

2N-6

実環境中に存在する文字情報の自動抽出

遠海由夏<sup>†</sup> 須藤 智<sup>†</sup> 恩田 憲一<sup>†</sup>

尚美学園大学芸術情報学部情報表現学科<sup>†</sup>

1. はじめに

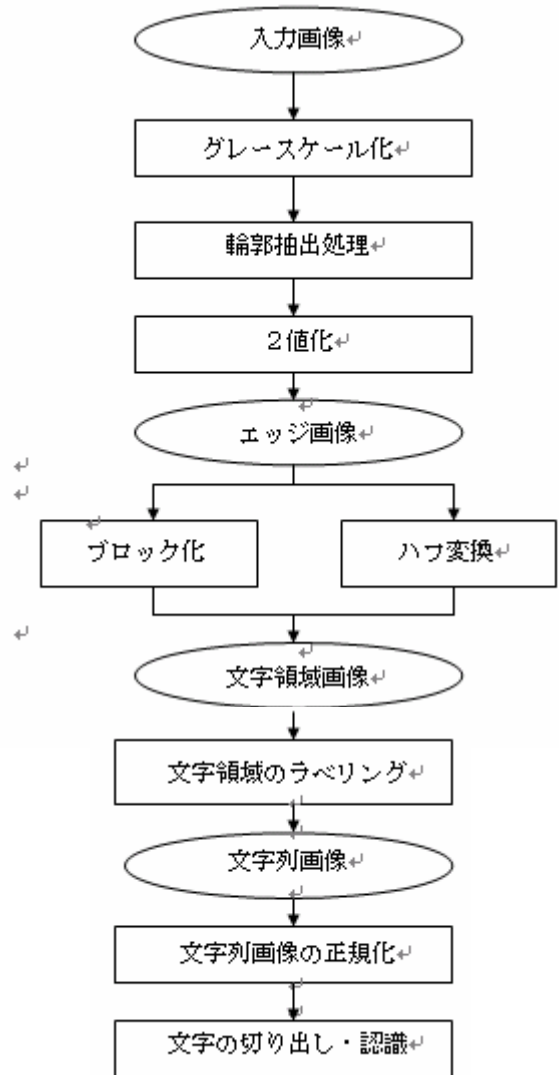
私たち人間は、普段の生活の中で視覚を通じて多くの情報を取得し、それに基づいて行動を決定している。しかし、視覚に障害のある人たちは、聴覚や触覚で補う他に方法が無い。そのため公共の場では、点字ブロックや、音声による案内も行える地図など、障害者を支援する設備が敷かれているが、それらをもってしても、健常者が視覚から得られる情報量に比べれば、不十分である。

そこで本研究では、視覚障害者支援研究の一環として、コンピュータを用いて、シーン中の案内板、看板などの文字情報を、撮影した画像から自動的に抽出、認識を行うことで、利用者をナビゲートするシステムの基盤を作ることを目的としている。

2次元の文字認識は、OCRに代表されるように、既に手法が確立されているが、それに対し3次元での文字認識はまだはっきりと確立されておらず、特に文字を正面から捉えていない場合、現在の手法では認識が困難である。本研究ではそのような文字情報の抽出に主眼を置く。

環境内の文字情報を抽出する研究としては、様々なアプローチがあるがその1つに文字領域の空間周波数と輝度コントラストを用いる手法[1]がある。画像を部分領域に分割後、各部分毎にフーリエ変換を用いて空間周波数を求める手法だが、これは部分領域のサイズに抽出精度が影響を受けやすく、また輝度コントラストは証明条件や色に影響を受けやすいために抽出精度落ちてしまう問題があった。また、それを踏まえて、このような特徴に頼らず、文字の持つ一般的特徴を用いて文字領域を絞り込む手法[2]がある。カラー画像から輪郭抽出を施した後、文字の外接矩形を求めるが、これはすなわちカメラが文字を正面から捉えているという前提条件の下にのみ利用できる手法となる。

2. 処理の流れ



抽出すべき文字と正対していない場合にも対応させるため、関連研究[1][2]のような抽出方法では不十分となる。そのため、微分画像を2のべき乗四方のブロックにまとめ、エッジの集中している領域を抜き出す操作と、同じく微分画像からハフ変換によって直線を抽出する操作、この2つを用いて文字の領域と傾き具合を推定する。

ある程度、文字候補領域が絞れたところで、その各領域にラベリングを行い、直線で文字列と思われる領域を囲む。囲んだ四方形をアフィン変換によって正規化することで、文字を正面

Automatic extraction of characteristic information in real environment.

