

マルチメディアデータ検索のための言語情報を援用した カラー画像中の文字認識*

中村 昌志[†] 鷺坂 光一[‡] 山崎 憲一[§] 尾内 理紀夫[¶]

NTT 基礎研究所^{||}

概要

マルチメディアデータベース構築のためには、マルチメディアデータに対してインデックスを付与しなければならない。その第一歩として、カラー画像にインデックスを付与するために、画像中の文字を認識する手法について検討した。文字認識は、白黒画像中の文字認識法である補完類似度認識法をベースに、それをカラー画像中のカラー文字認識へと拡張した。その結果、高い文字認識率が得られ、また、連続した文字の誤認識率も低く、誤認識された文字の日本語辞書を主体とした言語情報による訂正も有効であった。

1 はじめに

データベース内の情報が、テキストから音、画像といったメディア情報、あるいはそれらが複合したメディア情報へと多様化し、それらを格納したデータベースが互いにネットワークで結合される時代が到来しつつある。その結果、このような世界中に分散された大量のマルチメディアデータベースから所望の情報を効率良く獲得する手法の重要性が増しつつある。

効率的な情報検索のためには、まず、情報そのものに適切なインデックスを自動的に付与しておく必要がある。インデキシングの際には、音、画像といったデータの場合、そこから如何にしてテキスト情報を抽出するかが問題となる。つまり、音声データならば、その音声を如何にして認識し、テ

キスト化するか、画像データならば、その画像中の文字を如何にして認識しテキスト化するかである。画像データからその画像の意味を自然言語で自動的に抽出するという課題もこの延長線上にある。

そこで、我々は、ネットワークで結合されたマルチメディアデータベースの中から、所望の情報をインタラクティブに検索するシステムの構築の第一歩として、データベースのインデックス情報として用いることを念頭に、カラー画像データ中の文字を認識する手法について検討、評価した。

以下、従来の2値化文字認識手法である補完類似度法のカラー画像中の文字認識への拡張法、補完類似度法の高精度の認識結果に基づいて、誤認識された文字を日本語辞書を主体とした言語情報を援用して訂正し、認識率を向上する手法、それら手法の適用実験と評価について述べる。

2 補完類似度法について

補完類似度は2値ベクトル間の類似度を計る尺度である。2値化された画像データのビットマップパターンと、あらかじめ用意しておいた参照ビットマップパターンの各々を2値特徴ベクトルとみなし、それらの間の補完類似度から個別文字認識する方法が補完類似度法である。補完類似度法で

*Recognition of Characters in Color Images Improved with Language Information Intended as an Indexing Method for Multimedia Data Retrieval

[†]Masashi NAKAMURA

[‡]Mitsukazu WASHISAKA

[§]Kenichi YAMAZAKI

[¶]Rikio ONAI

^{||}NTT Basic Research Laboratories

は、参照ビットマップパターンとして様々な文字のビットマップパターンを用意しておき、画像データ中のパターンとの補完類似度が各文字毎に定められたしきい値を越えたときにその文字であると見なす。文字毎のしきい値には、その文字パターンと他の様々な文字パターンとの間の補完類似度の中で最大値のものを使用する。

補完類似度法について詳細は文献 [4]、文献 [5] に述べられているが、かすれや汚れなどの雑音に強く、白黒反転文字も認識できるなどの長所を有する文字認識法である。反面、同じ文字でもフォントの種類の相違などの僅かな差異で認識率が低下するという短所もある。これらの結果として、認識対象の文字フォントがあらかじめ分かっている場合には、汚れやかすれにかかわらず高い認識率を示し、連続して文字を誤認識する可能性は極めて低いという特徴を有する。

3 文字認識

3.1 認識条件

カラー画像中に書かれたカラー文字を認識する場合には、前処理として、画像中の文字領域をあらかじめ切り出ししておく必要がある。本稿では、この文字領域の切り出しについては処理済みであるものとする。また、画像中の文字は単色であり、文字に用いられる字形は既知であるものとする。本稿の実験では、X Window 用の JIS の 24 ドットフォントを実験用画像データと参照ビットマップパターンに使用し、補完類似度を測定している。

