特徴点のサンプリングによる省メモリ化、特徴量の高次元化による識別性改善、冗長な特徴量の削除による安定性改善を行った。有効性を検証するため、1030 万ページデータベースに対して 1003 枚の画像を検索質問とした検索実験を行った。その結果、処理時間 35ms/query、使用メモリ 44GB の下で検索精度 99.4%を得た。

##### (d) 検索に基づく Reading Lifelog

試作システムを図 1 に示す。これは NAC 社製アイトラッカを上記の LLAH と連結したものである。その結果、視線に応じて文書画像が実時間で検索され、文書のどの部分が注視されているのかを記録することに成功した。ただし、注視箇所の推定については未だ誤差が大きく、この克服が課題として残されている。



(a) 装着時の外観

(b) システム動作画面

図1 検索に基づく Reading Lifelog の試作システム

## (2) 大町グループ

### (a) ベース認識装置開発

本グループでは、ベース認識装置で重要な役割を果たす、情景画像から文字領域を抽出する手法について検討した。フォント等個々の文字に依存しない処理を実現するために、文字らしさを表わす特徴として色の均一性およびエッジに注目し、これらをもとに文字領域を抽出する手法を検討した。その結果、実時間で処理することを考慮するとエッジ情報がより有効であることが分かり、エッジ抽出による文字領域抽出を高速かつ高精度に行なう手法について検討を重ねた。エッジを抽出する際に有効な色空間、および、高速化のための並列処理手法を検討した。その結果に基づき、毎秒 3 フレーム程度の処理速度で文字領域を抽出することができるシステムを試作した。さらに全方位カメラを購入し、全方位画像を取得するシステムを構築した。全方位画像から文字領域を抽出する手法については検討中である。

### (b) 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を行なった。特定のフォントの文字画像を数枚サンプルとして与えたときに、それらの画像を小さいパッチに分割し、並べ替えることで与えられた文字以外の文字フォントを自動的に構築する手法を提案した。その並べ替えの指針の一つとして文字の骨格情報および画像の形状情報を表す **Shape Context** を組み合わせて用いた。実験の結果、サンプルの特徴をある程度保持した文字画像が構築できることを確認した。一方で、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する手法を開発した。この手法により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が可能になるものと考えており、引き続き検討している。

## (3) 内田グループ

### (a) ベース認識装置開発

本プロジェクトの基盤となる文字認識技術を構築するという課題に対し、3つの異なる視点から取り組んだ。具体的には、局所特徴に基づく文字認識[C-1,C-2]、Massive 文字認識[C-3]、および多重仮説による実環境文字認識について試みた。

局所特徴に基づく文字認識とは、文字を局所領域すなわち断片に分解し、断片毎の認識結果を統合する手法であり、変形に頑強な文字認識手段と成りうる。今年度は、幾つかの統合手法を比較検証するとともに、特徴空間における断片の分布について詳細な観察をおこなった。

Massive 文字認識については、100 万個に近い正解付き手書き文字パターン(辞書パターン)の利用が、どの程度、そしてどのように認識精度の向上に資するのかを実験および各種統計データから検証した。結果、誤認識を半減させるためには、辞書パターンをおよそ2桁増やさなくてはならないことなどが判明した。

多重仮説による実環境文字認識とは、情景画像中の文字を検出・認識する際、単一の領域分割

法では必ず限界があるという予想に基づき、複数の方法で領域分割し、全領域それぞれについて単文字認識を試みる方法である。今年度、分割法および単文字認識の実装が終了した。

#### (b) Reading Lifelog

Reading Lifelogの基盤技術となる、情景内画像からの文字検出について、主として、局所特徴に注目した手法、および環境コンテキストを用いた手法という二つのアプローチを試み、定量的評価を行った。

前者は、明確に文字だけを切り出すことは困難という考えに基づいた方法である。具体的には、画像中のコーナー等に局所領域を多数検出し、その局所領域毎に文字・非文字判定をするものである。局所領域だけしか見ないという状況ながら、75%を超える識別率を得ている。今後、周辺の局所領域の結果と併せて多数決などを取るなど、結果の安定化と高精度化が課題となる。

後者は、文字の存在可能性(事前確率)を周辺環境の状況から計算して利用することで、検出の精度向上、特に誤検出の削減を目的としている。実験の結果、Recall、precisionを共に10%向上できることを確認した。

#### §4. 成果発表等

##### (4-1) 原著論文発表

###### ● 論文詳細情報

- [A-1] Takuya Kobayashi, Masakazu Iwamura, Koichi Kise, “Recognition of Affine Distorted Characters by Using Affine-InvariantLocal Descriptors”, Proceedings of The 2nd China-Japan-Korea Joint Workshop on Pattern Recognition (CJKPR2010), pp.74-77, 2010
- [A-2] Kazutaka Takeda, Koichi Kise, Masakazu Iwamura, “Multilingual Document Image Retrieval Based on a Large-Scale Database”, Proceedings of The 2nd China-Japan-Korea Joint Workshop on Pattern Recognition (CJKPR2010), pp.42-46, 2010
- [C-1] Seiichi Uchida and Marcus Liwicki, “Part-Based Recognition of Handwritten Characters”, Proceedings of The 12th International Conference on Frontiers in Handwriting Recognition, pp.545-550, 2010 (DOI: 10.1109/ICFHR.2010.90)
- [C-2] Masahiro Fukutomi, Koichi Ogawara, Yaokai Feng, and Seiichi Uchida, “ Sequential Pattern Recognition by Combining Local Classifiers”, 2nd China-Japan-Korea Joint Workshop on Pattern Recognition (CJKPR2010), pp.209-212, 2010.
- [C-3] Wenjie Cai, Yaokai Feng and Seiichi Uchida, “Massive Character Recognition with a Large Ground-Truthed Database”, Proceedings of 26th Symposium on Applied Computing, Document Engineering, 2011.

##### (4-2) 知財出願

- ① 平成22年度特許出願件数(国内 1 件)
- ② CREST 研究期間累積件数(国内 1 件)