

カメラを用いた文字認識・文書画像解析の現状と課題

黄瀬 浩一[†] 大町真一郎^{††} 内田 誠一^{†††} 岩村 雅一[†]

[†] 大阪府立大学大学院工学研究科 〒599-8531 大阪府堺市学園町 1-1

^{††} 東北大学大学院工学研究科 〒980-8579 仙台市青葉区荒巻字青葉 6-6-05

^{†††} 九州大学大学院システム情報科学研究院 〒812-8581 福岡市東区箱崎 6-10-1

E-mail: [†]{kise,masa}@cs.osakafu-u.ac.jp, ^{††}machi@aso.ecei.tohoku.ac.jp, ^{†††}uchida@is.kyushu-u.ac.jp

あらまし デジタルカメラの小型化, 高解像度化, 一般化に伴い, これを入力手段として用いた文字認識・文書画像解析への期待が高まりつつある. 本稿では, デジタルカメラを用いた文字認識・文書画像解析について現在の代表的な技術を俯瞰するとともに, バーコードなどの関連技術との比較によって, 文字認識, 文書画像解析の立場や将来への課題を浮き彫りにする. また, 将来に向けた試みの一つとして, 著者らが「バーコードリーダ並みに手軽で高精度な文字認識」を目指して行っている研究「複比を用いた文字への情報埋め込み」の一端についても紹介する.

キーワード カメラベース, 文字認識, 文書画像解析, 情報埋め込み, 複比

Current Status and Future Prospects of Camera-Based Character Recognition and Document Image Analysis

Koichi KISE[†], Shinichiro OMACHI^{††}, Seiichi UCHIDA^{†††}, and Masakazu IWAMURA[†]

[†] Graduate School of Engineering, Osaka Prefecture University Sakai-shi, Osaka, 599-8531 Japan

^{††} Graduate School of Engineering, Tohoku University

6-6-05 Aoba, Aramaki, Aoba-ku, Sendai-shi, 980-8579 Japan

^{†††} Faculty of Information Science and Electrical Engineering, Kyushu University

Hakozaki 6-10-1, Higashi-ku, Fukuoka-shi, 812-8581 Japan

E-mail: [†]{kise,masa}@cs.osakafu-u.ac.jp, ^{††}machi@aso.ecei.tohoku.ac.jp, ^{†††}uchida@is.kyushu-u.ac.jp

Abstract Pervasive use of handy digital cameras with higher resolution is now defining new roles of character recognition and document image analysis as a mean of analyzing camera-captured images. In this report, we survey state-of-the-art of research and technologies of camera based character recognition and document image analysis. We also describe the current position and future prospects of character recognition and document image analysis in comparison with related technologies such as “barcodes”. In addition, we briefly introduce our research entitled “embedding information on characters using cross ratios” whose final goal is to make character recognition as easy and accurate as bar-code reading.

Key words Camera-based, Character recognition, Document image analysis, Embedding information, Cross ratio

1. ま え が き

近年, デジタルカメラやデジタルビデオレコーダは小型化や高解像度化を伴いながら急速に一般化している. 例えば, デジタルカメラの解像度は小型でも 500 万画素程度となっており, 携帯電話付属のカメラでさえ 300 万画素を超えるものもある. このような流れに伴い, 用途は従来のスナップ写真から, 文字で書かれたメモの取得や情報の記録にも拡大しつつある. すなわち, デジタルカメラは, 文字や文書を記録するためのパーソ

ナルな携帯コピー機として活躍しはじめている.

さて, 上記のように取得した画像は, 少数のうちにはよいが, 数が多くなるにつれて必ず整理の問題を引き起こす. 現在は手作業で整理可能な量であっても, 近い将来に破綻をきたすことは想像に難くない. この問題を解決する方策として, 文字認識や文書画像解析への期待が高まっている.

文字認識や文書画像解析への入力手段としてデジタルカメラを捉えると, その携帯性やボタン一つの操作性は魅力的である. これは, スキャナは使いにくいと感じる一般ユーザでも, デジ

タルカメラならば抵抗が少ないと考えられるからである。反面、処理技術の立場から眺めると、画像の認識・解析にはかなりの困難が伴うことも事実である。スキャナという2次元統制環境下で取得した画像と比べ、デジタルカメラの画像は3次元の非統制環境で取得されるものであるため、従来の文字・文書の分野では起こりえなかった様々な問題点が生じてくる。

