

ピントのずれに頑健な文字認識手法の提案

伊村 政志[†] 岩村 雅一[†] 日浦 慎作^{††} 黄瀬 浩一[†]

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

^{††} 広島市立大学 大学院情報科学研究科 〒 731-3194 広島市安佐南区大塚東 3 丁目 4 番 1 号

E-mail: [†]imura@m.cs.osakafu-u.ac.jp, [†]{masa,kise}@cs.osakafu-u.ac.jp, ^{††}hiura@hiroshima-cu.ac.jp

あらまし 文字認識技術には、ピントのずれによる認識率低下が問題として存在する。この問題の原因は、文字の特徴が潰れ、パターンが類似することにある。そのため本研究では、符号化開口を用いることでピントがずれても文字がその特徴を残すようにし、ピントが合わない場合の文字認識を可能にする。しかし、ぼけの大きさはレンズの焦点面からのずれにより変化するため、認識には距離情報を既知にする必要がある。そこでパターン認識に特化した制約として、画像に写っている文字がテンプレートのいずれかであるという条件の下で文字の認識を目指す。ぼけ画像はピントのあった画像とぼけ関数の畳み込み積分で表現できる。前述の制約を用いれば、画像がぼけていない場合にはテンプレートのいずれかが存在するということになり、ぼけ画像とテンプレートの逆畳み込み積分により、距離を推定できると考えられる。しかし、実際は折り返し誤差などの信号処理問題やノイズの影響により、正しく距離を推定できない。本論文では逆畳み込みに代わる手法として、(1) テンプレートを走査させて相関係数を求めることで距離を推定し、それを基にテンプレートをぼかすことで画像との比較から文字を認識する手法と、(2) 予めぼけた文字のパターンを複数用意して認識する手法を提案する。

キーワード 文字認識, 符号化開口, コンピュータショナルフォトグラフィ, テンプレートマッチング, 逆畳み込み積分

1. はじめに

近年、撮影画像中の文字を認識する技術が発展している。この技術はカメラ付き携帯電話やスマートフォンのアプリケーションとして一般にも普及し始めており、情景画像中の文字翻訳など様々なサービスを提供している。しかし、文字を認識するシステムはぼけを始めとした撮影画像の劣化に脆弱である。通常、カメラに取り付けられている絞りで、ピントがずれると文字の特徴が潰れ、その形状が似通ってしまい認識精度を著しく低下させてしまう。

文字のぼけに対しては、オートフォーカスなど、撮影の段階でピントがずれないように撮影システムを構築するという方法がある。しかし、認識対象の前にガラスなどの遮蔽物が存在する場合、そちらにピントが合ってしまう、認識対象に必ずピントが合うというわけではない。そこで本研究では、画像に計算処理を加えることで鮮明な画像を得るコンピュータショナルフォトグラフィ [1] の技術の 1 つである符号化開口 [2] を用いることで解決を図る。これは絞りの形状を設計しぼけの形状を変化させることで、ぼけが発生しても計算処理により鮮明な画像を得る技術である。この技術を用いると、ピントがずれても画像には文字の特徴が残るため、文字認識に活用できると考えられる。

しかし、画像のぼけはレンズの焦点距離からのずれにより変化する。ぼけた画像を通常の文字認識手法で認識

するために、入力画像がどれだけぼけているか、その指標となる距離情報が必要になる。距離情報を得るための既存手法として、ステレオカメラなど特殊なデバイスを用いることで、一度に入手した複数枚の画像を用いる手法が存在する [3]。一方で、河本らは 1 枚の画像から距離情報を得る手法として、対象をバーコード画像に限定し、画像は白と黒の 2 値で表現されているという制約の下でヒストグラムを用い、その分離度から距離を推定した [4]。このように、1 枚の画像から距離情報を得るには、元画像に対して強い制約を設ける必要がある。文字認識では、認識対象の背景色が未知であり、また 2 値で表現されてるとも限らないため、上記のような色による制約を用いることはできない。そのため本研究ではパターン認識に特化した制約として、画像に写っている文字はデータベースに存在するテンプレートのいずれかであるとする。

ぼけ画像は、ピントのあった画像とぼけ関数による畳み込み積分により表現可能なため、上記の制約を設けた場合、テンプレートとの逆畳み込み積分により得られる画像から、距離情報を計算で求めることができると考えられる。しかし逆畳み込み積分は画像をフーリエ変換し、除算することにより求めるため、実際は折り返し雑音などの信号処理問題や、画像中のノイズにより正しく求めることができない。本論文では逆畳み込み積分に代わり、テンプレートを走査させて相関を求めることで距離を推定して文字を認識する手法を提案するとともに、予めぼ

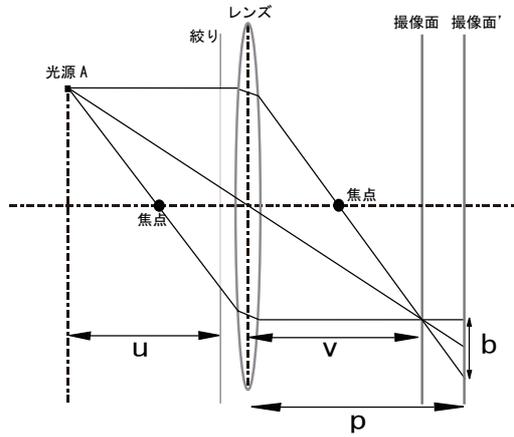


図1 レンズカメラのモデル

けた文字のパターンを複数用意してマッチングにより認識する手法を提案する。

2. 符号化開口

2.1 ぼけ発生と除去

我々が普段使用する、一般的なレンズカメラをモデル化したものを図1に示す。焦点距離が f であるレンズによりカメラから u 離れた光源 A を撮影する場合、ピントの合う距離 v は次式により求まる。

