情報処理学会論文誌　Vol. 51　No. 9　1520–1529 (Sep. 2010)

文字認識信頼度と視覚的特徴量を用いた画像内文字検索システムの提案

益子　宗[†1]　星　鉄矢[†2]　平野廣美[†1]

従来からインターネット上の電子商店街（サイバーモール）において，モール管理者が電子商店に違法な誇大表現が掲載されていないかどうかテキスト検索技術を用いて日々監視することが行われてきた．しかし，近年テキスト表現を埋め込んだ画像を利用することが多くなり，既存のテキスト検索だけでは発見できないことが問題となっている．そこで，本稿では文字検索誤りの程度を表す文字認識信頼度，画像内文字列の視覚的特徴量，出現頻度を考慮することで，検索キーワードに基づく画像の新たなスコアリング手法を提案し，画像内文字検索システムを構築した．

A System for Searching Texts in Images by Using Reliability Factor and Saliency

Soh Masuko,[†1] Tetsuya Hoshi[†2] and Hiromi Hirano[†1]

In electronic shopping malls consisting of a lot of retail shops, the mall managers have to watch for illegally exaggerated expression uploaded by the shop owners. Though this has been done by using text search technologies, many shops now tend to use images embedded with various texts, which cannot be searched by text search technologies. We propose a novel text recognition system that scores images based on reliabilities of recognized characters with respect to a search text, visual saliency, and term frequency. We applied the system to a high amount of images used in shops of a working Internet electric shopping mall.

†1 楽天株式会社楽天技術研究所
Rakuten Institute of Technology, Rakuten, Inc.
†2 ビーグル株式会社
Beagle, Inc.

1. はじめに

近年，楽天市場や Yahoo! ショッピングといった複数の店舗のページ（電子商店）を 1 つのサイトにまとめて様々な商品を販売する電子商店街（サイバーモール）が普及し利用されるようになった[1),2)]．その一方で，健康食品や化粧品の効果を消費者に過剰に期待させるような誇大表現を電子商店に掲載し，薬事法の違反により出店者が逮捕される事件が発生するなど社会問題化している．そのため，サイバーモール管理者には電子商店に掲載されている不正な文言を見つけ出し，出店者に対して警告するなどの対応が求められている．

これまでモール管理者はあらかじめ選出した不正な表現となりやすいいくつかのキーワードを組み合わせてテキスト検索し，該当するキーワードが使われている文章を確認することで，電子商店に不正な表現が掲載されていないかを監視してきた．しかし，ブロードバンドの普及にともない，多くの電子商店では様々なフォントやグラデーションにより装飾された文字を埋め込んだ商品説明画像が多用されるようになり，テキスト検索では監視ができなくなってきている．

そこで，あらかじめ人手によりテキスト化された画像に含まれる文字列を検索することが考えられるが，サイバーモール上に存在する大量の画像内文字をテキスト化することはコストの面から実用的であるとはいい難い．文書画像の文字認識に用いられる OCR（Optical Character Reader）技術を利用して画像から得られた認識結果を利用して検索[3),4)] することも考えられるが，サイバーモール上の商品説明画像のように複雑なレイアウト構造や飾り文字を含んだ画像に対しては多くの認識誤りを含むため，認識結果をそのまま検索に利用することは困難である．また，従来の画像内文字の検索では単一の文書画像内に含まれる単語を検索することにとどまっており，サイバーモール内に存在する複数の商品説明画像を横断的に検索した場合，検索キーワードにどの程度適合した画像なのかを判断できず，モール管理者は大量にリストアップされた検索結果画像の中から不正な表現を見つけなければならないため非常に効率が悪い．

そこで本稿では画像内の文字認識誤りに起因する検索漏れを低減させるために，文字認識結果を多重化することで再現率を高める画像内文字の検索手法を提案する．さらに，検索キーワードに該当する画像内文字列の文字認識信頼度と視覚的特徴量，出現頻度を検索結果の精度の指標とした検索キーワードに基づく新たな画像のスコアリング手法を提案する．最後に提案手法を用いたプロトタイプシステムを作成し，実際のサイバーモール内の電子商店に存在する約 56 万枚の画像を対象に実験を行った．その結果，テキスト検索では見つ

けることができなかった不正表現を含む画像を平均して 60%の再現率，平均検索時間が約 350 ミリ秒で見つけることができ，本稿で提案した画像内文字検索システムが画像内に不正表現を含む電子商店のページを検知することに有用であることを示した．

2. 関 連 研 究

従来からインターネット上の画像を見つけ出す需要は多くあり，著作権管理や違法コピー防止といった観点からの取り組みとして，企業のロゴマークやアイドルの写真が不正利用されていないかを画像特徴量をもとに検索するシステム[5] や，画像周辺のテキスト情報を利用して任意の検索単語に関連する画像を提示する画像検索システムも実用化されている[6]．また，画像特徴量だけでなく画像内に埋め込まれた電子透かしを利用し，画像の不正な 2 次利用を防ぐ著作権管理（DRM）技術の取り組みも行われている[7),8)]．しかしそれらは，画像内の文言を検出することを目的としていないため，本研究で対象としている画像内に不正な表現を含むかどうかを判定することは困難である．

一方で，商用の OCR ソフトを利用して画像から得られたテキストの認識結果を検索できるコンシューマ向けの製品が実用化されている．しかし，これらの製品では画像内の文字の検索機能は提供されているが，複数の画像を並べ替えるためのスコアリングについて検討されておらず，ユーザが求める画像への効率的なアクセスが困難である．また，それらの製品に利用されている OCR エンジンの多くは文書画像のようにあらかじめレイアウトが決まった画像を対象としたものが多く，サイバーモール上の商品説明画像のような複雑なレイアウト構造や飾り文字を含んだ画像に適用した場合，多くの認識誤りを含むためそのままの利用は困難である．