このような問題点を整理するとともに、現在の技術の中で代表的なものを俯瞰しておくことは、今後の発展に重要となる。このような考えから、本稿では、カメラを用いた文字認識・文書画像解析の現状と将来の課題について著者らなりの考え方をまとめる。紙数の制限から、本稿は、網羅的なサーベイではなく著者らのポジションペーパー的な色合いが濃いものとなっている点を予めお断りしておく。網羅的なサーベイに興味のある読者は、Doermannらによるサーベイ [1] が参考になる。ユビキタスコンピューティングとの関係に興味のある読者は山田らによる解説 [2], [3] から有用な情報が得られる。

加えて本稿では「文字認識をバーコードリーダの精度や手軽さに高める」という目標のもとで著者らが進めている研究の一端も紹介する。これは、複比 [5] という不変量を用いて文字に情報を埋め込むことにより、認識精度を高めようという試みである。なお、埋め込みビット数と認識率やリジェクト率との関係については別稿 [4] で報告する。

以下、2. でまずカメラを用いた文字認識・文書画像解析の主な技術を概観する。次に3. において、バーコードやRFIDなどの関連技術と比較することにより、文字認識・文書画像解析の立場を浮き彫りにする。また、4. で将来への課題をまとめると共に、5. で我々の情報埋め込み法を簡単に紹介する。

2. カメラを用いた文字認識・文書画像解析

2.1 環境に対する前提

カメラを用いた文字認識・文書画像解析が従来のものと大きく異なる要因は、処理対象となる画像を取得する環境が、2次元の統制環境から3次元の非統制環境に移ることによる。そこで、個々の処理について概観する前に、まず、前提がどのようになっているかについてまとめておこう。

違いは大きく、撮像に関する違いと対象に関する違いに分類できる。以下、各々について述べる。

(1) 撮像に関する違い

- 画像に与えられる幾何歪み・・・スキャナで取得した画像においては、画像に与えられる歪みは、回転、併進、拡大縮小といった相似変換の範囲に収まっている。一方、カメラで撮影した画像に対しては、射影変換 [5] までを考慮する必要がある。射影変換が加えられた画像においては、元の対象の平行線がもはや平行ではなくなることから、処理はより困難となる。

- 解像度・・・A4サイズの文書を300dpi, 400dpiという通常の解像度でスキャンした場合、画像の画素数は、各々900万画素、1500万画素となる。一方、カメラ付き携帯電話やデジタルカメラの解像度は、各々数十万～500万画素、300～800万画素程度である。このように単純に画素数を比較しただけでも、カメラの解像度が低いことがわかる。

- 照明・フォーカス・手ぶれ・・・スキャナでは無縁であった不均質な照明、フォーカスのずれ、手ぶれなどが、カメラでは通常のこととして画像に加えられる。

(2) 対象に関する違い

スキャナでは、紙面の文字や文書を撮影することがほとんどであるが、カメラでは紙面に加えてシーン中の文字、物体表面上の文字が撮影対象となる。これにより、オクルージョン、複雑な背景、非平面上の文字・文書など、スキャナでは生じなかった問題が出てくる。

カメラを用いた文字認識・文書画像解析は、上記のような困難への挑戦であると言える。

2.2 画像処理

2.2.1 2値化

スキャナで取得された文書画像において、背景と文字の分離のために、2値化がしばしば採用される [6]。カメラで撮影された文書画像においても同様の目的で採用されることが多い。しかし、光源の非一様性、印字面の湾曲・反射、撮像系の光学的歪み [7] により、背景と文字の濃淡値は一定せず、両者の正確な分離は困難である [2]。さらに文字が様々な彩色されていたり、複雑な背景を持つ場合、問題は更に難しくなる。

2値化を含め、カメラで撮影された文書画像に対する背景・文字分離法は、以下のように整理される。

- 局所2値化 [8], [9]。
- 濃淡値ヒストグラムの複数分割に基づく画像分解 [10]
- 複雑度など文字領域性評価尺度に基づく画像分解 [11]
- 色クラスタリングによる画像分解 [12]
- エッジ連結領域に対する局所的な色クラスタリング [13]
- 局所的2値化・輝度情報・色情報を統合する方法 [14]

以上の手法で分離・抽出されるのはあくまで文字領域の候補に留まる。従って、2.3節で述べられる文字切り出し・認識処理を経て、その文字らしさは再び評価されることが多い。

2値化以外にも、文字・テキストの強調技術として、高解像度化、コントラスト補正、フォーカスのずれ、手ぶれ補正が必要になる場合がある。これらの詳細は文献 [1] にある。

2.2.2 幾何歪みの補正

スキャナによる文書画像に生じる幾何歪みは全体的な傾き(スキュー)程度だが、カメラによる文書画像の幾何歪みは複雑で多様なものとなる。これら幾何歪みは、(i) カメラ側に起因した歪みと、(ii) 文書側に起因した歪みに大別される。それらの補正技術は共に dewarping と称され、様々な検討されている。