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \quad (1)$$

このとき、レンズから撮像面までの距離が v と一致するとピントのあった画像を得ることができるが、異なる場合にはレンズを通った A からの光は次式で表される直径 b の円として撮像面上に投影される。

$$b = \frac{a}{v} |v - p| \quad (2)$$

ここで、 p はレンズから撮像面までの距離を、 a はカメラの絞りの大きさを表している。このとき、 b の大きさが撮像素子1つ当たりの大きさを超えると画像にぼけが生じる。その形状は絞りの形に依存しており、一般的なカメラの場合絞りの形は円であるため、ぼけの形状も円となる。これを円形開口と呼ぶ。

このようなぼけはレンズからの撮像面までの距離により、その大きさが変化する。この特徴を考えると、ぼけ画像はある関数に則った劣化画像であると言える。この関数をぼけ関数 (PSF) と呼び、撮影されたぼけ画像を t 、ぼけ関数を f 、元画像を s 、画像に含まれるノイズを n と表すと、

$$t(x, y) = f(x, y) * s(x, y) + n(x, y) \quad (3)$$

と畳み込み積分の式として表現することが可能である。式 (3) から、ぼけを除去した画像 \hat{s} はぼけ画像とぼけ関数の逆畳み込み積分により求まることがわかる。畳み込み積分の式はフーリエ変換すると積の式に変化するので、

式 (3) をフーリエ変換すると、

$$T(u, v) = F(u, v) \cdot S(u, v) + N(u, v) \quad (4)$$

となり、式 (4) からぼけ除去画像のフーリエ変換 \hat{S} は、

$$\hat{S}(u, v) = \frac{T(u, v)}{F(u, v)} = S(u, v) + \frac{N(u, v)}{F(u, v)} \quad (5)$$

として表すことができる。以上より、 \hat{S} をフーリエ逆変換することによりぼけを除去した画像 \hat{s} を得ることができる。

ピントのずれによるぼけ関数は開口形状と相似の関係を持つため、 f を式に直した場合以下のように表すことができる。

$$f(x, y) = \frac{1}{M^2} a\left(\frac{x}{M}, \frac{y}{M}\right) \quad (6)$$

ここで、 $a(x, y)$ は開口形状を、 M は $a(x, y)$ に対する $f(x, y)$ の倍率を指す。なお、レンズの透過光量は一定であるため、 $\frac{1}{M^2}$ はそのための正規化項として表されている。もし開口形状が円である場合、 F は0または非常に小さい値を多く含んでしまうため、式 (5) 右辺においてノイズが強調されてしまい、 \hat{S} の値が不安定なものになってしまい、結果としてぼけを除去した画像 \hat{s} もノイズが多くなってしまう。

2.2 符号化開口

2.1 節で述べたとおり、円形開口では逆畳み込みの精度が不安定なため、ぼけの除去により鮮明な画像を安定して得られるような開口形状を設計することが必要となった。これが符号化開口と呼ばれるものである。符号化開口には、MURA コード [5]、Veeraraghavan らの手法 [6]、Levin らの手法 [7] を始め、何を撮影するか等、目的毎に様々な形状が存在しており、これをカメラの絞りに取り付けることにより、撮影された画像に発生するぼけを変化させる。その中でも、Veeraraghavan の開口形状は可視光を撮像する一般的なレンズカメラのための絞りとして設計されている。

しかし、焦点距離からのずれにより式 (6) の M は変化するため、ぼけ関数 f が未知であるという問題が残る。距離を推定しなければ画像復元のための式 (5) も用いることができない。武田らは焦点距離の異なるレンズによるステレオカメラを用い、複数枚の画像を同時に入手することで、距離情報を得た [3]。しかし、ステレオカメラのような特殊なデバイスを用いて距離を推定する一方で、1枚の画像から距離を求めることは容易ではない。河本らは被写体をバーコードに限定し、背景色とインク色が2値であるという制約の下で輝度ヒストグラムの分離度から距離情報を求めた [4]。このように、1枚の画像から距離を推定する場合は、原画像に対し強い前提条件を用意する必要が出てくる。本研究の目的はパターン認識である。扱う画像が2値であるとは限らず、その色も未知であるため、本研究では色情報に代わる別の制約を設ける必要がある。