<sup>†</sup> TOUMI Yuka, SUDO Satoshi, and ONDA Norikazu

Shobi University  
Faculty of Informatics for Arts  
Department of Digital Expression

から捉えた画像を生成する。

そこから先の、文字認識はOCRなど別モジュールに任せることとする。

### 3. 提案手法

入力画像をグレースケール化し、エッジ抽出後、2値画像にすると、文字の存在する領域にはエッジが多く集中する特徴が見えてくる。そこで、余計なエッジを除去し、文字候補領域を見つけ出すために画素をブロック毎にまとめる作業を行う。

隣接した4\*4画素、或いは8\*8画素を一つの画素ブロックとみなし、そのブロック内にエッジ部分(白画素)が多く含まれる場合はそのブロック自体をエッジ領域とみなし白一色に変換、それ以外を黒く塗りつぶす。そうすることでエッジの集中している領域を抽出する。

しかし、これだけでは文字の傾きを算出することは出来ない。そこで、エッジ画像に対しハフ変換を施す。これによって画像内に存在する直線群の傾き具合から、文字の傾きを推定する。

ここまででブロック画像を作成し、エッジの集中している場所を抽出したが、このままでは文字以外の領域も出てしまうため、完全な文字領域画像にはならない。そこで、文字候補領域画像から文字領域を確定していく作業が必要になる。

そこで、画像をいくつかの正方形に分割し、その中のエッジ密度を算出して高いものを文字の存在する領域と仮定し、絞り込みを行う。

更に、文字のある部分には直線も集中しているという特徴を利用し、ハフ変換によって作成した直線抽出後の画像に対しても、同様にエッジ密度を計算する。

ここで2つの2値画像が作成されるため、その画像の和をとり、文字の存在する正方形座標を算出、その結果をブロック画像に適用することで文字領域を確定する。

文字領域は、抽出した時点ではいびつな形をしているため、ハフ変換から求めた傾きを元に、直線で四方を囲み、文字列としてまとめる。これを各文字領域全てに対して行う。

最後に、文字列の傾きを正規化し、文字を正面から見ているような画像に変換する。ここではアフィン変換を用いる。

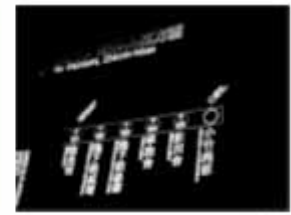
### 4. 実験

実験は、特にGPSの使えない環境での案内情報の発見、という目的のため、駅構内に焦点を絞って実験を行った。デジタルカメラを用いて撮影した画像に対して文字抽出実験を実施した。

撮影した画像の条件として、ホームの案内板など、画像内に必ず文字列が1列以上存在することとした。文字の位置や傾き、証明条件などは制限しなかった。画像サイズは640×480とした。



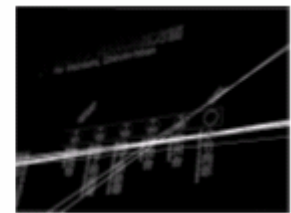
(a)原画像



(b)微分画像



(c)ブロック画像



(d)直線抽出画像

実験の結果どの画像でも、ほぼ人間が認識できる範囲の文字を抽出することが出来た。ただ、強い直線的エッジを持つ非文字領域が存在する場合など、画像によっては文字領域以外の領域が抽出される場合もあった。

また、画像によっては正しく傾きが算出できない場合もあるため、この辺りはまだ改良の余地が見られる。

### 5. まとめ

本研究によって、カメラがどの位置から文字領域を捉えても、人の視覚に近い精度で文字領域の抽出が可能になった。それによって文字領域の位置を知ることの出来ない視覚障害者も快適に利用できるシステムの基盤となるだろう。

### 参考文献

- 1) 劉 詠海, 山村 毅, 大西 昇, 杉江 昇, “シーン内の文字列領域の抽出について”, 電子情報通信学会論文誌, Vol. J81-D- , No. 4, pp. 641-650, (1998-04)
- 2) 鈴木 悠司, 平岩 裕康, 竹内 義則, 松本 哲也, 工藤 博章, 山村 毅, 大西 昇, “視覚障害者のための環境内の文字情報抽出システム”, 電気学会論文誌(C), vol. 124-C, no. 6, pp 1280-1287, (2004-06)