そのような問題を解決するために，文書画像を対象として認識誤りを含む認識結果を利用した文字列検索手法が提案されている．たとえば近藤らの研究では検索用キーワードでデータベースから文書画像を検索する際に，あらかじめ用意しておいた類似文字テーブルを用いて検索用キーワードを拡張することで誤認識の影響を吸収する手法を利用している[9)]．また，太田らの研究では単一の文字認識結果に対してコンフュージョンマトリックスなどの誤認識の傾向を用いて，検索キーワードとの関係を複数の可能性を考慮して検索を行っている[10)]．しかし，それらの手法ではあらかじめ類似文字テーブルやコンフュージョンマトリックスを用意する必要があり，用いる文字抽出手法や文字認識手法に応じてそれらの変換表を変更する必要があるため，汎用的な手法とはいい難い[11),12)]．

このように従来の不正画像を検知する研究では画像内の文字表現に注目した取り組みは少なく，また，既存の文書画像を対象とした文字認識技術を利用した検索手法では電子商店に存在する商品説明画像のように複雑なレイアウトを持ち，様々なフォントやグラデーションにより装飾された文字を埋め込んだ商品説明画像に対しては，本稿で目的としている結果を得ることは難しい．

3. 提案手法の概要

本稿では上述した問題点を克服するために，次の方針に基づいて画像内文字検索システムを開発した．

(1) 画像内文字の認識は文字の抽出，文字切り出しに起因する文字認識誤りを生じるため，文字認識処理により得られた第 1 候補だけでなく，複数候補を利用することで認識結果を多重化し再現率を向上させる．

(2) 多重化された認識結果を検索した場合，文字認識結果の第 1 候補のみを用いて検索する場合に比べ誤検出が発生し適合率が低下する．そのため，文字候補順位を利用した文字認識信頼度を検索キーワードにマッチする度合いとして用いる．

(3) 画像内の検索キーワードに該当する文字列の色やサイズといった視覚的特徴と画像内でのキーワードの出現頻度を，文字認識信頼度と同時に計算することで検索キーワードに基づく画像のスコアを求める．

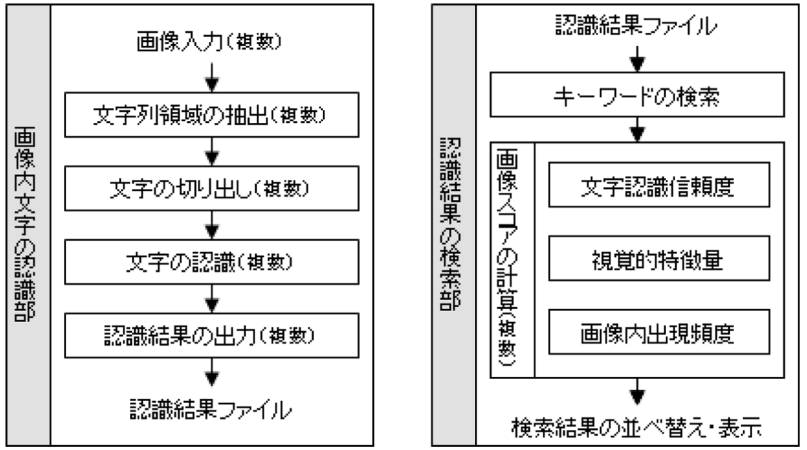

図 1　画像内文字検索システムの概要
Fig. 1　An overview of a system for searching texts in images.

図 1 にシステムの構成を示す．まず，画像内文字の認識部では複数の入力された画像に対して，文字認識処理した認識結果を外部ファイルに出力する．次に，検索部では入力された検索キーワードを認識結果ファイルの中から探し，検索キーワードを含む画像のスコアを計算する．最後に画像のスコアをもとに検索結果をソートし表示する．画像内文字の認識部については 4.1 節，認識結果の検索部は 4.2 節，検索キーワードに基づく画像のスコア計算については 5 章で詳細を述べる．

4. 画像内文字の認識と検索

4.1 文字の抽出と認識

3 章で述べた画像内文字の認識部では，画像内に含まれる文字列領域を抽出し，同領域から文字画像を 1 文字ごとに切り出し文字認識処理を行う．まず，画像内の文字を抽出するために対象画像をグレースケール画像に変換した後，判別分析法により閾値を決定し 2 値画像に変換する[13]．次に 2 値画像にラベリング処理を行い，得られた画素連結要素をピッチ，縦横比，角度を用いて領域を連結し横方向と縦方向に並んだ文字列画像を抽出する[14]．最後に，形態素解析を用いて文字列画像を分割し，得られた文字画像に対して文字認識処理を行った．ただし，文字認識に用いる特徴量は文字の輪郭を利用した方向線素特徴を用いた[15]．さらに本手法では，特徴量のユークリッド距離が近いものから対象文字画像に対する文字候補に順位付けを行い，第 N 位までの多重化された認識結果を保持することで，文字認識誤りによる検索漏れを回避する．本手法によって得られた文字認識結果の例を**表 1** に示す．表 1 中の文字座標は画像左上を原点 $(0, 0)$ とした場合の文字画像の (x 座標，y 座標，幅，高さ) を示し，認識結果 C の j 番目の文字候補第 n 位を C[j][n] と表現する．たとえば，表 1 の認識結果 C における C[1][1] と C[1][2]，C[10][1] は，それぞれ「そ」と「予」，「高」となる．ただし，本稿で対象としている商品画像は背景や文字飾りがあるなど文書画像と比べ複雑なレイアウトであるため，画像から精度良い文字認識結果を得ることは挑戦的な課題である．そのため，本稿では画像内文字の認識精度を改善することには注目しないが，本節で利用した手法は他の認識手法に容易に置き換えることができ，今後拡張することが可能である．

4.2 多重化された認識結果からのキーワード検索

本手法では 4.1 節で得られた N 個の文字候補の中から任意の文字列を**図 2** に示す手順により検索を行う．まず，検索キーワードの 1 文字目 Keyword[1] に該当する文字が認識結果 C の j 番目に含まれるかを調べるために，C[j][n] と Keyword[1] を比較する．次に検索

表 1　画像内文字の認識結果の例
Table 1　Example of recognition results.