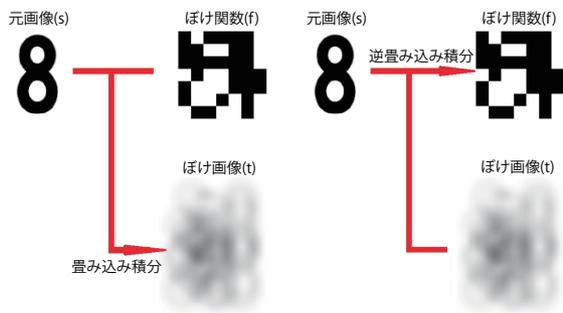


図 2 ぼけ発生の仕組み 図 3 逆畳み込み積分による処理手順

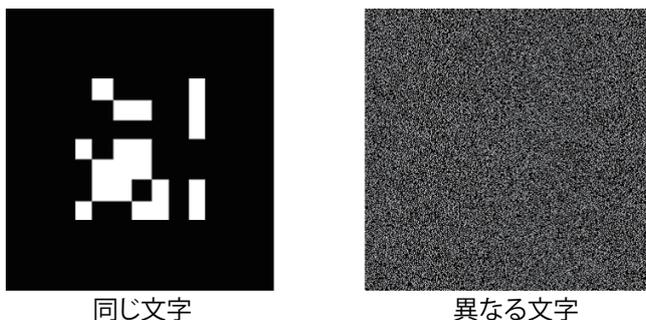


図 4 テンプレートの逆畳み込み積分で得られる画像 (理想)

3. 提案手法

本節ではぼけに頑健な文字認識手法のための制約、その制約を用いた計算処理による距離推定手法と問題点、そして提案手法について説明する。

3.1 本研究で用いる制約

2.2 節で述べたとおり、パターン認識を目的に画像から距離を推定するためには、色情報を用いない別の制約を設ける必要がある。文字認識はそもそも認識対象とデータベースに存在するテンプレートとのマッチングにより行われている。そこで本研究では、写っている文字がデータベースに存在するテンプレートのいずれかであるという制約を設ける。

3.2 計算による距離の推定

上記制約を設けると、式 (3) の元画像 s がテンプレートのいずれかとなる。ぼけ画像 t は撮影で得られた画像であり、図 2 のように畳み込み積分で表現できるので、ぼけ関数 f は図 3 のように、逆畳み込み積分で求まると考えられる。式 (3) をフーリエ変換し、式変形することで、

$$\hat{F}(u, v) = \frac{T(u, v)}{S(u, v)} = F(u, v) + \frac{N(u, v)}{S(u, v)} \quad (7)$$

となる。 s がぼけ画像に写っている文字と同じ字種のテンプレートである場合、 \hat{F} をフーリエ逆変換した \hat{f} は開口形状と同じ形であると考えられる。一方で s がぼけ画像に写っている文字とは異なる場合、 \hat{f} の形も開口形状

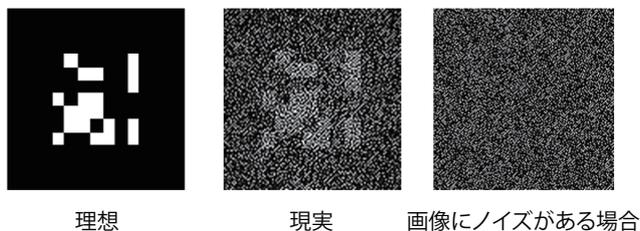


図 5 逆畳み込み積分により得られる画像の理想と現実

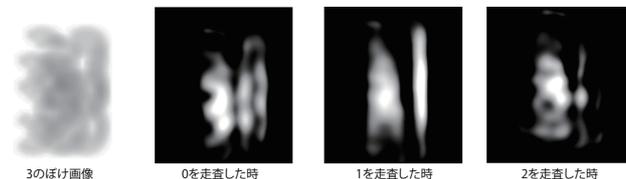


図 6 ぼけ画像とテンプレートごとに相関で生成される画像

と異なる。以上よりぼけ画像に対して、データベースに登録されているテンプレートとそれぞれ逆畳み込み積分を行う。すると理想では、ぼけ画像と同じ文字の場合は図 4 左のような、違う文字の場合は図 4 右のような画像が得られるため、その画像から距離を推定するとともにテンプレートが写っているかどうかを判別できるのではないかと考えられる。

しかし 2.1 節で述べた通り、式 (7) において S に 0 またはそれに近い値が多く含まれるとノイズが強調されてしまい、距離情報を正しく推定できない。また、式 (7) では信号を除算する処理を行なっているため、折り返し雑音などの信号処理問題も併せて発生する。以上より、鮮明にぼけ関数を得たいものの現実を図 5 のようになり、計算によりぼけ画像から文字を認識しようとしても上手くいかない。

3.3 提案手法 1, 2

3.2 節で述べたとおり、テンプレートとの逆畳み込み積分を用いてぼけ画像から距離を推定することは困難である。そのため逆畳み込み積分を使わずに距離を推定する手法が必要となる。

ここで、ぼけの形は絞りの形状に依存し、その大きさは焦点面からの距離に従い変化するという点に着目する。もし、元画像が既知であるならば、ぼけ画像との相関係数を求めた時に複数のピークが存在し、なおかつそのピークの分布、勾配は絞りの形状に近いと考えられる。そこで本論文ではテンプレートとの相関を用いる手法を提案する。

本節では提案手法 1 と 2 について述べる。提案手法 1 では 3.3.1 節、3.3.2 節、3.3.3 節の順に処理を行い、提案手法 2 では 3.3.4 節、3.3.2 節、3.3.3 節の順に処理を行う。以下にて具体的な処理の内容について説明する。

