



Title	市街地図における文字・記号抽出の一手法
Author(s)	渡部, 崇也; Watanabe, Takaya; 北島, 秀夫 他
Citation	北海道大學工學部研究報告, 164, 83-87
Issue Date	1993-05-28
Doc URL	https://hdl.handle.net/2115/42363
Type	departmental bulletin paper
File Information	164_83-88.pdf



市街地図における文字・記号抽出の一手法

渡部 崇也 北島 秀夫
白川 智昭 小川 吉彦
(平成 4 年 12 月 25 日受理)

Extraction of Characters and Symbols from City Maps

Takaya WATANABE, Hideo KITAJIMA, Tomoaki SHIRAKAWA and Yoshihiko OGAWA
(Received December 25, 1992)

Abstract

This paper introduces a new algorithm for extraction of characters and symbols from other graphical primitives in city maps.

Their extraction is an essential step in automatic processing of a city map drawn on paper.

The algorithm is intended to solve problems arising from characters/symbols being in contact with other primitives.

It is based on masking operations followed by other separation steps.

Results of computer experiments are included to show the effectiveness of the algorithm.

1. ま え が き

近年、社会の多様化・高度化にともなって、地図情報を計算機上に蓄積しておき、高度な利用を図ることを目的とした、地図のデータベース化技術が注目を集めている。これを実現するためには、まず地図画像上に表現されている図形情報を計算機に入力することが必要となる。

地図情報を計算機に入力する方法として、タブレット板と座標指示器によって構成されるディジタイザをもちいた手入力による方法と、スキャナ等を用いた自動読み取り装置による方法がある。ディジタイザによる方法では、膨大な時間と労力を必要とし、誤りや入力もれが発生する可能性も高い。そこで、地図画像を自動的に読み取り、短時間に効率よくデータを入力する技術が要求されている。

これを実現するためには、各地図要素の種別を認識することが必要となるが、文字・記号要素と他の図形要素とはまったく異なった扱いをしなければならないため、認識に先立って、これらを分離することが有効である。つまり、文字・記号要素を図形要素から分離・抽出することができれば、現在の文字認識技術を用いて自動読み取りを行うことが可能となる。しかし、文字・記号要素が図形要素と接触している場合、これらの分離は困難とされている。本稿では、接触部分をマスク処理を用いて分離した後、各領域について判定を行うことによって、文字・記号を抽出する手法を提案し、国土地理院の 1/25,000 地形図に対して抽出実験を行った結果について報告す

る。

2. 概 要

地図をスキャナで 300 dpi 程度で読み取った際の、図形要素と文字・記号要素との接触部分は、ほとんどの場合 1 画素程度である。そこでまず、図形要素と文字・記号要素が接触していると思われる部分を 3×3 画素のマスク処理を用いて分離する。次に分離された各領域について、塗りつぶし部分と思われる領域、ハッチング部分と思われる領域を除去する。さらに外接矩形の大きさが一定範囲内の領域を抽出する。ただし、このままでは、二つ以上の領域に別れた文字（分離文字）を一つの文字として認識することができないため、膨張・収縮処理によって、これらを融合する。この結果より再度分離枠を設定し、これを文字・記号の候補領域とする。これらの候補領域について以下の条件により文字・記号であるかどうかを判定する。

- ・ループを含むものは文字・記号である。
- ・特徴点(端点, 分岐点, 交差点)を多く含むものは文字・記号である。

3. 抽出手順

3.1 接触要素の分離

図形要素と文字・記号要素が接触していると思われる部分を以下の方法で分離する。

- (1) 原画像全面について、図 1 に示すマスク（およびこれらを回転させたもの）にあてはまる部分の中心画素を分離点候補とする。
- (2) 次の条件を満たす分離点候補を除去することにより、領域を分離する。
 - ・各分離点候補で領域を分離した場合、もとの 1 領域が 2 つ以上の領域に分割される。
 - ・各分離点候補で領域を分離した場合、両方の領域の輪郭長がしきい値 $T1$ 以上となる。
 - ・各分離点候補からしきい値 $T2$ 以下の距離に他の分離点候補が存在しない。