3.2 認識の手続き

3.2.1 前処理

補完類似度法は 2 値化画像に対する文字認識法である。この方法をカラー画像中の文字認識に適用するためには、まず、認識対象の画像を 2 値化する必要がある。

減色 認識対象となる画像中の文字は単色であるとしてその文字の色を推定する。しかし、実際のビデオなどから画像データを取り込む場合は、単色に見える文字でも色にある程度のばらつきが生じるため、前処理として画像を減色処理する。このた

め本稿の実験では、RGB それぞれ 8bit から成る原画像の色を上位 3bit に減色する処理を施し、色むらを抑えた後に文字色の推定を行なっている。

文字色推定 画像中の文字色の推定方法として、画像の色ヒストグラムをとり、上位の色を文字色とみなす方法が考えられる。しかし、ひらがな、カタカナをはじめとする画数の少ない文字が書かれた画像を対象とした色ヒストグラムでは、文字を囲む領域全体の画素数に対する、文字自身の画素数は必ずしも多くないので文字の色がヒストグラムの上位を占めるとは限らない。

そこで認識対象となる領域から文字のエッジ抽出を行ない、エッジ近辺の画素のみの色ヒストグラムを生成することにする。その結果、主に文字と背景の境界付近の画素を母集団としたヒストグラムが生成され、文字色が安定してヒストグラムの上位を占めるようになる。ここで推定された文字色の画素を 1、その他の色の画素を 0 とする 2 値化により 2 値化画像を生成する。

文字色はヒストグラム中の最上位を占めるとは限らないが、ヒストグラム中の上位数色以内には安定して入る。本稿の実験では、サンプル当たり平均で 1.93 位に正解の色が入った。そこで、上位数色それぞれに対応する 2 値化画像を生成し、それら全ての画像に対して文字認識を行なう。文字色に対応する 2 値化画像には汚れが少なく、高い精度で文字を認識できる。これに対し、他の色に対応する 2 値化画像からはほとんど全く文字を認識できないため、候補文を生成することがなく、文全体の認識結果にはほとんど影響をあたえない。

3.2.2 文字認識

2 値化された画像中の文字領域に対する文字認識としては、補完類似度法を適用する。2 値化画像内から切り出した 1 文字分の大きさのビットマップパターンとシステムがあらかじめ持っている文字フォントのパターンとを照合し、計算された補完類似度の値があらかじめ計算しておいたしきい値を越えた文字をその領域の文字の候補として挙げる。ここで用いる補完類似度のしきい値は文献 [5] の方法にしたがって文字毎に決めている。また、ビットマップパターンの切り出しに関しては、文献 [5] のずらし照合同様に、文字領域上で 1 文字分の大きさの観測窓を移動させ、観測窓内のピッ

トマップパターンを切り出す方法を用いる。このような処理の結果、画像中の各文字領域における文字の候補が出てくる。

3.3 認識実験結果

表1は実験に用いた例文の統計データである。120の例文は主として、週刊誌、新聞、技術系単行本から採った。カラー画像は6枚用意し、その中から無作為に選択し、X Window用のJISの24ドットフォントを用いた単色文字の例文を、そのカラー画像中に合成して実験用のサンプルとした。単色文字の色は白、赤、青、黄の4色の中から無作為に選択した。

表 1: 文字認識の結果

サンプル文数	120 例 (2604 文字)
サンプル当たり平均文字数	21.7 文字
個別文字当たり認識率	93.4 % (2431/2604)
サンプル当たり平均候補文数	22.2 文
1 文字単独誤認識件数	140 件 (140 文字)
2 文字連続誤認識件数	15 件 (30 文字)
3 文字連続誤認識件数	1 件 (3 文字)
サンプル当たり平均連続誤認識件数	0.13 件

サンプルからの個別文字の認識率は93.4%となった。認識できない文字に関してはその文字の候補を複数挙げ、それら全ての組み合わせで文全体の候補を作った。この候補文の数はサンプル当たり平均で22.2文であった。一文中の連続誤認識件数はサンプル当たり平均で0.13件であり、その大部分は一文字の誤認識であり、2文字以上連続して誤認識する率は、10.3%であった。この結果、言語情報を援用して誤りを訂正する見通しが立った。表2に文字認識についてのデータを示す。