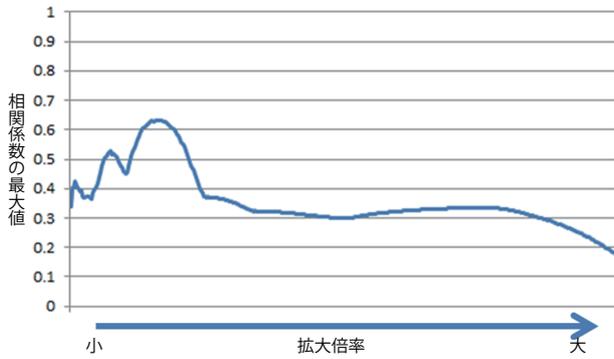


図 7 拡大倍率と相関係数の最大値の関係

3.3.1 相関による画像生成

ぼけ画像に対し、データベースに存在するテンプレート画像を各々走査し、相関係数を求める。その際に相関の式として、正規化相互相関の1つである Zero-mean Normalized Cross-Correlation (ZNCC) を用いる。これはテンプレート及び画像の輝度値の平均値を0にして求める正規化相互相関である。ぼけ画像を t 、テンプレートを s 、それぞれの輝度値の平均を \bar{t} , \bar{s} とおくと、ZNCC により求まる相関係数 r は以下の式で表される。

$$r(x, y) =$$

$$\frac{\sum_{i=-\frac{P}{2}}^{\frac{P}{2}-1} \sum_{j=-\frac{Q}{2}}^{\frac{Q}{2}-1} ((t_{(x+i,y+j)} - \bar{t})(s_{(x+i,y+j)} - \bar{s}))}{\sqrt{\sum_{j=-\frac{P}{2}}^{\frac{P}{2}-1} \sum_{i=-\frac{Q}{2}}^{\frac{Q}{2}-1} (t_{(x+i,y+j)} - \bar{t})^2 \sum_{j=-\frac{P}{2}}^{\frac{P}{2}-1} \sum_{i=-\frac{Q}{2}}^{\frac{Q}{2}-1} (s_{(x+i,y+j)} - \bar{s})^2}}$$

ここで P, Q はそれぞれテンプレートの横と縦の長さを指す。そして図 6 のように各座標 (x, y) ごとに求まった相関係数を輝度値とする画像を保存する。(相関係数が高いほど白い) この処理を走査したテンプレートごとに行い、相関を表す複数の画像を作成する。

3.3.2 距離の推定

距離を推定することは絞りの拡大倍率、つまりは式 (6) の M を求めることと等しいので、生成した相関の画像を基に、開口形状の拡大倍率を推定する。

処理の流れを説明する。開口形状の画像を拡大していき、拡大する度に相関画像に対し走査して、ZNCC を用いて相関係数を求める。そして相関係数の最大値およびその座標を拡大倍率ごとに保存する。このとき、横軸に拡大倍率を、縦軸に相関係数の最大値をとるグラフを生成すると、図 7 のように複数の極大値を持つ。そこで、この極大値を距離情報の候補として、走査したテンプレートごとに座標とともに複数保存する。

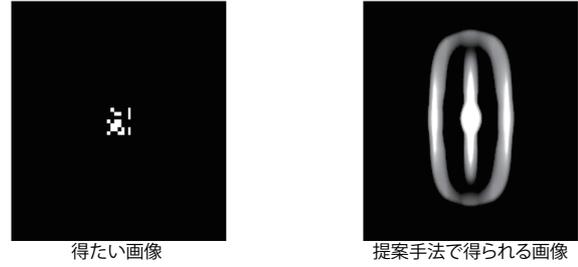


図 8 本来得たい画像と提案手法で得られる画像 (文字:0)

3.3.3 ぼけた画像の文字認識

距離情報が求まると、ぼけ画像からぼけを除去して文字認識を行うことも、ぼけ画像をそのまま認識することもできると考えられるが、今回はぼけたまま認識を行う。

3.3.2 節で求めた距離情報の候補を基に、対応するテンプレートを式 (3) を用いてばかす。その際、式に含まれるノイズ n は考慮しない。これにより、テンプレートの種類×保存した距離候補の数を持つ、ぼけたテンプレートのデータベースが作成される。こうして作成したデータベースで、距離情報とともに保存した座標情報の下、ぼけ画像とマッチングを行い、文字を認識する。

3.3.4 提案手法 1 の問題点と提案手法 2 による解決

3.3.1 節にて、テンプレートを走査し相関係数を求め、それを画像として保存したが、ピントの合っている画像や、ぼけが小さい時の画像に対してテンプレートを走査させると、図 8 右のような画像が得られる。この画像から距離を推定しようにも周囲に存在するノイズがそれを阻害すると考えられる。

このノイズの発生原因は、相関係数を求めて画像として保存する際の相関係数と輝度値の関係にある。提案手法 1 では相関係数と輝度値の関係を線形にしている。相関係数はピクセルごとの輝度値の類似度の総和であるため、類似度が高いピクセルの数が増えれば、それだけ相関係数も高くなる。輝度値との関係を線形にすると少し相関係数が高くなっただけで輝度値も同様に高くなり、ぼけが小さい時に図 8 のように文字ごとに特有のパターンが発生し、距離推定を困難にしている。そのため、ぼけの度合いに関係なく文字を認識するためには、文字ごとに存在するこの特有のパターンを除去する必要があると考えられる。このノイズを除去するため、相関係数と輝度値の関係が 2 次関数的な増加になるよう変化させる。これにより、相関係数が高い箇所を際立たせ、ノイズを除去することが可能になる。その際、得られる画像は図 9 のように変化する。なお、ぼけが大きい時に得られる相関係数は小さい時に比べて低く、そのまま輝度値とすると、全体的に暗い画像となるため、輝度値を -1 から 1 の範囲に正規化する。こうして生成された画像から 3.3.2 節の処理により距離を推定し、3.3.3 節の処理により文字を認識する。



