

カメラ付きヘッドマウントディスプレイを利用した 書類題目認識に向けた一検討

Document Title Recognition System using Head Mounted Display with Camera

西村 広光†

Hiromitsu Nishimura

1. まえがき

デスクワークで膨大な書類を管理する際に、資料を分類整理することの労力は非常に大きい。それは弱視などの障害を抱えているものにとっては、極めて甚大な労力が必要となる。本研究では、眉間位置にカメラを搭載したヘッドマウントディスプレイ(HMD)を利用して、弱視などの障害を持った方がデスクワークを行う際に、書類を目前に通すだけでその書類タイトルを認識して情報を提示するシステムについて検討を行っている。

カメラを利用した文字認識においては、現状でまだ多くの課題が残されている[1]。そのなかでも本論文では、得られた低画質画像から書類のタイトル部分を検出する処理について検討を進め、システムを試作し評価実験を行った。文字にいくつかの制約条件を付与することで、認識可能なパターンとしてタイトル部のパターンを検出することに成功した。

2. HMD を利用した書類題目認識

本研究で検討している HMD を利用した書類題目認識システムについて詳説する。本研究で利用する HM は、カメラ機能を搭載した AR Vision HMD タイプを利用することとした。この HMD は通常の HMD に加え、両眼中央にカメラを搭載しており、カメラで撮影した画像を PC に取り込み、処理した結果を HMD の出力として表示できるものである。主な仕様を表 1 にまとめる。

表 1. AR Vision の主な仕様

カメラ	カラー, NTSC, 1/3 インチ CMOS 60fps interlaced NTSC
フォーカス	固定
ディスプレイ	2xSVGA マイクロディスプレイ 480,000Pixels
視野	対角約 40 度 (4:3, 水平 32 度, 垂直 24 度)

AR Vision を利用して、目前に手で書類を持ってかざすことで、その書類を認識し、それをデジタルデータとして取得し、AR 空間上の仮想書類庫にデジタル書類の出し入れを可能にするようなシステムの構築を最終的な目標としている。本提案手法は、デスクワークの書類整理において大きな効果を発揮することだけでなく、弱視などの資格障害者のデスクワーク支援のシステムとしての応用も検討して開発を進めているものである。本発表では、その最も根幹部分である AR Vision からの映像取得を行い、目前で短

時間静止させた書類のタイトル部分を検出するシステムについて報告する。

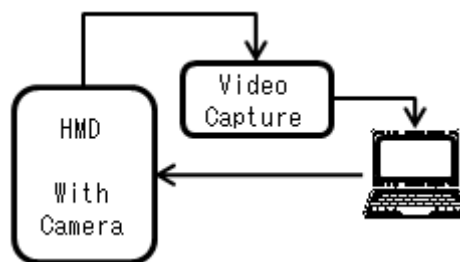


図 1. 構築システムのデータの流れ

3. 構築システム

構築システムは、図 1 に示すように、AR Vision カメラから出力したアナログ信号をビデオキャプチャ(IO-Data 製 GV-USB)を利用して USB 信号でコンピュータに取り込み、処理した後にアナログで AR Vision モニタに出力することとした。各部の信号がアナログでしか受け渡しができないのは、HMD の入出力の仕様に依存したためである。なお、本システムの構成から、取り込める画像の最大サイズは、横 720×縦 480 画素であり、SVGA よりも小さい。

図 2 左に、コンピュータに取り込んだ画像を 2 値化処理した結果を、図 2 右に、最も上に書かれた多さの文字部分を 300% に拡大した画像を示す。

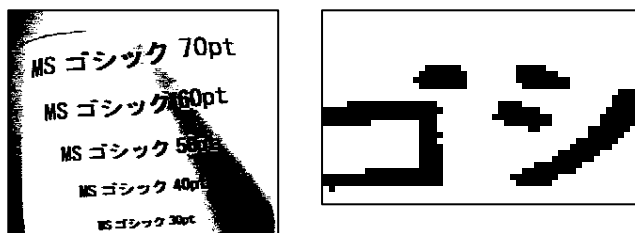


図 2. AR Vision から取り込んだ画像

図 2 は十分な環境光がある状態で A4 用紙に印刷した 10pt ずつ大きさを変えた文字を AR Vision のカメラで読み取ってコンピュータに取り込んだ画像である。このときのカメラと用紙との距離は、カメラの視界内に、ちょうど A4 縦の長さがすべて見えるような位置である。この図からわかるように、カメラの解像度が十分でないため、40pt 程度以上の文字であれば A4 全体の距離から文字部分を認識可能なパターンとして抽出することができると考えられる。しかし、HMD 搭載型カメラで、手に持った書類の文字を取り込んだ場合、双方を静止させようとしても完全に静止することがなく、図 2 右のようなブレを含むような画像でコンピュータに取り込むことになる。

† 神奈川工科大学 Kanagawa Institute of Technology

前述のように取得した画像に対して、図3に示す処理の流れによってタイトルである文字パターン領域を抽出する処理を実装した。

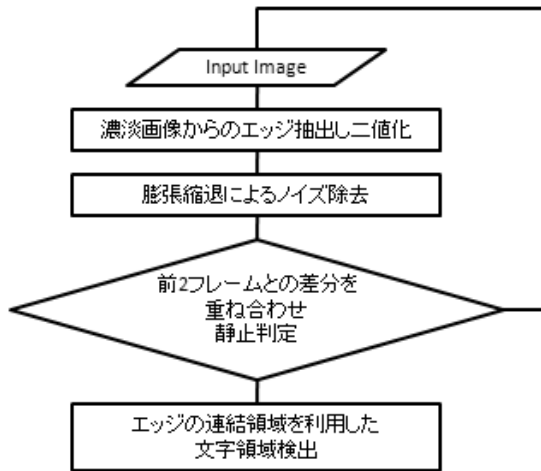


図3. 文字抽出までの処理の流れ

本システムはリアルタイムでの実装を最終的な目標としているため、高速に処理を行う必要がある。そこで入力カラー画像を濃淡画像にし、エッジ情報を抽出し、そのエッジを2値化して情報量の削減を図った。これは、紙面に書かれた黒い文字を想定した場合には、文字領域に明確なエッジ情報が検出できると考えたためである。