表 2: 文字認識単独での結果

正解文が一意に定まった場合	28 文
複数の候補文中に正解が含まれる場合	92 文
正解率	23.3 %

図2,3に実験に用いたサンプル文例を含んだ原画と、その2値化画像を、表3,4に、そこからの文字認識結果を示す。

補完類似度法による文字認識の結果、認識された文字とその文字の文字領域中の位置とが出力される。認識に用いられたしきい値も同時に出力される。‘位置’は認識された文字のX座標値、‘類似度’は算出された補完類似度、‘しきい値’は認識に用いられたしきい値である。‘類似度’が大きいほど良い認識結果である。この例の場合は表6の複数の候補文に正解が含まれる場合に相当する。

表 3: 文字認識例

オ	:	位置	10,	類似度	222,	しきい値	144
ー	:	位置	34,	類似度	154,	しきい値	113
フ	:	位置	56,	類似度	144,	しきい値	114
ブ	:	位置	57,	類似度	167,	しきい値	122
ン	:	位置	58,	類似度	188,	しきい値	122
ン	:	位置	82,	類似度	160,	しきい値	104
を	:	位置	106,	類似度	227,	しきい値	148
挙	:	位置	130,	類似度	257,	しきい値	167
行	:	位置	154,	類似度	238,	しきい値	154
し	:	位置	178,	類似度	175,	しきい値	158
た	:	位置	202,	類似度	213,	しきい値	198
ホ	:	位置	226,	類似度	217,	しきい値	193
ボ	:	位置	227,	類似度	192,	しきい値	147
テ	:	位置	250,	類似度	198,	しきい値	166
ル	:	位置	274,	類似度	209,	しきい値	135
日	:	位置	298,	類似度	238,	しきい値	197
航	:	位置	322,	類似度	276,	しきい値	181
...							

次に文字の位置と文字幅から認識出力の各文字が何文字目であるかを判定し、文字候補を列方向に記載すると表4のようになる。

表 4: 文字候補

オ	ー	ブ	ン	を	挙	行	し	た	ホ	テ	ル	日	航	...
		ブ							ボ					
		フ												

4 言語情報の援用

文字認識に補完類似度法を用いた結果、連続して文字を誤認識することが少なく、高い文字認識率が得られ、言語情報を利用可能な条件が整った。以下、言語情報を援用し、複数の候補文の中から、もっとも日本語文としてふさわしい候補文を選択することについて述べる。

4.1 形態素解析

形態素解析は、辞書と形態素間の接続規則を用いて、文字列を形態素単位に区切り、その品詞や活用形等を明らかにする言語解析処理の一つである。

文字認識では単語辞書を用いることで、単語レベルで誤認識を減らす処理を行なうが、本研究では日本語文の認識を目的にしているため、まず単語ごとに区切る必要があり、日本語形態素解析を用いる。

本研究ではEDRの日本語辞書(文献[8])と形態素接続規則をもとにして、日本語形態素解析を行なうことで、分割された形態素が日本語辞書に記載されているかどうか、また、形態素間の接続規則をもとに日本語の接続規則を満たしているかを調べる。候補の選択に際しては、接続規則ごとにコストを付けて、各コストの総合が最小になる文を日本語文として最適のものとして選択した。

ただし、日本語形態素解析だけで意味のレベルまでは立ち入らない場合、単語が辞書に載っておりかつ品詞が同じ時には一意に結果が定まらない場合があるという問題がある。

4.2 解析結果

文字認識単だけでは一意に認識できなかったサンプル文(92文:表2参照)に対し、前記の形態素解析による言語情報を援用した結果を表5に示す。

表5: 形態素解析による結果

正解文が一意に定まった場合	79文
複数の候補文中に正解が含まれる場合	12文
正解文が含まれない場合	1文
正解率	85.9%

文字認識により得られた候補文字の組合せで候補文を生成し、それぞれの候補文に形態素解析による評価を行なった出力例を表6に示す。文中の「|」は形態素の区切りを表し、各文の直後に表示される「総合コスト」が日本語としての繋ぎの良さ(「総合コスト」が小さいもの程よい)を、「未定義語数」は文中の辞書にない言葉の数を、「形態素数」は形態素の数を表す。この例文の場合、総合コストが最小になった候補文が2つあり、その中に正解文が含まれるので、表5では「複数の候補文中に正解が含まれる場合」に分類されている。総合コストが最小の候補文に正解文が含まれない場合は、表5では「正解文がふくまれない場合」に分類される。なお、「正解文が含まれない場合」の1文の原因は、サンプル文中に日本語辞書に載っていない未定義の動詞があり、正しく形態素解析出来なかったためである。