No.j	文字座標	認識結果（n=1～30 ※文字候補順）
1	200,326,21,26	そ,予,干,亡,ぞ,モ,辛,子,素,キ,中,午,チ,ヤ,ナ,壬,ヂ,ャ,平,千,乎,幸,吃,セ,七,垂,華,手,玉,ぇ
2	228,326,25,26	の,ｃ,ｏ,０,Ｏ,ぬ,わ,Ｃ,ｗ,ゆ,ゅ,Ｑ,刀,Ｇ,力,め,わ,Ｄ,曲,ひ,幻,眼,巧,ｑ,な,乃,嬢,む,壌,巾
3	258,326,27,26	安,妾,妥,妄,要,垂,巽,宙,錠,ま,異,塞,去,宝,黄,芝,云,支,珪,民,審,寅,菱,曳,興,衷,宍,表,変,突
4	290,326,26,26	全,金,合,会,生,至,台,圭,企,奄,歪,主,杢,今,令,楚,垂,余,土,里,空,竺,堂,建,弁,串,奔,舎,翌,営
5	321,326,27,26	性,牲,姓,惟,住,世,佐,仕,桂,陸,蛙,柱,健,牡,杜,挫,佳,任,は,注,裡,接,甘,娃,催,侶,壮,睦,雄,隆
6	354,326,18,26	と,ヒ,Ｌ,Ｅ,七,ム,ム,占,ど,亡,ピ,６,ｂ,包,む,ｈ,セ,８,杜,止,也,た,よ,女,右,壮,ょ,生,ｔ,廿
7	379,326,26,26	品,吊,晶,足,昂,串,毘,嘉,頗,男,昆,常,兄,呂,話,帯,虫,甜,鼎,窃,岱,舵,枯,摘,弔,賠,出,賭,用,思
8	411,326,25,26	貿,賃,賀,宵,寛,貫,貸,管,貨,胃,貞,賛,首,皆,腎,賞,貢,背,賃,貴,曽,真,音,箕,晋,曾,資,笛,買,育
9	442,326,26,26	の,０,力,Ｏ,ｗ,刀,ぬ,ｏ,Ｑ,わ,ゆ,Ｃ,ｃ,曲,巧,ゆ,眼,ｑ,幻,乃,櫓,め,カ,Ｇ,賦,巾,穣,ヵ,的,餌
10	474,326,26,26	高,言,葛,盲,曹,吾,寓,菖,宮,富,官,曽,旨,首,喬,害,唐,亭,賞,胃,冨,音,島,吉,貫,昌,普,貢,嵩,畠
11	506,326,17,26	さ,き,包,る,白,ざ,を,圭,笛,古,ま,忘,占,色,労,甘,宙,も,ょ,ａ,ぎ,右,８,承,索,由,自,主,3,Ｅ
12	530,326,24,26	か,が,九,井,ァ,ア,丸,分,柾,ガ,Ａ,八,凋,朴,穴,ケ,介,サ,廿,曲,林,沖,ｄ,ひ,ｂ,カ,允,抽,油,功
13	559,326,16,26	ら,ち,5,ｂ,６,Ｂ,Ｅ,８,日,3,Ｓ,も,弓,ｈ,る,う,右,ぢ,包,石,且,Ｇ,好,埼,ろ,Ｆ,ぅ,南,方,ム

```
入力: Keyword 検索キーワード,C 認識結果, N 文字候補数
出力: 検索成功の場合 similarity 検索結果の文字認識信頼度
i=1;
totalrank=0;
for j = 1 .. C.length do
 for n = 1 .. N do
  if C[j][n]==Keyword[i] then
   i=i+1;
   totalrank = totalrank +n;
   goto FOUND_MATCH_FOR_I;
  else if n==N then
    j=j-i+1;
    i=1;
    totalrank=0;
  end if
 end for
FOUND_MATCH_FOR_I:
 if(i==Keyword.length) then
   similarity = totalrank / Keyword.length;
   goto FOUND_KEYWORD;
 else if(j==C.length) then
   return false;
 end if
end for
FOUND KEYWORD:
return similarity;
```

図 2　多重化された認識結果を用いた検索手順
Fig. 2　Proposed search method using multiple recognition results.

表 2　検索キーワードと文字認識信頼度の関係（“絶対痩せる” を検索した場合）
Table 2　Relation between a search query word and reliability of recognized characters.

	文字候補 n					
画像①	1	2	3	4	5	…
絶	絶	難	組	稚	稚	継…
対	対	射	封	効	苅	吋…
痩	痩	癯	喪	痺	癒	煙…
せ	せ	ぜ	甘	ゼ	廿	セ…
る	る	ろ	ら	ち	承	埼…

	文字候補 n					
画像②	1	2	3	4	5	…
絶	絶	難	組	稚	稚	継…
対	対	射	封	効	苅	吋…
痩	痩	癯	喪	痺	癒	煙…
せ	せ	ぜ	甘	ゼ	廿	セ…
ろ	ろ	6	る	5	ら	ム…

キーワードの 2 文字目以降 Keyword[i] が C[j][1] から C[j][N] までに存在するかを順に調べる．最終的に検索キーワードの長さ Keyword.length 分文字が発見できた場合，検索が成功するものとする．

また，本手法では文字認識結果の第 1 候補のみを用いて検索する場合に比べ，検索漏れの低減（再現率の向上）が期待できるが，同時に認識誤りを多く含む検索誤りが増加する（適合率が低下する）ことが考えられる．そこで，本手法では 4.1 節で得られた文字候補順位を利用し文字認識信頼度 *similarity* を式 (1) により求めることで検索キーワード t に対する一致度（0.0〜1.0）を表現する．ただし，第 1 候補のみで検索キーワードにマッチした文字列の文字認識信頼度は 1.0 となる．

$$similarity(t) = Keyword(t).length / totalscore(t) \quad (1)$$

なお，式 (1) 中の $Keyword(t).length$ は検索キーワード t の長さ，$totalscore(t)$ はマッチした文字候補の順位の合計とし，検索と同時に文字候補順位を *totalrank* として加算し，最終的に文字認識信頼度 *similarity* を得る．

たとえば，**表 2** に示した 2 つの画像 ①，② のそれぞれの文字認識結果から「絶対痩せる」というキーワードで検索した場合の文字認識信頼度は $5 \div (1+1+1+1+1)$, $5 \div (1+1+1+1+3)$ により計算され，それぞれ 1.00 と 0.71 となる．これにより，文字認識信頼度が低い画像は誤検索されている可能性があり，文字認識信頼度が高い画像は検索キーワードをより正確に含んでいるという指標として利用できる．そのため，大量の画像の中から検索キーワードを含む画像をリストアップする際に文字認識信頼度をもとに検索結果をソートすることで，検索誤りが少ない結果の画像を優先的に提示することが可能となる．