カメラに起因した幾何歪みとしては、2.1(1) で述べたように、カメラと文書が正対していない場合に発生する射影変換歪みが挙げられる。文献 [15]～[17] では、文字列方向やテキスト領域の外接矩形等から、元々平行だったと思われる直線状部分を異なる2方向(水平・垂直の2方向)について求め、それらから射影変換歪みを推定・補正している。

一方、文書側に起因する幾何歪みは、2.1(2) で述べたように、文書の非平面性に因るものであり、射影変換のような線形変換単独では表現できないことが多い。この多様性に対処するため、ステレオ測定により文書の表面形状を取得する試みがなされて

いる [18]。他に、カメラで取得された文書中の曲線状部分をトラッキングにより複数検出し、それらを湾曲したテキスト行とみなして直線化することで、文書全体の非線形歪みを補償する方法が提案されている [19]。また、対象を限定して考えることで、より正確に歪み補償する試みもなされている。その例として、見開いた書籍表面を対象とした幾何歪みモデル [20], [21] がある。また、この見開いた書籍画像において、撮像環境が限定できる場合は、Shape-from-shading に基づく手法も利用できる [22]。

2.2.3 複数画像の統合処理

一つの文書を一旦複数の静止画 (フレーム) として断片的に記録し、再度それらを統合する試みが検討されている。これにより、1 フレーム内に収まらないような文書も扱えるようになる。また、この統合処理が実現すれば、カメラを文書に近づけて、ずらしながら撮影できるので、たとえ解像度の低いカメラであっても、カメラで撮影された文書画像の認識・理解に使用できることになる。さらに、複数フレームの重複を利用することで、2 値化の精度向上 [23] や超解像処理 [24] が可能となる。これらと類似して、低解像度画像を複数統合することで、高解像度画像から得られるような文字特徴量を近似合成する方法も提案されている [25]。

複数フレームを統合し、文書全体を含むような大きな画像を作る処理は、モザイクングと呼ばれる。隣接フレームをオーバーラップさせながら重ね合わせることになるが、カメラ位置の変化により、各フレームは異なった座標系で撮像されている。すなわち、各フレーム毎に異なった幾何変換を受けていることになる。従って、統合する際には、フレーム間で対応点を探索し、射影変換 [26] やアフィン変換 [23] を補正しながら互いに位置合せを行う必要がある。この対応点探索については、文書独自の性質を利用する方法も提案されている [26]。また、画像側の幾何変換を直接求めるのではなく、撮像時のカメラ外部パラメータを推定する方法も提案されている [24]。

こうした画像レベルの統合処理とは別に、一旦各フレーム毎にフレーム内の文字列を認識した後、その認識結果を隣接フレーム間で統合するという方法がある [9]。

2.3 文字切り出し・認識

2.3.1 文字切り出し

カメラを入力デバイスとする場合、扱う対象によって文字切り出しの手法は大きく異なる。カメラをスキャナに代わる文書の入力デバイスと考え、机上の文書を固定されたカメラを用いて電子化するような場合 [27] には、2.2 節で述べた画像処理手法により、従来のスキャナを用いた場合と同等程度の文書画像が得られ、従来の文字切り出し手法が適用できる。一方、任意の情景画像から看板や標識等の文字を切り出す場合には、複雑な背景から文字領域をいかに特定するかが大きな課題となる。後者の場合、対象に制約を加えることで精度を上げる立場と、制約をできるだけ緩くして様々な場合に対応する立場がある。

アプリケーションを限定して制約を厳しくした手法としては、例えば自動車のナンバープレートの認識がある。ナンバープレートでは使用される文字のカテゴリやフォント、規格が決

まっているため、文字配置規則を用いた高精度な抽出 [28] が可能である。また、道路案内標識を対象とし、標識の構成規則に基づき矢印の位置と文字領域との対応付けを行う方法も提案されている [29]。

制約をやや緩和した方法としては、文字と背景は明度値や色相が大きく異なる等の条件を用いてまず文字らしい領域を抽出し、文字列や文字自体の性質を利用して文字列領域を絞り込む手法がある。文字列の性質を利用したものとしては、文字列では文字が一直線上に配置されているという仮定を用いる方法 [11], [12] が一般的である。一方、文字の一般的な性質を使用した手法としては、文字線と背景の濃度差を利用した手法 [8]、文字と背景との色対比を利用した手法 [30]、空間周波数の分析と輝度コントラストの評価に基づく手法 [31]、ラブラシアン・ガウシアンフィルタを用いて線状領域を抽出する手法 [32]、文字列領域の統計的な性質を利用する手法 [33]、多重解像度でのエッジと色分布を利用した手法 [17] 等が検討されている。