まれる場合」に分類されている。総合コストが最小の候補文に正解文が含まれない場合は、表5では「正解文がふくまれない場合」に分類される。なお、「正解文が含まれない場合」の1文の原因は、サンプル文中に日本語辞書に載っていない未定義の動詞があり、正しく形態素解析出来なかったためである。

表6: 候補文の解析例

|オープン|を|挙行|し|た|ホテル|日航|東京|は|お台場|人気|で|連日|満員|

未定義語数 = 0, 総合コスト = 277, 形態素数 = 14

|オープン|を|挙行|し|た|ホテル|日航|東京|は|お台場|人気|で|連日|満員|

未定義語数 = 0, 総合コスト = 277, 形態素数 = 14

|オープン|を|挙行|し|た|ホテル|日航|東京|は|お台|湯|人気|で|連日|満員|

未定義語数 = 0, 総合コスト = 300, 形態素数 = 15

|オープン|を|挙行|し|た|ホテル|日航|東京|は|お台|湯|人気|で|連日|満員|

未定義語数 = 0, 総合コスト = 300, 形態素数 = 15

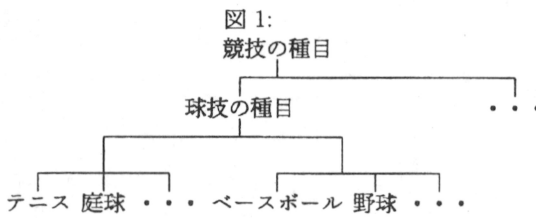
...

4.3 単語間の概念距離の援用

EDRの辞書の中に概念辞書がある。EDR概念辞書は、概念間の上位、下位、兄弟関係を規定している。そのような関係の一例を図1に示す。競技の種目という概念の下位概念の一つが、球技の種目となっている。EDR概念辞書の最上位概念は「概念」である。

本稿では、兄弟関係の概念距離は0、上位、下位関係は一段について概念距離1としている。よって、野球とテニスの概念距離は2である。

表6において先頭の2候補文は総合コストがともに277である。ここにおいて、オープンとオープンの、他の名詞(含サ変動詞の語幹)との概念距離を計算すると、候補オープンは51、候補オープンは72となり、 $51 < 72$ であるから、候補オープンの選択が可能となる。ただし、表5の「複数の候補文中に正解が含まれる場合」の12文中で、概念



距離が有効なのは1文であり、他は、単語辞書に固有名詞未登録なもの4文、人間が見てもどれが正解か判定できないもの4文、それ以外の3文はさらに深い意味処理をすれば判定可能なものである。

5 他研究との比較

文字認識の結果に言語情報を援用することで、認識率を改善する手法に関しては多くの研究が行なわれてきた(文献[1], [2])。また、実際のカラー映像データから文字情報を抽出し、それに言語情報を援用して、インデキシングの確度を向上させる研究も行なわれている(文献[7])。さらに、言語情報として共起関係という意味情報の援用も小規模ではあるが試みられている(文献[3])。これらの研究と、本研究の差は、

- 雑音に強い補完類似度法を初めてカラー画像に適用した
- 誤認識文字の訂正のための形態素解析において、形態素間の接続にコストを付けている
- 大規模辞書をベースにした意味情報の援用を試行していること

である。

6 考察

本稿では、白黒画像中の文字認識法である補完類似度認識法を、文字色推定と2値化とにより、カラー画像中のカラー文字認識へと拡張することが可能であることを初めて示すとともに、93.4%の文字認識率を達成した。