5. 検索キーワードに基づく画像スコアの計算

5.1 不正表現を含む画像の特徴

4.2 節で求めた文字認識信頼度をもとに検索結果をソートし表示することで，検索誤りの少ない画像から優先的に確認することが可能となる．しかし，検索キーワードを含む画像であっても，画像が使われる文脈によっては不正な表現とならない場合も多く見られる．そのため，サイバーモール上に大量に存在する検索キーワードを含む画像を文字認識信頼度順にリストアップするだけでは不正な表現を含む画像を効率的に見つけることが困難である．そのため，本章では画像内の検索されたキーワードの色やサイズといった視覚的特徴と画像内でのキーワードの出現頻度をもとに，検索キーワードに基づく画像スコアを求める．

予備実験としてモール管理者が事前に不正な画像であると判断した 674 枚の画像を目視で確認した．その結果，不正表現を含む画像には，

(1)　不正単語が視覚的に目立つものが多い
(2)　不正単語の出現頻度が高い
(3)　画像内に複数の不正単語が含まれる

という傾向があることが分かった．そこで，これらの知見をもとに，① 画像内の文字列の視覚的特徴，② 画像内キーワード出現頻度の 2 つの要素から画像の重要度を計算する．ただし，視覚的特徴はキーワードに該当した画像内文字列のサイズと色，画像内キーワード出現頻度は画像内に含まれる検索キーワードの数とする．

このように画像の重要度を計算することで，同一文字列が含まれる画像でも，小さい文字で説明されている画像に比べ，タイトルなどの大きな文字で表記されている画像の場合にスコアが高くなり，より視覚的に目立つと同時に不正である可能性が高い表現を含む画像を見つけることが期待できる．① 画像内の文字列の視覚的特徴については 5.2 節で述べ，5.3 節では 4 章で述べた文字認識信頼度と ① 画像内の文字列の視覚的特徴，② 画像内キーワード出現頻度を利用した画像のスコアリング手法について述べる．

5.2 文字列の視覚的特徴量の計算

人間は周囲の視覚刺激の中で異なる属性を持っている刺激に対して無意識に視線を向けることが多い[16]．W3C の Techniques For Accessibility Evaluation And Repair Tools によると明度差 125 以上，色差 500 以上が読みやすい色の組合せであるとされている．ウェブコンテンツ制作においても読みやすいコンテンツとするためには文字色と背景色の明度差や色差によるコントラストを確保する必要があることが知られている[17]．また，槙らの研

表 3　文字列の視覚的特徴量（saliency）
Table 3　Visual saliency of a text.

		文字サイズ		
		20pt 以下	20pt〜30pt	30pt 以上
明度差	124 以下	low	low	medium
	125〜157	low	medium	high
	158 以上	medium	high	high

究では文字と背景の色彩をそれぞれ 40 通りに変化させた 1,600 サンプルの評定結果から配色の明度差が読みやすさに大きく関わっていることを示している[18]．

そこで，5.1 節で得られた「不正単語が視覚的に目立つものが多い」という知見をスコアリングに反映させるために，674 枚の不正表現を含むサンプル画像内の不正な文字列とその他の画像に含まれる文字列のサイズとコントラストに注目し分布を計測した．その結果，画像のサイズに関係なく文字サイズ 30 pt 以上はタイトルや見出し，20 pt 以下の文字サイズは説明文に多く用いられている傾向があった．また，検知したいキーワードはタイトルや強調される部分に多く見られ，サイズが小さい場合でも背景色と文字色のコントラストが高く目立ちやすい色使いをされていることが分かった．これらの知見から，本手法では画像に含まれる検索キーワード t の視覚的特徴量 $saliency(t)$ を，**表 3** に示すように文字色と背景色のコントラスト差と文字サイズが大きくなるほど特徴量が大きくなるように high・medium・low の 3 段階で特徴量を設定した．ただし，明度差の範囲は W3C で定義されている読みやすい明度差 125 と，高本らの研究によって得られた白内障の人にとっての読みやすいと感じられる “おおむね十分の境界” である 158 を基準に決定した[19]．

また，文字サイズは 4 章で述べた文字抽出時に得られた文字画像領域の縦横の大きさを利用し，文字色と背景色は文字画像領域に含まれる文字領域と背景領域に対して代表色選択法[20] を用いて取得した．代表色の選択の手順は，まず文字領域と背景領域の各領域に対し画素値を RGB 色空間から L*a*b* 色空間に変換した後，すべての画素を 1 辺 w の立方体に分割した L*a*b* 色空間に写像し小領域に含まれる画素の数を調べる．その結果，小領域の画素数が周りにある 26 近傍のそれぞれの小領域に含まれる画素数に比べて最も多い小領域を代表色とした．ただし，複数個所が発生する場合はそれらいずれかの領域を代表色とした．また，明度（L）は式 (2) により求め，明度差は文字色と背景色のそれぞれの明度値の差の絶対値とした[21),22]．

$$L = 0.298912R + 0.586611G + 0.114478B \tag{2}$$

5.3　文字特徴量と tf-idf による画像のスコア付け

画像に含まれるキーワードの出現頻度を考慮するために，画像に含まれるキーワードの tf-idf を計算する．tf-idf は文章中の特徴的な単語を抽出するためのアルゴリズムとして知られ，主に情報検索や文章要約などの分野で利用される指標である[23]．tf は文章中の単語の出現頻度であり，idf は多くのドキュメントに出現する語は重要度を下げ，特定のドキュメントにしか出現しない単語の重要度を上げるための逆出現頻度である．本手法ではこの tf-idf の考え方を画像内文字に拡張し，文字列の視覚的特徴量と文字認識信頼度に組み合わせて用いることで画像スコアを計算する．