2.3.2 文字認識

文字領域が特定された後、文字認識を行う際にも、陰影等による光学的変換、歪み、解像度の低さ等による画像劣化が認識精度を下げる要因となる。また、従来の綺麗な文書画像と比較して切り出しの不正確さに起因する誤りも多発する。画像劣化に対応するために、2.2 節で述べた画像処理手法により高精細な (2 値) 画像を得て従来の文字認識を使用する方法 [34] の他、カメラにより取得された文字画像の特性を考慮した種々の認識手法が提案されている。

カメラによって取得された画像の品質の悪さを補うための手法としては、画像の劣化指数を求め、劣化指数に応じて辞書を変化させる手法 [35] や、動画の複数フレームの画像を用いて部分空間を構築する手法 [36] が提案されている。文字画像の非均一性による影響を取り除くための方法としては、輝度レベルの正規化を行った後ガボールウェーブレット変換により特徴を抽出する手法 [17] がある。切り出しの不正確さに対応するために、文字の部分画像を含む様々な変形のパターンを登録しておく手法 [37] が検討されている。また、複数フレームを用いて認識精度を向上させる方法として、認識結果を文字列レベルで統合する手法 [9] や、モザイクングと認識を DP で同時最適化する手法 [38] 等が提案されている。

さらに、看板文字等の情景画像中の文字は、人の目を引き付けたりデザイン性を高める等の理由で、既成のフォントにとられない様々な飾りが施されている場合がある。このような飾り文字を認識する試みとして、複数パラメータでばかしを加えた画像を用いて文字の本質的な構造を抽出する手法 [39] が提案されている。

2.4 文書画像解析

最後に、より大きな単位の解析として、文書のレイアウト構造や論理構造を解析する処理 (以下、文書画像解析と呼ぶ) を考えてみよう。上記の画像処理や文字切り出し・認識の研究事例に比べて、カメラで撮像された文書画像解析に相当する研究事例は少ない。理由としては、現在の研究の中心が画像処理や文字切り出し・認識にあることに加え、現状では解像度が不足

していることが挙げられるであろう。更に、通常の文書を対象とする場合、dewarping が成功すれば従来技術が適用可能であることも遠因と考えられる。実際、文字や文字列ではなく、段落やページといった単位の解析を行っている場合でも、処理目的は dewarping であることが多い [15], [16]。

では、将来にわたっても、カメラで得た画像に対して、“文書レベル”の解析は不要なのであろうか。我々は次の3つの可能性を挙げたい。まず一つ目は、カメラの性能向上によって、解像度不足が解消されることである。カメラによる画像取得は容易であるため、スキャナに近い解像度が得られれば、文書全体が撮影される機会も増えると考えられる。次に、文書画像検索 [40] など、情報へアクセスするキーとしての利用がある。文書が持つ構造を利用することによって、単語レベルを超えた効果的なアクセスが可能であると考えられる。最後に、看板や標識など、従来の紙面とは異なるが独自の構造を持つものの解析である。これらの対象から意味のある情報を取得するためには、文字が読めるだけでは十分とは言えず、それらの間に成り立つ構造を解析によって取得する必要がある [29]。

3. 関連技術

カメラを用いた文字認識・文書画像解析の目標は、他の関連技術によっても実現可能と考えられるかもしれない。ここでは、このような関連技術として、バーコード [41], [42], RFID (Radio Frequency Identification) [43], 電子透かし (例えば [44]) の3技術を取り上げ、比較検討する。

表1に概要を示す。ここで、情報表現と情報密度は、データを表現する方法とビットあたりの占有面積を表す。情報取得とは、取得すべき情報をユーザが選択可能か一括入力か、更には同時に複数を読み取ることが可能かどうかを表す。外観は、情報表現によって外観に影響があるかどうかを意味する。コストとは情報を記録する際のコスト、データ量とは記録できるデータの量、機械と人間の可読性とは、機械や人間にとってデータが読み取りやすいかどうかを表す。他と比べて特長と思われる箇所は太字で、問題点と考えられる箇所は下線で表している。以下、表に従ってバーコードから順に述べる。

バーコードは、機械にデータを読み取らせる手段として最も普及しているものである。バーコードの特徴としては、外観に影響を及ぼすものの^(注1)、印刷だけで実現可能であるためコストが低い点が挙げられる。バーコード利用の中心は、ID用データや情報コンテンツへのポイントの表現であるが、2次元バーコードではコンテンツを記録可能な数キロバイトといった大量のデータを表現することもできる。可読性という点で見ると、機械による可読性は高いが、人間には読み取れない。また、情報取得は一括であって一部を選択的に読むことはできない。加えて、読み取りは逐次的であって、複数のバーコードがある場合には基本的には数だけ読み取りを繰り返す必要がある。