図 9 相関係数と輝度値の関係を変化させた時の差

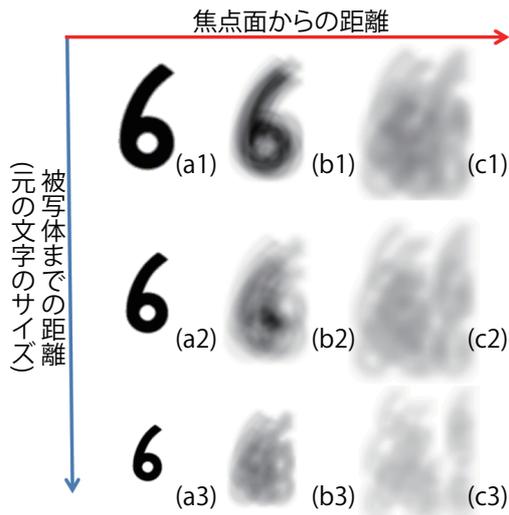


図 10 2つの距離から得られる画像のぼけたパターン

3.4 提案手法 3

提案手法 3 は提案手法 1, 2 と異なり, 距離情報の推定を行わず認識する手法である. 提案手法 1, 2 は文字のスケールが既知であるという前提条件が必要となる. しかし実問題では図 10 のように被写体までの距離と焦点面からの距離の 2 つの距離からぼけた画像は生成される. ゆえに実問題に対応するにはこの 2 つの距離を同時に推定する必要がある.

ここで, 2 つの距離により生成される文字のぼけパターンは, 形だけを見ると 1 つのパラメータの変化で表されることに着目する. 例えば, あるぼけた文字に対し, 文字のスケールと, 焦点面からの距離をそれぞれ倍にすると, そのぼけたパターンは形が一致する. そこで提案手法 3 では, 本来図 10 のように 2 つの距離により変化するものを, 焦点面からの距離にのみ限定し, ぼけた文字のパターン (以後, 劣化テンプレートとする) を複数用意することでの認識を目指す.

処理の流れを説明する. まず, 焦点面からの距離, つまり開口形状の拡大倍率を複数用意し, これを基に全てのテンプレートをぼかす (図 10 の横軸に対応). これにより図 10(a1), (b1), (c1) のような劣化テンプレートが生成される. これらをそれぞれ拡大縮小することで図 10

1234567890

図 11 実験で用いるテンプレート

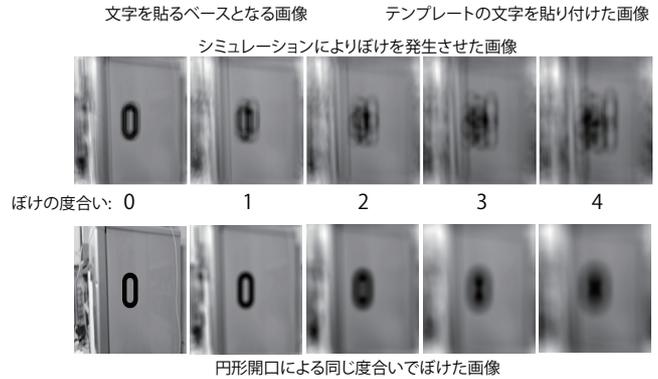


図 12 クエリに用いた画像例

の縦軸 (a2~c3) を作成する. その後ぼけ画像に対して走査することで文字を認識するというものである.

4. 実験と考察

本節では, シミュレーションによる実験で提案手法の有効性を検証するとともにその考察について述べる.

4.1 実験条件

本実験で用いたテンプレートは図 11 に示す 0 から 9 の数字画像 10 枚である. クエリとして図 12 のようにグレースケール画像に各テンプレートを貼り付け, シミュレーションによりピントのあっている 0 段階目から大きくぼけた 10 段階目までの 11 段階にぼかした画像 110 枚を用意した. 開口形状の拡大倍率は 1 段階目から 10 段階目まで, それぞれ 3 倍, 6 倍, 9 倍, ..., 30 倍となっている. これはぼけの度合いが 1 上がるたび, 焦点面からの距離が等間隔に離れていることを意味する. 実験に用いた絞りの形状は Veeraraghavan の開口形状である.

4.2 実験結果

上記実験条件のもと提案手法 1, 提案手法 2, 提案手法 3 の認識精度を調べる.

本実験に際し, 比較手法として逆畳み込み積分による文字認識手法を用いる. 具体的には, 3.3.1 節で生成した相関による画像ではなく, 3.2 節の式 (7) に記したぼけ画像とテンプレートで逆畳み込み積分により画像を生成する. そしてこの画像から 3.3.2 節の処理により距離を推定し, 3.3.3 節の処理で文字を認識する.