まず，5.2 節で述べた画像に含まれる検索キーワード t の視覚的特徴量 $saliency(t)$ と文字認識信頼度 $similarity(t)$ から，画像内に含まれる m 番目の文字列 (t, m) の文字特徴量 $termscore(t, m)$ をそれぞれ式 (3) により求める．ただし，重み α で文字認識信頼度と視覚的特徴量の調整をする．

$$termscore(t, m) = (1 - \alpha) \cdot similarity(t, m) + \alpha \cdot saliency(t, m) \tag{3}$$

次に，文字列 t の出現頻度に応じて画像スコアを高くするために，式 (4) により画像内に $tf(t)$ 個含まれる文字列 t のそれぞれの文字特徴量の 2 乗和を求め，検索キーワードによる画像のスコアとする．

$$score(t, image) = \sum_{i=0}^{tf(t)} \{termscore(t, m)\}^2 \tag{4}$$

また，複数の検索キーワードで AND 検索，OR 検索を行う場合の画像スコアは，クエリ $\mathbf{q}$ に含まれる複数の検索キーワード t の画像スコア $score(t, image)$ に $idf(t)$ の値を掛け合わせた数値の総積，総和を式 (5)，(6) により計算することで求めることで容易に拡張可能である．ただし，検索キーワード t の $idf(t)$ は検索対象の総画像数（A）と t を含む画像数（S）を用いて式 (7) により求められ，$idf(t)$ は検索語 t を含む画像が少ないほど大きな値となり希少語であることを示す．

$$score(\mathbf{q}, image) = \prod_{t \in \mathbf{q}} idf(t) \cdot score(t, image) \tag{5}$$

$$score(\mathbf{q}, image) = \sum_{t \in \mathbf{q}} idf(t) \cdot score(t, image) \tag{6}$$

$$idf(t) = \log(A/(S + 1)) + 1 \tag{7}$$

tf-idf の考え方では，文章が長くなるほど検索キーワード t を含む確率が高くなるため，

文章量に応じて tf を調整することが一般的である．そのため，本手法でも画像内に含まれる文章量の指標として，文字認識後の文字列の長さや画像のサイズを用いて重み付けを行うことが望まれる．しかし，本稿で対象としている商品説明画像は複雑な背景やレイアウトを持つため，文字認識時に背景を文字として誤認識するなど認識結果にノイズを含む場合が多く，一概に文字認識後の文字列の長さを画像内の文字量の指標として利用することは難しい．また，画像サイズと画像内に含まれる文字量は一定でないため，$600 \times 10{,}000$ pix の超巨大画像のスコアが非常に低くなる場合や，20×100 pix 程度の小さなバナー画像のスコアが急激に高くなる場合がある．そのため，本稿では画像内に含まれる文章量による重み付けを行わないものとした．

6. 評価実験

6.1 文字候補数による画像内文字検索の精度

文字候補数 N によりどの程度，画像内文字検索の精度が変化するかを評価するために，文字候補数を 1〜60 まで 5 刻みで変化させ，**表 4** に示した不正表現を含む画像で用いられやすい 66 個の検索キーワードを利用して画像内文字検索を行った．

実験ではあらかじめモール管理者が，楽天市場の「医薬品・コンタクト・介護カテゴリ」内で検知した 674 枚の不正表現を含むサンプル画像を用い，4.1 節で述べた手法により画像内文字を認識し，認識結果を得た．ただし，文字カテゴリとして英，数，記号，ひらがな，カタカナ，漢字（JIS 第 1 水準）を含む 3,410 文字を利用し，辞書を作成するために電子商店で多く利用されている「HGS 創英角ポップ体」「HGP 行書体」「MS ゴシック」の 3 つのフォントを利用した．また，表 4 に示した検索キーワードを用いて目視によりカウントした検索キーワードを含む画像数（S）と，文字候補数を変化させ検索した結果得られた正解画像数（T），誤って検知された画像数（E）を比較し，平均の再現率（Recall）と適合率（Precision），F 値を式 (8)，(9)，(10) により求めた（**図 3**）[24]．

$$Recall = T/S \tag{8}$$

$$Precision = T/(T+E) \tag{9}$$

$$F = (2 \cdot Recall \cdot Precision)/(Recall + Precision) \tag{10}$$

図 3 から文字候補数を増やすことで適合率が下がり，再現率が上がる傾向が見られ，本稿で提案した文字認識結果を多重化することで検索漏れを低減することが可能であることが分かる．また，文字候補数が 30 付近で再現率が約 60%で安定し，30 以降では適合率が 90%より低くなり検索結果に誤認識結果が増加するため，本稿で用いた文字認識手法では文字候補を第 30 位まで利用するものとした．

表 4　実験に用いた検索キーワード
Table 4　Search query words used in the experiment.

白肌,爆乳,細胞,発毛,抑毛,巨乳,豊満,便秘,脂肪,除毛,医薬品,花粉症,色白肌,白い肌,若返り,白内障,抗老化,毒出し,クスミ,冷え性,医療用,お通じ,ビタミン,アレルギー,プロ痩せ,抗酸化力,ボケ防止,切りキズ,細胞活性,細胞賦活,シワ防止,解毒作用,便秘緩和,アトピー,不老長寿,抗酸化力,脂肪溶解,ボケ防止,抗酸化作用,脂肪を燃焼,ＤＥＴＯＸ,脂肪が増加,抗ウイルス,代謝アップ,老化を防ぎ,カップＵＰ,メタボリック,細胞を活性化,新陳代謝促進,老化を遅らせ,血液サラサラ,抗アレルギー,バストアップ,セルライト除去,老化と病気防止,基礎代謝アップ,セルライトを軽減,アンチエイジング,メタボリック症候群,活性酸素を取り除く,肌細胞を活性化させ,細胞をイキイキとさせ,バストラインを豊かに,メタボリックシンドローム,バストに脂肪を盛り付ける,抗メタボリックシンドローム

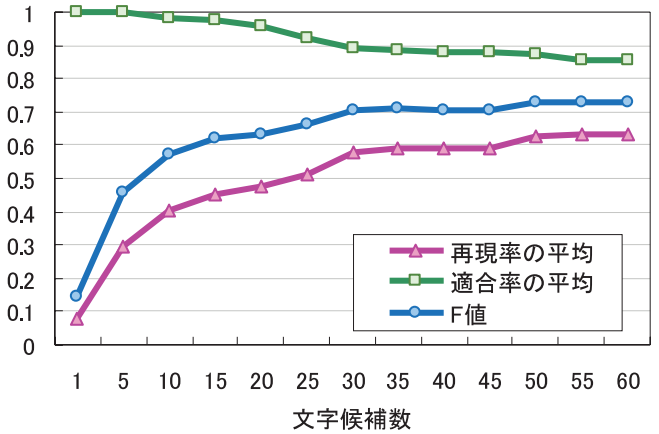

図 3　文字候補数と画像内文字検索の精度
Fig. 3　Recall-precision relation given by the number of recognition candidates.