RFIDはコピキタスコンピューティングなどの概念とともに注目を集める技術であり、バーコードの代替として捉えられる

ことも多い。バーコードとの違いは、コストが高いものの、同時に複数のデータを読み取ることが可能な点にある。

電子透かしは、写真などの領域にデータを埋め込む技術である。印刷の場合には数バイト程度のデータ量となるが、外観をほぼ損ねずに情報を表現可能である。ただし、埋め込みの存在は人間には知覚できず、どこを読めば情報が取得できるのかが分からないため、読み取りに問題が生じることも考えられる。

以上と比較して、文字による情報表現の最も大きな特徴は、人間に可読であるという点 [2], [3] であろう。他の情報表現形態では、どのような情報が取得されたのか、ユーザである人間が直接知ることはできないが、文字は人間にも読めるため、これが可能である。すなわち「モノに情報を貼り付ける」だけではなく、「貼り付けた情報を知る」ことが可能である。この特徴によって、取得したい情報をユーザが選択することも実現できる。さらに文字は「人間が読むもの」なので外観を悪化させないという利点もある。その反面、文字の機械可読性は他と比較して低いといわざるを得ない。機械による文字の利用が十分促進されていない理由は、この点にあると考えられる。

4. 将来に向けての課題

我々が目指すべき目標は、人間と機械の双方に可読性の高い情報表現形態である。文字を用いてこれを実現するためには、機械に対する可読性を向上させなければならない。では、具体的にどのような方策が可能であろうか。2.で述べたトピックの探求が重要であることは言うまでもないが、従来のスキャナを用いた文字認識・文書画像解析の発展過程を省みれば、これに加えて、少なくとも次のような可能性があると考えられる。

(1) 劣化モデルの検討

文字画像や文書画像の劣化モデル (defect models, degradation models) は、スキャナを用いた文字認識や文書画像解析の開発、信頼性向上に有効である [45], [46]。同様のことは、カメラを用いた場合にも当てはまると考えられる。実際、事例はまだ多くないものの、試みは報告されている [35], [47], [48]。

(2) OCR フォント再考

文字認識が現在よりかなり困難であった研究初期のころ、先人達はなんとか「認識」を実現するため、OCR フォントという考え方を導入した。これは、機械への可読性を高めるため、認識対象である文字に手を加えようとするものである。カメラを用いた文字認識が簡単ではない現在、同様の考え方が成り立つのではないだろうか。

5. 複比を用いた文字への情報埋め込み

バーコード並みの認識率を備えた文字認識が可能になれば何かが変わるか。また、それはどう実現されるものか。ここでは、筆者らのこのような素朴な疑問を契機として開始された研究事例を紹介する。本研究の目標は、バーコードリーダでバーコードを読み取るような手軽さ・精度により、カメラを用いて文字を認識することである。見方を変えると、人間にも読めるバーコードを目指すことも言える。これは、3.で述べた「機械可読性向上」を実現する一形態であり、4.で議論した「カメラを

(注1): DataGlyph [42] のように外観への影響を極力排除したコードもある。

表 1 関連技術との比較

	情報表現	情報密度	情報取得	外観	コスト	データ量	機械可読性	人間可読性
バーコード	専用領域	低～中	一括, 逐次	影響あり	低	多	高	無
RFID	チップ	高	一括, 同時	影響なし	高	多	高	無
電子透かし	写真などへの埋め込み	低以下	一括, 逐次	影響少	低	少	中	無
文字・文書	文字	低	選択的, 逐次	影響なし	低	多	低	有

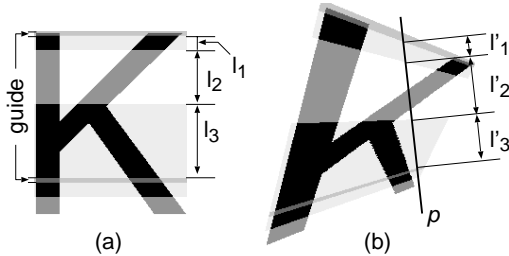


図 1 (a) 複比パターンを埋め込んだ文字画像および (b) その射影変換 ($\delta = 32$)