抽出した強いエッジ領域のノイズ除去を行うため、膨張・縮退処理を行うこととした。これにより、詳細な文字輪郭をつぶして、文字の周辺領域でエッジが検出されるように処理することができる。

カメラの前に書面をかざす場合には、移動してきた用紙が一度静止し、また移動していくという画像の流れのなかから、静止しているときだけを見つけ、文字領域を抜き出せば良い。しかし、HMDの前に手に持った紙を静止させようとしても完全な静止はありえない。そこで画像が大きく変化しているか、それとも変化が少ないのかという指針で静止しているかどうかの判定を行うこととした。具体的には、前2フレームと現在のフレーム間でそれぞれ求めているエッジ画像の差分を取り、重ね合わせた画像の中に抽出された差分画素の合計数の大小により静止かどうかを判定することとした。

文字領域の抽出は、重ね合わせた差分画像をさらに膨張処理を行った画像に対し、ラベリングにより連結領域の画素数が一定範囲であり、かつその連結領域を内包する矩形領域内の画素を一定割合以上含む領域を文字候補領域として検出することとした。文字を膨張処理でつぶして検出を行っているため、ここで検出される領域は、1文字単位よりも、行という単位で検出されることを想定している。包含矩形領域内に検出した画素が一定以上含んだ場合としたのは、図形のようなものを誤って検出することを防ぐためである。

4. 実験結果

テスト用のサンプルパターンを用意し、構築した文字領域抽出システムの評価実験を行った結果を図4-6に示す。図4.5の左はテスト用サンプルの画像を2値化したものであり、それぞれ文字領域を検出した結果が右の図である。

本資料では、結果の可視化のため、文字領域検出結果を検出した文字を内包する白い領域で表示してある。この白くあらわした領域の元パターンを取り出すことで文字領域を取り出すことができる。図6は、図4の実験をカメラと紙面の距離を変更した場合の結果である。

実験結果から、適切なカメラとの距離を保った場合には、タイトル表記のようにやや文字も大きく他の領域と分離された文字の並びに関しては、AR Visionからの入力から文字領域を正しく切り出すことに成功した。しかし、図5のように本文の文章が何行にもわたって書かれている場合には、入力画像の画質が低いことが原因で正確に文字領域を検出するには至らなかった。

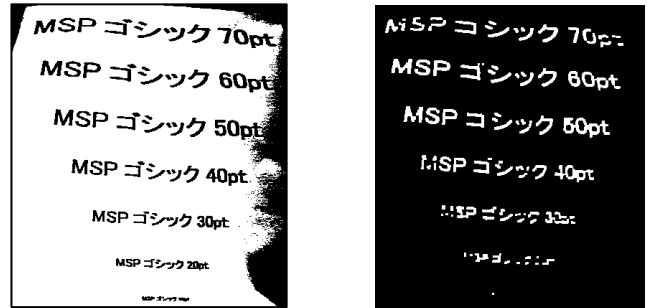


図4. 対象パターンと文字領域抽出結果(成功例)

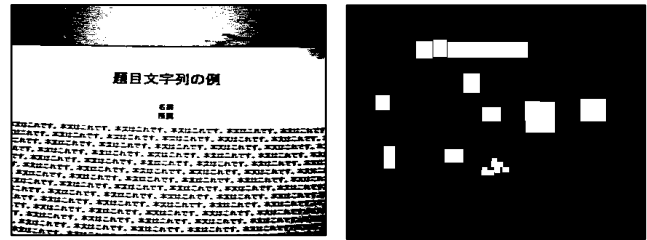


図5. 対象パターンと文字領域抽出結果(失敗例)

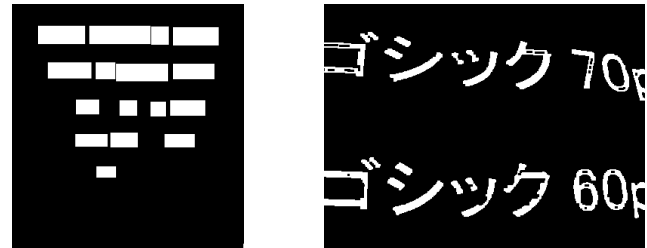


図6. 対象パターンと文字領域抽出結果(距離変動)

5. おわりに

本研究は、カメラ付きHMDのAR Visionを利用して、目前にかざした印刷物の表題部分の文字領域を検出するシステムについて検討を行った。構築システムにより、距離に制限があるものの、見出しに書かれるような文字の大きさであれば、その領域を正しく検出することに成功した。今後の課題としては、現在ポータブルPCでは秒1枚程度しか処理ができていないため高層化を実現し、認識エンジンを組み込んでリアルタイムの資料電子化を実現したいと考えている。

参考文献

[1] 黄瀬, 大町, 内田, 岩村, "カメラを用いた文字認識・文章画像解析の現状と課題", 信学技報 PRMU2005-104(2005)