文字候補数 30 の場合の検索キーワードの長さと検索精度の関係を**表 5** に示す．同表により，検索キーワードが短い場合に検索誤りが生じ適合率が低くなる傾向が認められる．これは文字候補数を多くすることによって，誤認識された文字認識結果を検知する確率が高くなるためであり，検索キーワードの長さに応じて文字候補数を調整することで適合率を高くすることが可能である．さらに，再現率と検索キーワードの長さとの相関は見られなかったが，全体的に再現率が低くなる傾向が見られる．これは，サンプル画像にはアーチ状に配置された文字列や，斜体の文字，サイズが小さい文字といった，文字抽出・認識が困難なケースが多く含まれたためである．そのような高度な文字認識を必要とする画像内文字解析の精度向上は今後の課題である．

表 5　目視結果と検索結果との比較（$N = 30$ の場合）
Table 5　Manual search results vs. the search results of a proposed system.

検索キーワード	目視結果（枚）	検索結果		再現率	適合率	F 値
		正解（枚）	誤り（枚）			
細胞	60	21	0	0.35	1.00	0.52
便秘	11	3	1	0.27	0.75	0.4
脂肪	6	3	1	0.5	0.75	0.6
除毛	6	3	2	0.5	0.6	0.55
若返り	8	4	0	0.50	1.00	0.67
医薬品	23	5	0	0.22	1.00	0.36
花粉症	3	3	0	1.00	1.00	1.00
抗酸化力	8	1	0	0.13	1.00	0.22
ビタミン	9	4	0	0.44	1.00	0.62
アレルギー	6	3	0	0.50	1.00	0.67
抗アレルギー	2	1	0	0.50	1.00	0.67
バストアップ	31	9	0	0.29	1.00	0.45
セルライトを軽減	1	1	0	1.00	1.00	1.00
アンチエイジング	12	5	0	0.41	1.00	0.59
バストに脂肪を盛り付ける	1	1	0	1.00	1.00	1.00
抗メタボリックシンドローム	1	1	0	1.00	1.00	1.00

6.2　文字特徴量に基づく画像スコアの評価

文字認識信頼度と画像内の文字列の視覚的特徴と出現頻度を利用した画像スコアを用いることで，不正である確率が高い視覚的に目立つ文字列を含む画像を効率良く見つけることが可能かどうかを確認するために，サンプル画像としてあらかじめ**図 4** に示した 10 種類の画像を作成し画像スコアを求めた．画像内の文字色は #000000 の「MS ゴシック」のフォントを利用し，① ④ ⑥〜⑩ は文字サイズを 30 pt，③ ⑤ は 20 pt，② は 30 pt と 20 pt の両方を利用し，①〜⑤ ⑦ ⑧ ⑩ は背景色を #FFFFFF，⑥ ⑨ は #666666 とした．また，「絶対痩せる」と「絶対痩せろ」の各文字列の画像の認識結果に対し「絶対痩せる」というキーワードで検索した場合の文字認識信頼度は 4.2 節の表 2 に示したとおり，文字サイズにかかわらずそれぞれ 1.00，0.71 となった．

サンプル画像のスコアを，5.3 節で述べた文字認識信頼度と視覚的特徴量のバランスをとるパラメータ α を 0.0〜1.0 で 0.2 刻みに変化させて計算した結果を**表 6** に示す．ただし，5.2 節で述べた視覚的特徴量 $saliency(t)$ は式 (3) において $saliency(t)$ を 0.0 とした場合，画像内文字の視覚的な特徴をスコアに反映できないため，本稿では low を 0.5，high を 1.0，medium をその中間の値である 0.75 とした．

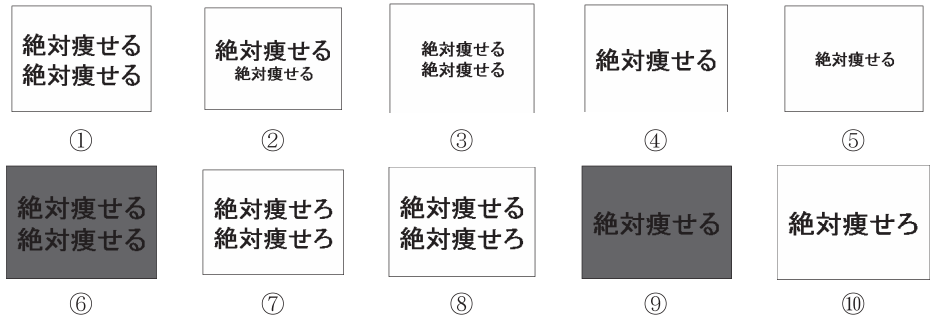

図 4　実験に用いたサンプル画像
Fig. 4　Sample images used in the experiment.

表 6　画像スコアの計算結果
Table 6　The image score results.