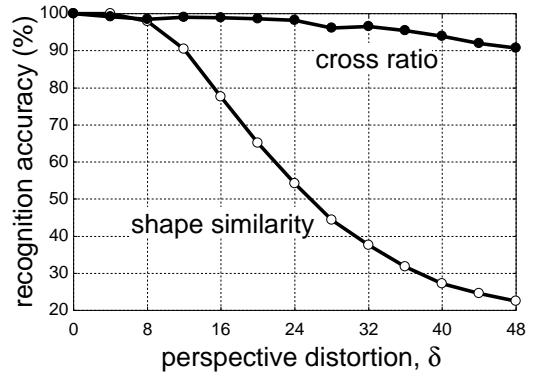


図 2 文字形状ならびに複比による文字認識率

用いた場合の OCR フォント」とも捉えられるものである。

5.1 方 式

各カテゴリのフォントに対し、ある複比を持つパターンを埋め込むことを考える。埋め込みの形式については様々考えられるが、ここでは背景も含め文字全体に 4 本の平行縞を 1 セットとした模様 (複比パターン) を埋め込むことを考える。図 1(a) はその例であり、同図 (b) はそれが射影変換を受けた結果である。同図では表現上、複比パターンが目立つようにしている。複比パターンの最初の縞はガイドである。残る 3 本で複比を表す。それぞれ縞の幅を l_1, l_2, l_3 とすると、このパターンによる複比は、 $r = (l_1 + l_2)(l_2 + l_3) / (l_2(l_1 + l_2 + l_3))$ で表される。縞の幅 l_1, l_2, l_3 を変えることで、埋め込む複比の値を制御できる。なお、縞の数を増やすことで複数の複比を連続的に埋め込むこともできるが [49]、ここでは予備的な検討として、以上のように一つの複比を埋め込むこととした。

この複比パターンから複比を抽出する際は、それを横断するような直線を引き (図 1(b) 中の直線 p)、その線上での複比パターンの間隔 (同図 l'_1, l'_2, l'_3) を求め、それらの複比を計算すればよい。理論上、どのように線を引いても、また射影変換を受けても、同じ複比 r が求まる。従って、この複比パターンを埋め込んだ文字画像をいかなる角度から撮影しても、量子化の影響がなければ常に同じ複比が取り出せることとなる。

複比が担う情報についても様々考えられる！「人間が読むもの」と「機械が読むもの」を同一化した情報表現のためには、各文字にそのカテゴリを表す複比を埋め込めばよい。次節ではこの実証実験を行なう。他に、文字列に対し、その内容とは別の情報を埋め込むことも考えられる。例えば、名前に自分の URL を埋め込むといった応用が考えられる。

5.2 実験結果と考察

複比を埋め込んだ英語大文字フォント (Arial) 26 画像を対象として、文字認識実験を行なった。各フォントのサイズは 200×200 画素である。本実験ではカテゴリ毎に異なる複比を埋め込むこととした。ガイド縞の幅および l_1 を 15 画素、 l_2 と

l_3 については $l_2 + l_3 = 135$, $15 \leq l_2$, $15 \leq l_3$ を満たす範囲で可変にした。すなわち本実験では l_2 と l_3 の境界位置を変えることで、26 種類の複比を表現する。境界位置については、文献 [49] のような理論的解析が必要と思われるが、現時点では単純に、境界として取りうる区間を等間隔に 24 分割し、これら分割点および区間の両端点計 26 点を境界位置とした。

各フォント画像について、その 4 隅の x, y 座標をそれぞれ $\pm \delta$ 画素だけ変位させて射影変換を施し、合計 256 個のテストパターンを生成した。前出の図 1(b) もその一例である。各カテゴリの標準パターンとしては、射影変換を施していない元々のフォント画像を用いた。

まず、文字形状を用いた単純類似度による認識実験を行なった。単純類似度は、外接矩形の線形正規化を行なった後、各画素の特徴量を 1 (文字線), 0 (背景) として計算した。図 2 に射影変換の程度 δ による認識率の変化を示す。このように単純に 2 画像を重ね合わせて類似度を評価した場合、 δ が少しでも大きくなると認識率は急激に低下することがわかる。従って、カメラで撮影された文書について文字形状を利用した認識を行なう場合、幾何歪みに関する劣化モデルの構築、およびそれに基づく劣化補償が重要と言える。

次に、テストパターンから複比を抽出し、それが埋め込んだ 26 種の複比のいずれに近いかによって識別を行なった。結果を図 2 に示す。相当程度の射影変換を受けた場合でも、非常に安定して識別できていることがわかる。このように、複比を埋め込んだフォントを利用することで、文字形状を用いなくてもこの程度の認識精度が得られることがわかる。ただし、今回の実験はあくまで計算機内でのシミュレーションであり、実際のカメラ画像の場合の精度は今後検討する必要がある。また著者らは、文字形状による認識結果を併用すれば、異なる複比の数 (埋め込む情報の量) をカテゴリ数以下に抑えられると考えており [4]、この点においても実験的な検証が必要と考えている。