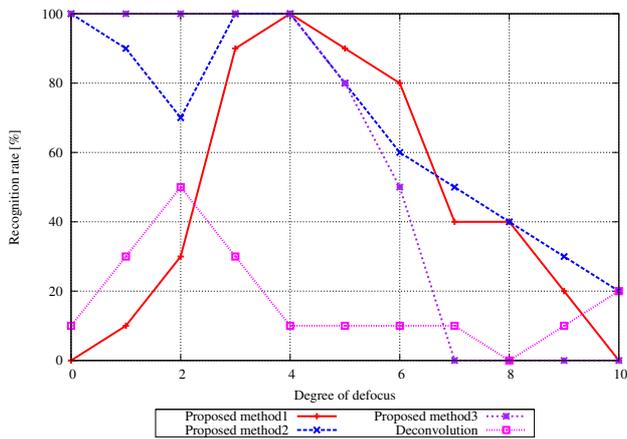


図 13 ノイズを付加しない時の実験結果

なお、提案手法3で用意した劣化テンプレートは、4.1節で述べたぼけの度合11段階全てである。提案手法1, 2では文字が写っている位置を特定し、提案手法3ではクエリのぼけの度合全てを把握している。そのためクエリと同じ文字を返していても、提案手法1, 2では位置が正しくなければ、提案手法3ではぼけの度合を正しく返していなければ、それぞれ誤認識とみなした。また、本実験は文字のスケールが既知のものとしている。そのため、提案手法3における拡大縮小の処理を今回は省略した。

4.2.1 ノイズがない場合

ノイズがない場合の実験の結果を図13に示す。縦軸が認識率、横軸がクエリ画像のぼけの度合を表す。逆畳み込み積分により得られる画像は、3.2節で述べたように荒い画像となるため、その影響により獲得した距離情報及び座標は真の値と大きく異なるものになってしまい、すべてのぼけの度合において、提案手法2より高い認識率を出すことができなかった。

ぼけの度合が3と4の時、提案手法1ではそれぞれ90%と100%、提案手法2ではともに100%と高い認識率を示した。認識に成功したクエリ画像例を図14に示す。提案手法で得られる画像の中央に開口形状に似た模様が確認でき、距離情報と位置情報共に推定が成功したことがわかる。ぼけの度合が0から2にかけて、提案手法1の認識率が0%、10%、30%と極めて低い一方、提案手法2の認識率は100%、90%、70%と比較的高くなった。この原因は3.3.4節で示したとおり、ぼけが小さい時のノイズを大幅に除去できたためであると考えられる。しかしそれでも完全に除去できたわけではなく、提案手法2ではぼけの度合が3の時、再び認識率が向上しており、グラフに谷が発生している。この原因は、提案手法により得られる画像の字種による差である。文字ごとに見た認識率の結果を図15に示す。提案手法2を見ると、字種が1の時に認識率が非常に低い。ぼけの度合が2のときの画像を図16に示す。文字が0のときは認識に成功、1のときは失敗している。原因は同じ字種

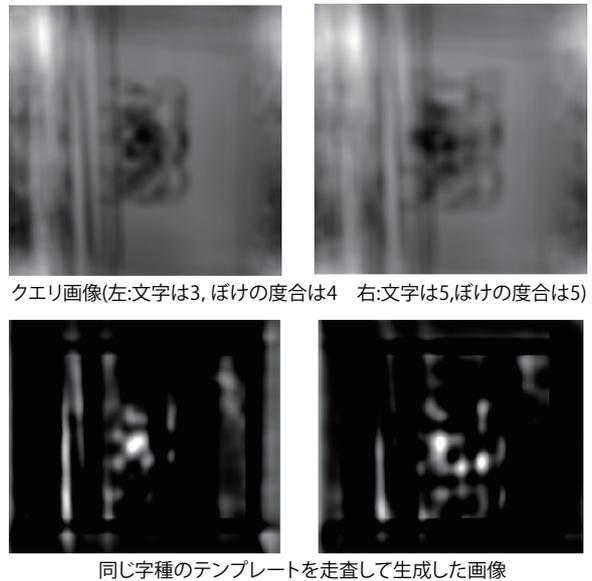


図 14 認識に成功したクエリ画像と提案手法で得られる画像の例

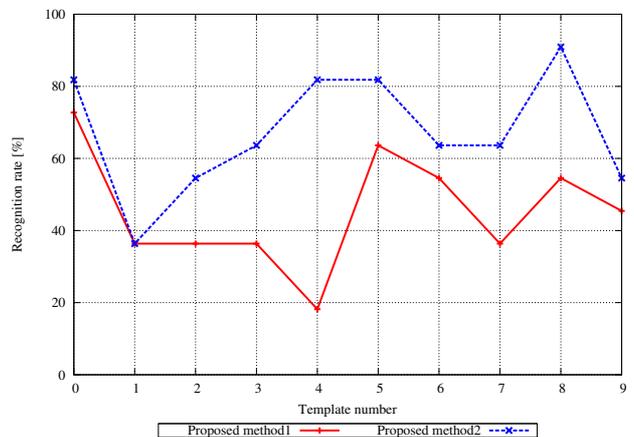


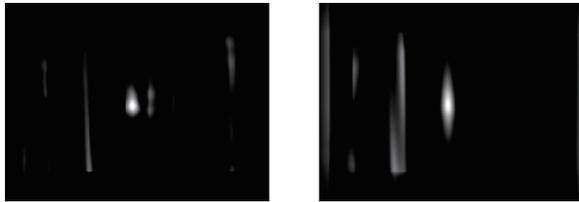
図 15 提案手法1, 2の文字ごとの認識率

のテンプレートを走査した時に生成される画像の差異にある。今回用いた1のテンプレート画像は縦長の棒であり、同じ字種で走査したとき、ぼけた文字が存在する箇所の相関係数が他の文字に比べ高くなっており、提案手法2において輝度値を正規化する際に悪影響を及ぼしている。その結果、図16右下のように、生成される画像に開口形状に似た形状が存在せず、距離情報と位置情報の推定に失敗したものと考えられる。これは生成する画像の輝度値を正規化する方法により改善可能と考えられるため、その方法の考案が課題として挙がる。