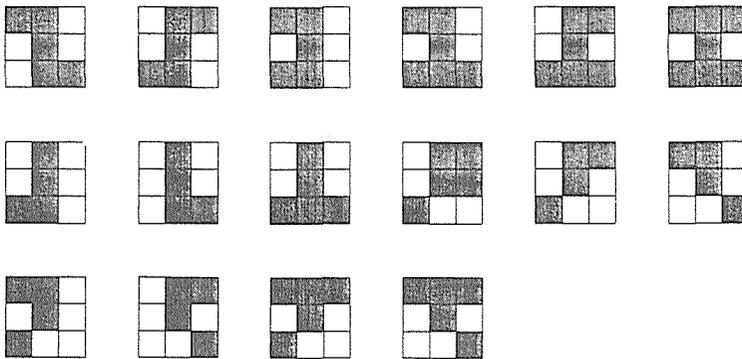


図 1 接触要素分離のためのマスク

3.2 塗りつぶし領域の除去

面状の部分を多く含む領域は塗りつぶし領域であるとして、以下の方法で除去する。

- (1) 各分離枠内の黒画素を着目画素とする。
- (2) 着目画素について、近傍 8 方向への黒画素のランを観測する。
- (3) このランの長さがしきい値 $T3$ 以上の方向が隣り合う 3 方向にある場合、着目画素は面状部

を構成する画素とする。

- (4) 面状部を構成する画素の全黒画素に占める割合がしきい値 T_4 以上の場合、塗りつぶし領域であると判定し、この領域を除去する。

3.3 ハッチング領域の除去

斜め方向の成分を多く含む領域はハッチング領域であるとして、以下の方法で除去する。

- (1) 各分離枠内の連結領域について輪郭追跡を行い、斜め方向の連結数をカウントする。
- (2) 斜め方向の連結数の全輪郭連結数に対する割合がしきい値 T_5 以上の場合、ハッチング領域であると判定し、この領域を除去する。

3.4 分離枠の設定

外接矩形の大きさが一定範囲内の領域を以下の方法で抽出する。

- (1) 輪郭追跡処理で求めた各領域の外周部の各点座標より、各領域に外接する分離枠を設定する。
- (2) この領域が次の条件を満たす場合、文字・記号の候補として抽出する。
 - ・分離枠の縦あるいは横方向の長さがしきい値 T_6 以上である。
 - ・分離枠の縦および横方向の長さがともにしきい値 T_7 以下である。

3.5 分離文字の融合と分離枠の再設定

二つ以上の領域に分かれた文字（分離文字）を以下の方法で融合する。

- (1) 各領域について n 回の膨張処理を行う。
- (2) 膨張処理と同一回数の収縮処理を行う。
- (3) 得られた領域について再度分離枠を設定し、これを文字・記号の候補領域とする。

3.6 各領域に対する判定

文字・記号の候補領域について以下の方法で最終的な判定を行う。

- (1) ループを含むものは文字・記号であるとして、以下の方法で判定する。
 - ・各分離枠内に分離枠と接しない白画素連結領域が存在する場合、文字・記号領域であると判定する。
- (2) 特徴点(端点, 分岐点, 交差点)を多く含むものは文字・記号であるとして、以下の方法で判定する。
 - ・各分離枠内の黒画素領域について細線化処理を施し、特徴点を抽出する。
 - ・特徴点をしきい値 T_8 以上含む場合、文字・記号であると判定する。

4. 抽出実験

4.1 実験方法

前節で述べた文字・記号抽出法を実際の市街地図に適用したので、本節でその結果を示す。処理対象とする地図は国土地理院の 1/25,000 地形図をスキャナによって 320 dpi で読み取り、二値化した後、雑音除去等を行ったもの (256×256 画素)である (図2)。

4.2 結果

原画像から接触要素を分離したものを図3, 塗りつぶし領域を除去したものを図4, ハッチング領域を除去したものを図5, 外接矩形の大きさによる判定を行ったものを図6, 分離文字を融合して分離枠を再設定したものを図7に示す。さらに、各領域について、ループによる判定を行ったものを図8, 特徴点による判定を行ったものを図9に示す。最終的な抽出結果は図10のようになる。本手法により、文字・記号が良好に抽出されている。