6. むすび

デジタルカメラの発展は、文字認識・文書画像解析に新しい課題を提示している。本稿では、最近の成果を中心に、この分野の現状と課題について、我々の考えを織り交ぜつつ述べた。また、文字の機械可読性を高めることが重要との観点から、複比を用いた情報め込みの研究事例を紹介した。今後、文字・文書の分野がより活発化することを願って、稿を閉じたい。

文 献

- [1] D. Doermann, J. Liang and H. Li: “Progress in camera-based document image analysis”, Proc. ICDAR’03, pp. 606–616 (2003).
- [2] 山田: “コピキタス情報インタフェースのための文字・文書メディア認識・理解”, 信学技報, PRMU2003-229 (2004).
- [3] 山田, 仙田: “携帯カメラを用いたコピキタス情報インタフェース”, 情報処理, **45**, 9, pp. 923–927 (2004).
- [4] 岩村, 内田, 大町, 黄瀬: “認識が与える情報量 ~ 認識率 100%まであと何ビット? ~”, 信学技報 (2005.3) [発表予定].
- [5] 佐藤: コンピュータビジョン, コロナ社, 東京 (1999).
- [6] O. D. Trier and T. Taxt: “Evaluation of binarization methods for document images”, IEEE Trans. PAMI, **17**, 3, pp. 312–315 (1995).
- [7] 新村, 白木, 武川, 澤木, 萩田: “デジタルカメラのズーム制御を利用した文字認識”, 信学論 (D-II), **J87-D-II**, 2, pp. 761–766 (2004).
- [8] 大谷, 塩: “情景画像からの文字パターンの抽出と認識”, 信学論 (D), **J71-D**, 6, pp. 1037–1047 (1988).
- [9] 仙田, 西山, 旭: “携帯カメラによる日本語文字認識の手法と実現”, 信学技報, PRMU2004-124 (2004).
- [10] S. Messelodi and C. M. Modena: “Automatic identification and skew estimation of text lines in real scene images”, Pattern Recognition, **32**, pp. 791–810 (1999).
- [11] 松尾, 上田, 梅田: “適応しきい値法を用いた情景画像からの看板文字列領域抽出”, 信学論 (D-II), **J80-D-II**, 6, pp. 1617–1626 (1997).
- [12] K. Wang and J. A. Kangas: “Character location in scene images from digital camera”, Pattern Recognition, **36**, 10, pp. 2287–2299 (2003).
- [13] 芦田, 永井, 岡本, 宮尾, 山本: “情景画像からの文字抽出”, 信学技報, PRMU2003-222 (2004).
- [14] 平山, 大町, 阿曾: “カラー情報を利用した情景画像中の文字列の高精度抽出”, 信学技報 (2005.3) [発表予定].
- [15] M. Pilu: “Extraction of illusory linear clues in perspectively skewed documents”, Proc. CVPR’01 (2001).
- [16] P. Clark and M. Mirmehdi: “Recognising text in real scenes”, IJDAR, **4**, pp. 243–257 (2002).
- [17] X. Chen, J. Yang and A. Waibel: “Automatic detection and recognition of signs from natural scenes”, IEEE Trans. Image Processing, **13**, 1, pp. 87–99 (2004).
- [18] M. S. Brown and W. B. Seales: “Image restoration of arbitrarily warped documents”, IEEE Trans. PAMI, **26**, 10, pp. 1295–1306 (2004).
- [19] C. Wu and G. Agam: “Document image de-warping for text/graphics recognition”, Lecture Notes in Computer Science (Joint IAPR International Workshops SSPR 2002 and SPR 2002), Vol. 2396, pp. 348–357 (2002).
- [20] H. Cao, X. Ding and C. Liu: “Rectifying the bound document image captured by the camera: a model based approach”, Proc. ICDAR’03, pp. 71–75 (2003).
- [21] T. Kanungo, R. M. Haralick and I. Phillips: “Global and local document degradation models”, Proc. ICDAR’93, pp. 730–734 (1993).
- [22] 和田, 浮田, 松山: “イメージスキャナを用いた書籍表面の 3 次元形状復元 (i) 近接照明下での shape from shading”, 信学論 (D-II), **J77-D-II**, 6, pp. 1059–1067 (1994).
- [23] H. Li and D. Doermann: “Text enhancement in digital video using multiple frame integration”, Proc. ACM Multimedia, pp. 19–22 (1999).