また同様に、ぼけの度合が5から10にかけて、提案手法1の認識率が90%、80%、40%、40%、20%、0%、提案手法2の認識率が80%、60%、50%、40%、30%、20%と、ぼけが大きくなるにつれ低下していった。ぼけが大きい時に認識に失敗したクエリ画像例を図17に示す。画像中に含まれる他の線などがぼけることにより、文字に重なり、それが距離推定を阻害したと



ぼけの度合が2の時のクエリ画像(左:0 右:1)



同じ字種のテンプレートを走査して生成した画像

図 16 ぼけの度合が2の時に認識できた文字(0)とできなかった文字(1)

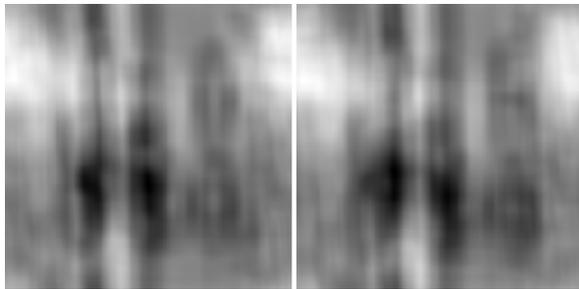


図 17 ぼけが大きい時の0の画像(ぼけの度合:左が9, 右が10)

考えられる。そのため、このように画像に含まれる線や模様の影響を無視できるように、距離推定手法及び認識手法を改良する必要がある。加えて、本実験で用いた Veeraraghavan の開口形状は、元来レンズカメラで撮影された画像からぼけを除去し、鮮明な画像を得るために設計されたものであり、本研究とはそもそもの目的が異なっている。そのため、必ずしも文字認識に、提案手法に適した形状だとは言えない。今後は文字認識に適した開口形状の設計も課題として挙げられる。

提案手法3はクエリ作成時に設けたぼけの度合、つまりクエリと同じぼけ方をした劣化テンプレートをすべて所持しているが、結果を見るとぼけの度合が5の時から認識率が低下し、7からは0%となってしまった。本節の最初に述べたとおり、提案手法3ではぼけの度合も正しく求めていなければ誤認識とみなした。ぼけの度合が5の時から認識率が低下するこの現象は、テンプレートをぼかして認識するというプロセスにおける限界であると考えられる。テンプレートをぼかして、劣化テンプレートを生成する際、テンプレートには少なからず余白が存在し、劣化テンプレートにおける余白の占める割合はぼけの度合が大きくなるほどこれもまた大きくなる。認識の際には正規化相互相関を用いているため、余白が占め



図 18 ノイズを付加したクエリ画像例

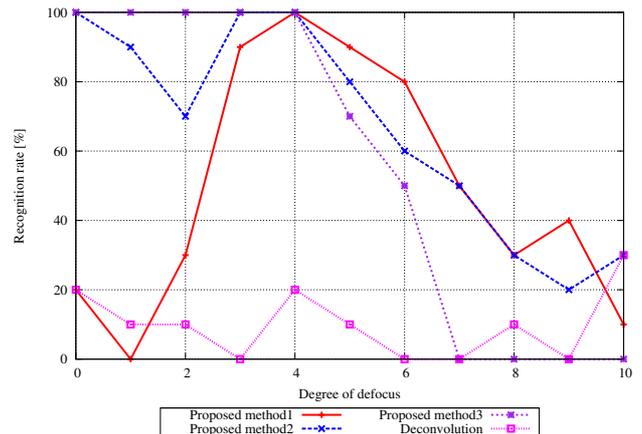


図 19 ノイズを付加した時の実験結果

る割合が大きくなるほど背景や文字周辺の物体の影響を受けやすくなった。その結果相関係数が低下し、クエリと同じぼけの度合でぼかしたにも関わらず、認識に失敗したものと考えられる。

4.2.2 ノイズがある場合

前節ではノイズのない理想的な状態で提案手法の有効性を検証したが、実際の状況では撮像素子によるノイズを考慮しなくてはならない。そこで提案手法がノイズに対して頑健か、そのロバスト性を検証する。クエリとして用意していた画像110枚に対して、図18のように、正規分布に従いノイズを付加し、これを新たなクエリとして同様に実験を行う。

実験結果を図19に示す。比較手法が認識率を低下させたのに比べ、提案手法1, 2, 3はともに認識率の低下が少なく、撮像素子によるノイズに頑健であることが示された。

4.3 文字のスケール変化への対応性

3.4節で述べたとおり、文字の認識において、文字のスケール変化(図10の縦軸に対応)への対応は欠かすことができない。そこで提案手法3が、文字のスケールが未知である実問題に対応できるか、その検証を行うとともに、文字のスケールがクエリと劣化テンプレートで異なる場合に与える影響について調べる。