文献[1]にも述べられているように、文字認識段階の精度が低いと、後の言語処理の段階に多大な処理量、および精度が要求されることが知られている。本研究においては、雑音に強い補完類似度

法をベースにした結果、言語情報の援用が可能な高い文字認識率を達成するとともに、連続文字を誤認識する割合が低いため、言語情報の援用による誤り訂正が有効となる条件が整った。

形態素解析を援用した結果、85.9%の誤り訂正率を達成することができ、マルチメディアデータの一つであるカラー動画画像中の文字によるインデキシングに関して見通しを得た。

7 今後の予定

本稿では文字領域の切り出しは済んでいるものとして実験用サンプルを用意したが、実際に精度良く文字領域を切り出す方法に関しては今後の課題である。また、本稿の実験では文献[5]と同様に1行分の文字領域に対し、ずらし照合による認識を行なっているため、文字領域内の個別文字を切り出す必要はなかった。しかし、個々の文字を高精度で切り出すことができれば文字認識処理の高速化が期待できるので、これも今後の課題である。

補完類似度法による文字認識ではデータ中の文字の字体が参照ビットマップパターンのもものと異なる場合に弱い。そこで、複数種のフォントを参照パターンとして持たせ、多種類の字体に対応する、即ち、マルチフォント対応についても課題である。

さらに、言語情報として意味情報を援用する手法に関しては、概念関係だけでなく、共起関係を考慮に入れるとともに、意味情報の援用に関する多くのデータを蓄積することも課題である。

謝辞

本研究を進めるに当たって御教示、御討論いただきました、基礎研究所澤木美奈子研究主任、佐藤孝治研究主任、コミュニケーション科学研究所萩田紀博研究企画部長と、本研究を御支援いただいた石井健一郎情報科学研究部部長に感謝します。また、多岐に渡って御討論いただきましたセマンティックコンピューティング研究グループの皆様にも感謝します。

参考文献

- [1] 三部裕史, 大森健児: 信頼性の低い文字認識結果に対する言語情報を用いた誤認識文字の

訂正, 情処論文誌, Vol.34, No.10, pp2117-2124, 1993.

- [2] 西野文人: 文字認識における自然言語処理, 情報処理, Vol34, No.10, pp1274-1280, 1993.
- [3] 紺野章子, 本郷保夫: 日本語 OCR の後処理に関する一手法, 信学技報, PRU92-21, pp.23-30, 1992.
- [4] 澤木美奈子, 萩田紀博: 補完類似度による劣化印刷文字認識, 信学技報 PRU95-14, pp.101-108, 1995.
- [5] 澤木美奈子, 萩田紀博: 補完類似度に基づく新聞見出し文字の領域流出と認識, 信学技報 PRU95-106, pp.19-24, 1995.
- [6] 大谷 淳, 塩 昭夫: 情景画像からの文字パターン抽出と認識, 信学論 D, Vol.J71-D, No.6, pp.1037-1047, 1988.
- [7] 茂木祐治, 有木康雄: ニュース映像中の文字認識に基づく記事の索引付け, 信学技報, IE95-I53, PRU95-240, pp.33-40, 1996.
- [8] (株)日本電子化辞書研究所: EDR 電子化辞書 1.5 版 仕様説明書, 1996.