- [24] 池谷, 中島, 佐藤, 池田, 神原, 横矢, 山田: “紙面を対象としたカメラパス推定によるビデオモザイクおよび超解像”, 信学技報, PRMU2003-223 (2004).
- [25] 小佐井, 星野, 岡本, 加藤, 山本: “低解像度画像からの文字認識手法について”, 信学技報, PRMU97-221 (1998).
- [26] A. Zappalà, A. Gee and M. Taylor: “Document mosaicing”, Image and Vision Computing, **17**, 8, pp. 589–595 (1999).
- [27] K. Yamada, K. Ishikawa and N. Nakajima: “A method of analyzing the handling of paper documents in motion images”, Proc. ICPR’00, Vol. 4, pp. 413–416 (2000).
- [28] 田邊, 川島, 丸林, 仲西, 塩, 大塚: “部分文字列の配置規則を考慮したナンバープレート領域抽出”, 信学論 (D-II), **J81-D-II**, 10, pp. 2280–2287 (1998).
- [29] 加藤, 小林, 長谷, 酒井, 米田: “道路案内標識中の矢印と文字領域の対応付けに関する実験的検討”, 画像電子学会誌, **31**, 4, pp. 542–552 (2002).
- [30] 周, 渡辺, 河岡: “情景画像からの看板領域と看板中の情報抽出 — 安定背景領域の推定 —”, 信学技報, PRMU2002-217 (2003).
- [31] 劉, 山村, 大西, 杉江: “シーン内の文字列領域の抽出について”, 信学論 (D-II), **J81-D-II**, 4, pp. 641–650 (1998).
- [32] 志久, 中村: “Log フィルタを用いた情景画像からの文字線検出”, 信学論 (D-II), **J87-D-II**, 8, pp. 1735–1739 (2004).
- [33] X. Chen and A. L. Yuille: “Detecting and reading text in natural scenes”, Proc. CVPR’04 (2004).
- [34] 森, 澤木: “低品質文字の認識手法とその応用に関するサーベイ”, 信学技報, PRMU2001-275 (2002).
- [35] 澤木, 村瀬, 萩田: “劣化推定に基づいた辞書の自動選択による本棚画像中の文字認識”, 映像情報メディア学会誌, **54**, 6, pp. 881–886 (2000).
- [36] 柳詰, 目加田, 井手, 村瀬: “携帯カメラによる動画画像を用いた低解像度文字の認識手法”, MIRU2004, **1**, pp. 321–324 (2004).
- [37] Y. Kusachi, A. Suzuki, N. Ito and K. Arakawa: “Kanji recognition in scene images without detection of text fields-robust against variation of viewpoint, contrast, and background texture -”, Proc. ICPR’04 (2004).
- [38] 宮崎, 内田, 迫江: “動画画像中のテキスト認識を目的としたビデオモザイク手法”, 信学技報 (2005.3) [発表予定].
- [39] S. Omachi, M. Inoue and H. Aso: “Structure extraction from decorated characters using multiscale images”, IEEE Trans. PAMI, **23**, 3, pp. 315–322 (2001).
- [40] 中居, 黄瀬, 岩村, 松本: “複比とハッシュに基づく文書画像検索”, 信学技報 (2005.3) [発表予定].
- [41] < <http://www.adams1.com/pub/russadam/stack.html> >.
- [42] D. L. Hecht: “Printed embedded data graphical user interfaces”, IEEE Computer, **34**, 3, pp. 47–55 (2001).
- [43] 増井: “スマートタグ連載開始にあたって”, 情報処理, **45**, 1 (2005).
- [44] 中村, 片山, 山室, 曾根原: “カメラ付携帯電話機を用いたアナログ画像からの高速電子透かし検出方式”, 信学論 (D-II), **J87-D-II**, 12, pp. 2145–2155 (2004).
- [45] H. S. Baird: “The state of the art of document image degradation modeling”, Proc. 4th IAPR International Workshop on Document Analysis Systems, pp. 1–16 (2000).
- [46] Y. Li, D. Lopresti, G. Nagy and A. Tomkins: “Validation of image defect models for optical character recognition”, IEEE Trans. PAMI, **18**, 2, pp. 99–108 (1996).
- [47] 石田, 柳詰, 目加田, 井手, 村瀬: “部分空間法による低解像度文字認識のための生成型学習法”, 信学技報, PRMU2004-7 (2004).
- [48] J. Sun, Y. Hotta, Y. Katsuyama and S. Naoi: “Low resolution character recognition by dual eigenspace and synthetic degraded patterns”, Proc. 1st ACM workshop on Hardcopy document processing, pp. 15–22 (2004).
- [49] 金澤, 松永, 金谷: “複比による識別のための最良マーカーパターンの設計: 理論解析”, 情報研報, 99-CVIM-115-13 (1999).