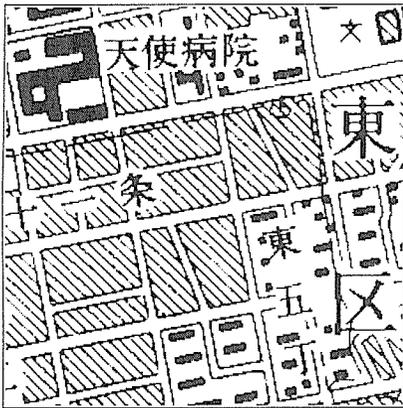


図2 原画像

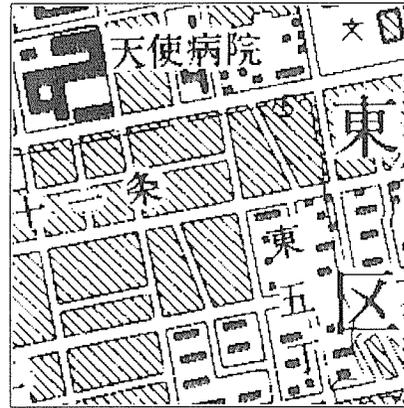


図3 接触要素の分離

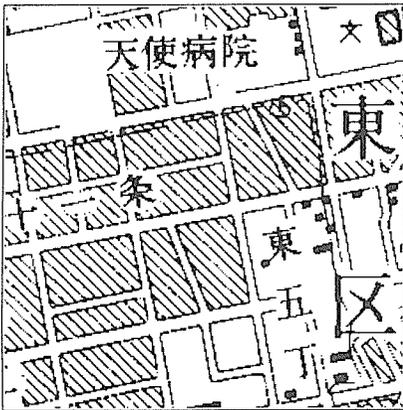


図4 塗りつぶし領域の除去



図5 ハッチング領域の除去

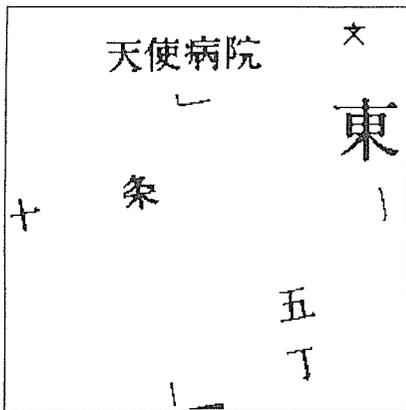


図6 分離枠の設定

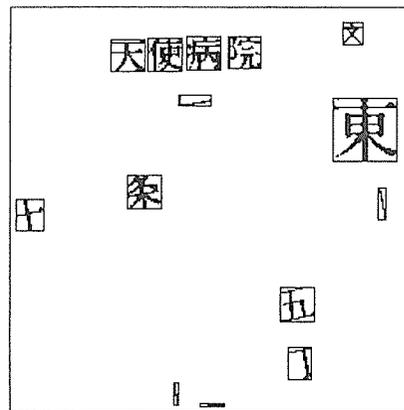


図7 分離枠の再設定

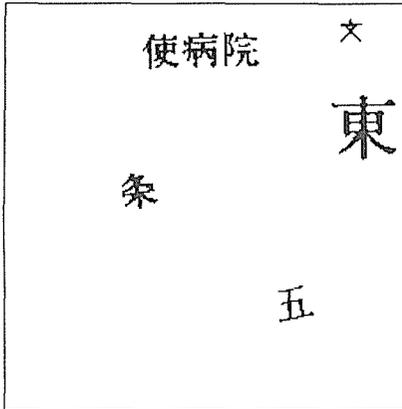


図8 ループによる判定

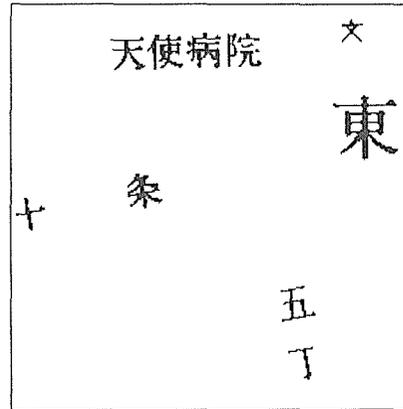


図9 特徴点による判定

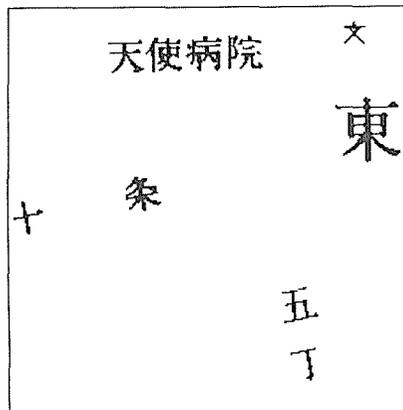


図10 抽出結果

5. ま と め

本稿では図形要素と文字・記号要素が接触しているような市街地図画像中から文字・記号を抽出する一手法を提案した。さらに、実際の地図を用いて本手法による抽出実験を行い、ある程度の接触があっても良好に抽出されることを示した。本手法では文字・記号抽出を一文字、一記号の単位で行っているため、画数の少ない単純な文字・記号を抽出できないという問題点がある。これは、隣接した文字・記号を文字列、記号列と認識することによって抽出が可能であると思われる。この点に関しては今後の検討課題としたい。

参 考 文 献

- 1) 中嶋, 安居院, 吉井, 今井: 地図・図面入力技術動向, (1988), 信学技報, IE 88-37
- 2) 寺嶋, 首藤, 川井, 渡辺: 図面に書かれた文字の切り出し方法, (1981), 情処全大, 6 C-4