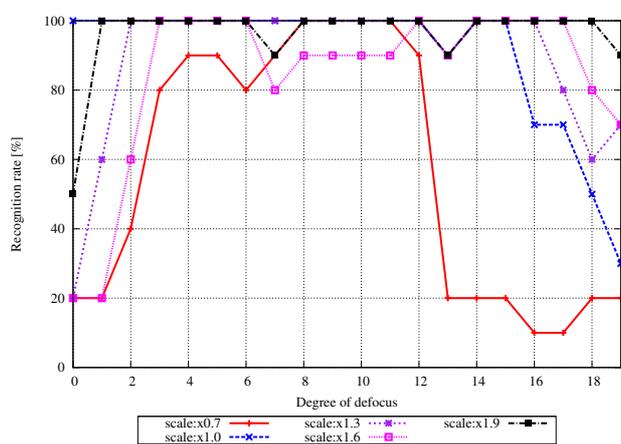


図 20 文字スケールを変化した時の実験結果

この実験で用いるクエリについて説明する。なお、この作成手順は 3.4 節で述べた劣化テンプレートの作成手順とは異なる。まず、図 12 のグレースケール画像に貼り付けるテンプレートのサイズを変化させる。サイズは元のテンプレートの 0.7 倍、1.0 倍、1.3 倍、1.6 倍、1.9 倍の 5 種類である (図 10 の縦軸に対応)。これにより、図 10 の (a1), (a2), (a3) が生成される。これらをシミュレーションによりピントの合っている 0 段階目から 19 段階目まで、20 段階にぼかし、図 10 の横軸 (b1~c3) を作成する。このとき、開口形状の拡大倍率は 1 段階目から 19 段階目まで、それぞれ 1 倍、2 倍、3 倍、…、19 倍となっている。

提案手法 3 のために用意する劣化テンプレートは、上記のぼけの度合での 0 段階目、3 段階目、6 段階目、9 段階目、12 段階目、15 段階目、18 段階目の計 7 種類である (図 10 の横軸に対応)。また、拡大縮小の処理における劣化テンプレートの拡大倍率は、0.6 倍、1.0 倍、1.4 倍、1.8 倍の 4 種類である (図 10 の縦軸に対応)。したがって、1.0 倍の時だけクエリと劣化テンプレートの文字のスケールが一致し、それ以外は一致しない。

結果を図 20 に示す。クエリと劣化テンプレートのスケールが一致する 1.0 倍の時、認識率が高くなった。しかし 4.2 節で述べたとおり、ぼけの度合が大きくなると背景の影響を受けやすくなり認識率が低下した。文字のスケールが一致しない場合は、ぼけの度合が小さい時と大きい時でそれぞれ認識率が低下した。特に 0.7 倍の時に、ぼけの度合が大きくなるにつれ認識率が著しく低下した。これは、文字のスケールが小さければ小さいほど、同じぼけの度合でもぼけはより大きく広がりること起因する。クエリは文字のスケールを変化させた後に文字をぼかした。提案手法 3 は文字をぼかした後にスケールを変化させた。処理手順が逆であるため、クエリのパターンと似た形状の劣化テンプレートが生成されず、認識率が低下したと考えられる。一方で文字のスケールが 1.3 倍、1.6 倍と拡大されるにつれ、ぼけの度合いが大きいに認識率が低下しにくくなった。これは先程の逆で、

文字のスケールが大きいほどぼけが広がりにくくなるため、劣化テンプレートで対応できたと考えられる。

また、文字のスケールが一致しない場合、ぼけの度合が小さいと認識率が低い。これはぼけの度合が小さい時に、提案手法が文字スケールの誤差に極めてナーバスであるということを示している。そのため、今後は文字のスケールを動的に把握する必要がある。

5. 結 び

既存の文字認識技術ではぼけた画像の認識が困難であるという問題に対して、符号化開口を用いて撮影された画像から、写っている文字がテンプレートのいずれかであるという制約の下、テンプレートとの相関により距離推定と文字認識を同時に行う手法と予め設けた距離候補でテンプレートをぼかしマッチングする手法を提案した。提案手法はピントのずれに頑健な認識に対しての可能性を示した。

今後は文字認識に特化した開口形状の設計や、文字のスケール変化に関する手法改良などが課題として挙げられる。

謝辞 本研究は科研費補助金挑戦的萌芽研究 (23650090) の補助を受けた。ここに記して感謝する。

文 献

- [1] 日浦慎作. コンピューショナルフォトグラフィ理解のための光学系入門. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2010-CVIM-171, No. 13, pp. 1-12.
- [2] 長原一. 符号化撮像. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2010-CVIM-171, No. 14, pp. 1-9, 2010.
- [3] 武田祐一, 日浦慎作, 佐藤宏介. 符号化開口ステレオによる被写界深度の拡張. 画像の認識・理解シンポジウム (MIRU2011) 論文集, pp. 1026-1033, July 2011.
- [4] 河本悠, 日浦慎作, 浅田尚紀. 符号化開口を用いたバーコード画像のぼけ除去. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2011-CVIM-177, No. 14, pp. 1-8, 2011.
- [5] S. R. Gottesman and E. E. Fenimore. New family of binary arrays for coded aperture imaging. *Applied optics*, pp. 4344-4352, 1989.
- [6] A. Veeraraghavan, R. Rasker, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhances cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*, Vol. 26, No. 3, p. 69, 2007.
- [7] F. Durand A. Levin, R. Fergus and W. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, Vol. 26, No. 3, p. 70, 2007.