No.	similarity	saliency	tf	score					
				α =0.0	α =0.2	α =0.4	α =0.6	α =0.8	α =1.0
①	1.00	1.00	2	2.00	2.00	2.00	2.00	2.00	2.00
	1.00	1.00							
②	1.00	1.00	2	2.00	1.90	1.81	1.72	1.64	1.56
	1.00	0.75							
③	1.00	0.75	2	2.00	1.80	1.62	1.45	1.28	1.13
	1.00	0.75							
④	1.00	1.00	1	1.00	1.00	1.00	1.00	1.00	1.00
⑤	1.00	0.75	1	1.00	0.90	0.81	0.72	0.64	0.56
⑥	1.00	0.75	2	2.00	1.80	1.62	1.45	1.28	1.13
	1.00	0.75							
⑦	0.71	1.00	2	1.01	1.18	1.36	1.56	1.77	2.0
	0.71	1.00							
⑧	1.00	1.00	2	1.50	1.59	1.68	1.78	1.89	2.0
	0.71	1.00							
⑨	1.00	0.75	1	1.00	0.90	0.81	0.72	0.64	0.56
⑩	0.71	1.00	1	0.50	0.59	0.68	0.78	0.89	1.00

まず，α が 0.0 の場合に注目すると，画像のスコアには文字認識信頼度のみが反映されるため，検索誤りが少ない結果の画像を優先的に提示することが可能となる．しかし，文字認識信頼度が同じ ①〜③ と ⑥ が同じスコアになり視覚的に目立たない ⑥ が上位にくる可能性がある．α は文字認識信頼度と視覚的特徴量のバランスをとるパラメータであるから，α の値を高くするほど視覚的特徴量を強く反映することができるが，α が 0.6 以上の場合に検索キーワードを含まない ⑦ と ⑩ のスコアが，検索キーワードを同数含む ⑥ と ⑤ を

それぞれ超えてしまっている．そのため検索結果の上位に検索誤りを含む結果が表示されてしまうことになる．同様に ④ ⑨ ⑩ を比較すると α が 0.0 の場合 ④ と ⑨ が同スコアになり，α が 1.0 の場合 ④ と ⑩ が同スコアになってしまう．

次に，① ⑦ ⑧ について比較すると α が 1.0 以外の場合に画像に含まれる検索キーワードが多いほど画像スコアが高くなっていることが分かる．これらの結果から，α の値を 0.2〜0.4 に設定することで検索キーワードを含まない画像のスコアを低く，また視覚的特徴量に応じて検索結果を良好にソートできていることが確認できる．

このように，文字認識信頼度だけでなく文字列の視覚的特徴量と出現頻度を考慮することで，視覚的に目立つ表現を含む画像のスコアを高くすることが可能であり，効率良く不正な表現を見つけることが期待できる．

6.3　画像内文字検索システム

前章までに述べた理論を実現した画像内文字検索システムを作成し，実際に楽天市場にある画像を対象に，検索時間の評価を行った．本システムは Tomcat 上で動くウェブアプリケーションであり，4.1 節で述べた多重化された認識結果から任意の文字列検索を高速に実現するために Lucene を用いた[25]．Lucene は Apache プロジェクトが管理するオープンソースの全文検索エンジンであり，インデックスを作成する API やインデックスを検索する API を備えた Java ライブラリとして公開されている．本システムではインデックスを作成するため Lucene に実装されている N-gram を用いた単語分割 Analyzer（uni-gram）を用いてインデックス作成をし，第 N 位までの文字認識候補の組合せの中から任意単語の検索を行った．ただし，インデックス作成時に文字色と背景色のコントラストと文字サイズから求められる視覚的特徴量を Field に持たせることで，検索キーワードに応じた画像スコアの計算を行う．開発した画像内文字検索システムの応答性能や検索精度といった実用性の確認を行うために，楽天市場の「ダイエット・健康カテゴリ」,「医薬品・コンタクト・介護カテゴリ」から取得した 567,667 枚の画像を対象に，あらかじめ画像内の文字認識を行った結果得られた認識結果をインデックス化している．**表 7** は測定を行った環境である．

実験では表 4 に示した 66 個の検索キーワードを用いて文字候補数が 1〜30 のインデックスを使用し，文字候補数に対する検索時間を確認した（**図 5**）．同図より，文字候補数に対する検索時間は $O(n)$ で増えている．同時に，標準偏差の値が大きくなることからキーワードの長さによって探索時間のばらつきが生じていることが分かる．また，文字候補数 30 の場合においても平均検索時間が約 350 ミリ秒であり，ストレスを感じさせないという意味で十分実用に耐えうる応答性能を実現できていることが分かる．ただし，平均検索時間は表 4 に示した 66 個のキーワードをクエリとして 10 回検索を行った際の平均の時間である．

表 7　画像内文字検索システムの環境
Table 7　System specification and experimental environment.

CPU	Intel (R) Core(TM)2 Duo CPU E4500:2.2GHz						
メモリサイズ	980MB RAM						
OS	Windows XP Pro						
画像数	567,667 枚						
インデックスサイズ(GB)	N=1	5	10	15	20	25	30
	2.2	2.8	3.6	4.4	5.2	6.0	6.8

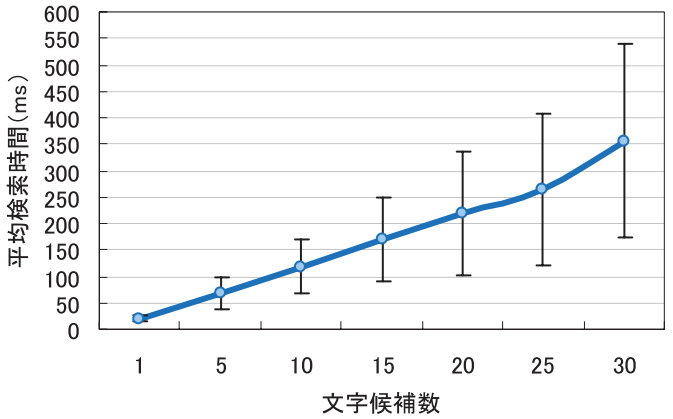

図 5　文字候補数と検索時間の関係
Fig. 5　Average search time given by the number of recognition candidates.