図 2: 原画像

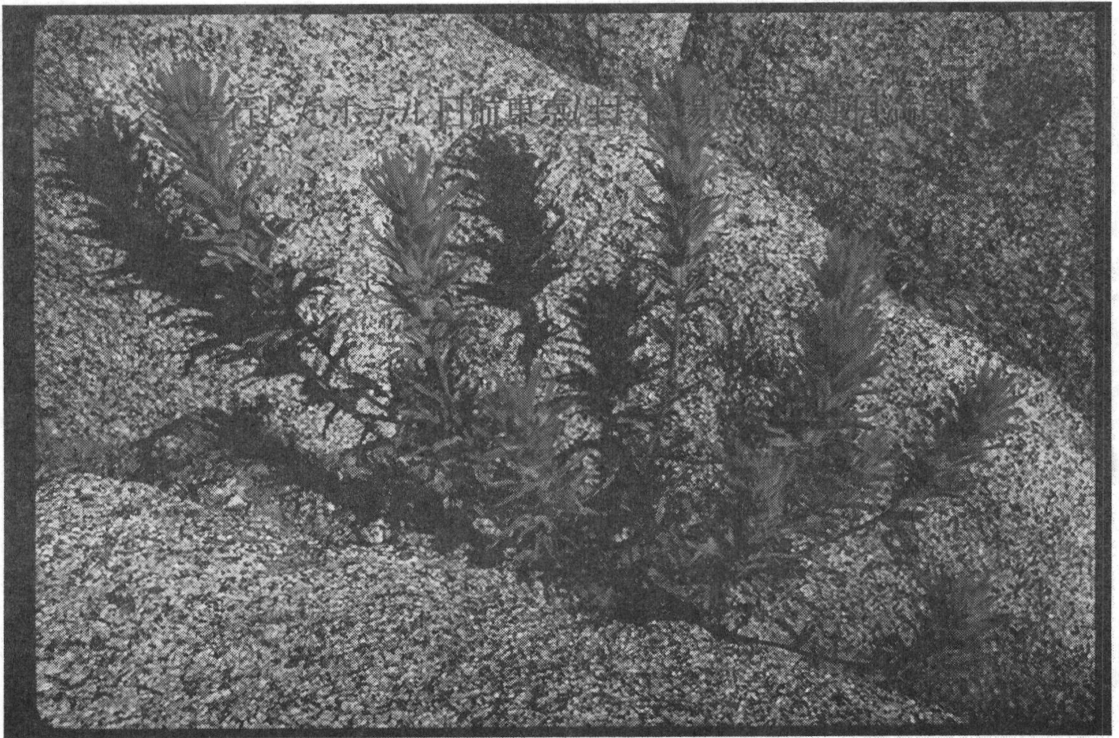


図 3: 2 値化画像

オープンを挙行したホテル日航東京はお台場人気で連日満員

本 PDF ファイルは 1997 年発行の「第 38 回プログラミング・シンポジウム報告集」をスキャンし、項目ごとに整理して、情報処理学会電子図書館「情報学広場」に掲載するものです。

この出版物は情報処理学会への著作権譲渡がなされていませんが、情報処理学会公式 Web サイトに、下記「過去のプログラミング・シンポジウム報告集の利用許諾について」を掲載し、権利者の検索をおこないました。そのうえで同意をいただいたもの、お申し出のなかったものを掲載しています。

https://www.ipsj.or.jp/topics/Past_reports.html

過去のプログラミング・シンポジウム報告集の利用許諾について

情報処理学会発行の出版物著作権は平成 12 年から情報処理学会著作権規程に従い、学会に帰属することになっています。

プログラミング・シンポジウムの報告集は、情報処理学会と設立の事情が異なるため、この改訂がシンポジウム内部で徹底しておらず、情報処理学会の他の出版物が情報学広場（＝情報処理学会電子図書館）で公開されているにも拘らず、古い報告集には公開されていないものが少からずありました。

プログラミング・シンポジウムは昭和 59 年に情報処理学会の一部門になりましたが、それ以前の報告集も含め、この度学会の他の出版物と同様の扱いにしたいと考えます。過去のすべての報告集の論文について、著作権者（論文を執筆された故人の相続人）を探し出して利用許諾に関する同意を頂くことは困難ですので、一定期間の権利者搜索の努力をしたうえで、著作権者が見つからない場合も論文を情報学広場に掲載させていただきたいと思います。その後、著作権者が発見され、情報学広場への掲載の継続に同意が得られなかった場合には、当該論文については、掲載を停止致します。

この措置にご意見のある方は、プログラミング・シンポジウムの辻尚史運営委員長 (tsuji@math.s.chiba-u.ac.jp) までお申し出ください。

加えて、著作権者について情報をお持ちの方は事務局まで情報をお寄せくださいますようお願い申し上げます。

期間：2020 年 12 月 18 日～2021 年 3 月 19 日

掲載日：2020 年 12 月 18 日

プログラミング・シンポジウム委員会

情報処理学会著作権規程

<https://www.ipsj.or.jp/copyright/ronbun/copyright.html>