7. ま と め

本稿では文字検索誤りの程度を表す文字認識信頼度，検索キーワードに該当する画像内文字列の視覚的特徴量，画像内キーワード出現頻度を考慮した検索キーワードに基づく画像のスコアリング手法を提案した．本システムを用いて，実際のサイバーモール内の電子商店に存在する約 56 万枚の画像を対象に実験を行い，提案手法によるスコアリングが不正文字検知に有用であり，画像内の文字を検知することで従来のテキスト検索だけでは検知することができなかった商品ページを検知できることが確認できた．

今後の課題として，フォントパターンを判別し画像スコアに反映することや，アーチ上に配置された文字列や背景と文字との分離が困難な文字列といった複雑なレイアウトを持った画像内文字認識の精度向上が求められる．また，テキスト検索結果との併用による画像スコアの算出手法についても検討する必要がある．

謝辞　本研究においてご教授いただきました東京大学米澤明憲教授，増原英彦准教授に深く感謝いたします．また，楽天技術研究所の西岡悠平，増田創，両名には多大なる技術的ご協力をいただきました．ここに感謝の意を表します．

参 考 文 献

1) 楽天株式会社：楽天市場．http://www.rakuten.co.jp/ (参照 2009-11-30)
2) ヤフー株式会社：Yahoo! ショッピング．http://shopping.yahoo.co.jp/ (参照 2009-11-30)
3) Lucas, S.M., Tams, A.C., Cho, S.J., Ryu, S. and Downton, A.C.: Robust word recognition for museum archive card indexing, *Proc. ICDAR2001 6th International Conference on Document Analysis and Recognition*, pp.144–148 (2001).
4) 丸川勝美，藤沢浩道，嶋　好博：認識機能の出力あいまい性を許容した情報検索手法の一検討—認識誤り特性に着目した検索手法の分析評価，信学論 D-II，Vol.79, No.5, pp.785–794 (1996).
5) 富士通株式会社富士通研究所：PRESS RELEASE インターネット上の不正利用画像を検出する技術を開発．http://pr.fujitsu.com/jp/news/2003/11/5-1.html (参照 2009-11-30)
6) グーグル株式会社：Google 画像検索．http://images.google.co.jp/ (参照 2009-11-30)
7) 安田　浩，青木輝勝，長田礼子：コンテンツ ID と電子透かしへの要求条件，電子情報通信学会ソサイエティ大会講演論文集，pp.215–216 (1999).
8) 平野秀幸，小谷誠剛，小野越夫：利便性を重視した著作権保護方式，信学技報，コンピュータセキュリティ，pp.31–36 (2000).
9) 近藤堅司，松川善彦，今川太郎，目片強司：文字認識誤りを含むテキストからのあいまい検索に関する一検討，信学技報，PRMU, 99(305), pp.69–75 (1999).
10) 太田　学，高須淳宏，安達　淳：認識誤りを含む和文テキストにおける全文検索手法，情報処理学会論文誌，Vol.39, No.3, pp.625–635 (1998).
11) Lopresti, D. and Zhou, J.: Retrieval Strategies for Noisy Text, *Proc. Symp. Document Analysis and Information Retrieval*, pp.255–270 (1996).
12) Tseng, Y.-H. and Oard, D.W.: Document Image Retrieval Techniques for Chinese, *Proc. 4th Symposium on Document Image Understanding Technology*, pp.151–158 (2001).
13) 大津展之：判別および最小 2 乗規準に基づく自動しきい値選定法，信学論 D，Vol.63, No.4, pp.349–356 (1980).
14) 芦田和毅，永井弘樹，岡本正行，宮尾秀俊，山本博章：情景画像からの文字抽出，信学論 D，Vol.J88-D2, No.9, pp.1817–1824 (2005).
15) 孫　寧，田原　透，阿曽弘具，木村正行：方向線素特徴量を用いた高精度文字認識，信学論，Vol.J74-D-II, No.3, pp.330–339 (1991).
16) Itti, L., Koch, C. and Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20, No.11, pp.1254–1259 (1998).
17) W3C: Techniques For Accessibility Evaluation And Repair Tools, W3C Working Draft (26 April 2000). http://www.w3.org/TR/2000/WD-AERT-20000426#color-contrast (参照 2009-11-30)
18) 槙　究，田中奈苗，留目真由香：読みやすさと配色の良さの両立—文字色と背景色の組み合わせの評価，日本色彩学会誌，Vol.29, No.1, pp.2–13 (2005).
19) 高本康明，永野行記：Web アクセシビリティ診断ツール：WebInspector，*FUJITSU*, Vol.54, pp.203–207 (2003).
20) 長谷博行，米田政明，酒井　充，丸山　博：カラー文書画像中の文字領域抽出を目的とした色分割についての検討，信学論 D-II，Vol.J83-D-II, No.5, pp.1294–1304 (2000).
21) 山口富士夫（監修）：実践コンピュータグラフィックス，日刊新聞社 (1987).
22) 日本放送協会：カラーテレビ教科書 [上]，日本放送出版会 (1977).
23) Salton, G. and Buckey, C.: Term-weighting approaches in automatic text retrieval, Readings in Information Retrieval (1997).
24) 北　研二，津田和彦：獅々堀正幹：情報検索アルゴリズム，共立出版 (2002).
25) Lucene. http://lucene.apache.org/ (参照 2009-11-30)

(平成 21 年 11 月 30 日受付)
(平成 22 年 6 月 3 日採録)

益子　宗（正会員）

平成 20 年筑波大学大学院システム情報工学研究科博士課程修了．博士（工学）．同年楽天（株）楽天技術研究所入社，現在に至る．平成 14 年 IPA 未踏ソフトウェア創造事業，開発代表者．平成 18〜20 年日本学術振興会特別研究員．平成 22 より情報処理学会グラフィクスと CAD 研究会運営委員．エンタテインメントコンピューティング，CG アニメーション等の研究に従事．ACM，電子情報通信学会，各会員．

星　　鉄矢

平成 18 年東京大学薬学部卒業．アクセンチュア株式会社を経て，ビーグル株式会社設立に至る．現在，代表取締役社長．また，楽天（株）楽天技術研究所にて客員研究員を務める．在学中，ソフトアガー重層法の考案等により抗生物質探索に貢献．著書に共著として，講談社ブルーバックス『パソコンで見る動く分子事典 Windows Vista 対応版』，翔泳社『FLASH OOP（Advanced Web design books）』，『FLASH OOP for ActionScript 3.0』等．

平野　廣美（正会員）

昭和 51 年京都大学工学部航空工学科卒業，日本 CAD，CSK 総合研究所，新日鉄ソリューションズを経て，楽天（株）楽天技術研究所勤務．現在に至る．脳科学をベースとした画像認識，人工知能等の研究に従事．人工知能学会会員．