手ぶれ情報を利用したカメラ入力型低品質文字の認識法

石田　皓之† 　高橋　友和† 　井手　一郎† 　目加田慶人†† 　村瀬　洋†

† 名古屋大学大学院情報科学研究科
〒 464-8601 愛知県名古屋市千種区不老町
†† 中京大学生命システム工学部
〒 470-0393 愛知県豊田市海津町床立 101
E-mail: †{hishi,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††y-mekada@life.chukyo-u.ac.jp

あらまし　近年，携帯ディジタルカメラ等のディジタル映像機器が普及しており，それらの機器入力による文字認識技術の実用化が期待されている．しかし，カメラを持つ手の揺れの影響により，文字の認識が困難となる場合が多い．手ぶれの影響を画像復元によって除去する手法は多数存在するが，サイズの小さい文字画像に対するぶれ除去は容易ではない．本稿では，画像復元処理に依存しない手ぶれ文字の認識手法を提案する．本手法では学習段階として，手ぶれを含む様々な劣化要因を生成型学習法によりシミュレーションする．本手法の認識は 2 つの段階から構成される．第 1 段階にて認識結果の候補を絞り，第 2 段階においては，カメラの動きから取得された手ぶれ情報をもとに誤りを起こしやすい字種を分離し，最終的な認識結果を得る．実験により，ぶれ情報の利用が，携帯カメラ入力型文字認識に有効であることを示した．

キーワード　手ぶれ，ディジタルカメラ，低品質文字，文字認識，生成型学習法

A recognition method for camera-captured low-quality characters using blur information

Hiroyuki ISHIDA†, Tomokazu TAKAHASHI†, Ichiro IDE†, Yoshito MEKADA††, and Hiroshi MURASE†

† Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan
†† School of Life System Science & Technology, Chukyo University
101, Tokodachi, Kaizu-cho, Toyota, Aichi, 470-0393 Japan
E-mail: †{hishi,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, ††y-mekada@life.chukyo-u.ac.jp

Abstract　Camera-based character recognition has gained attention with the growing use of camera-equipped cellular phones. In practical cases, however, images taken by hand-held cameras undergo motion blur due to the vibration of the hand. Since it is difficult to remove the motion blur from small characters via image restoration, we propose a recognition method without de-blurring. We introduce a Generative learning method in the training step to simulate blurred images by controlling blur parameters. This recognition method consists of two steps. The first step is based on the subspace method, and the second step re-classifies structurally similar characters using blur parameters estimated from the camera motion. We experimentally proved that the effective use of motion blur improves the recognition accuracy of camera-captured characters.

Key words　Motion Blur, Digital Camera, Low-quality Character, Character Recognition, Generative Learning Method

1. はじめに

携帯ディジタルカメラの普及に伴い，それらの機器を入力とする文字認識技術が近年注目を集めている [1]．カメラに用いられるセンサの高精度化，メモリの大容量化によって，画像の品質は近年大きく向上しているが，それでも高い認識率を得るこ

とは困難な場合が多い．撮影された画像中の文字は低解像度となりやすく，わずかな手の揺れによっても，ぶれを生じやすい．画質復元により，ぶれをはじめとする画像劣化の影響を除去する手法 [2]– [8] は多数提案されているが，サイズの小さい文字に対する有効性は十分に検討されていない．

我々は復元処理を用いない手ぶれ文字の認識手法を提案する．本手法は，ぶれを劣化要因として除去するのではなく，識別のための情報として積極的に利用することで認識率の向上を図る．さまざまな方向・強さの手ぶれを含む文字画像を学習しておき，認識時にはカメラの動きに伴う撮影画像中の文字位置の変化からぶれ情報を推定し，学習データとの類似度評価に利用する（図 1）．本手法では生成型学習法 [9] に基づく学習処理を行い，複数フレーム認識 [10] を用い認識する．本手法における認識は 2 段階からなる．第 1 段階では部分空間法 [11] を用い，認識結果の候補を絞る．部分空間法は単独で用いても有効な手法であるが，過度な劣化・変形を想定したサンプルを学習データとした場合に，形状の似た字種間で認識誤りが増加する．そのため本手法では認識第 2 段階を導入する．第 2 段階では撮影時に推定した手ぶれ情報を利用して，入力画像と生成した学習データとの類似性を評価し，最終的な認識結果を得る．

本稿は以下の章より構成される．2 章では，関連研究について述べる．3 章で生成型学習に基づく学習データの生成について述べる．4 章では，部分空間法に基づく認識第 1 段階について述べる．5 章では，認識第 2 段階について述べる．6 章の実験で，提案手法の有効性を検討し，7 章でまとめる・

2. 関 連 研 究

文字認識は，今まで多くの研究が積み重ねられてきた分野であるが，低品質な文字に対する認識性能は未だ不十分である．この問題への対処法の一つとして，画像復元 [2] がしばしば用いられてきた．低解像度文字をスムージングによって復元する手法 [3]，複数のフレームの情報を補間することで文字の高解像度化を図る手法 [5] が考案されている．一方，光学ぼけを受けた画像を復元するためには PSF(Point Spread Function) を利用することが有効である．他，Teager フィルタを用いた超解像手法 [6] も提案されている．また，カメラの動きによるぶれの除去のためには，ぶれ方向推定 [7] が必要であり，それをもとにした動きぶれ補正手法 [8] も考案されている．しかし，これらの手法を用いて認識対象の画像を復元するためには，認識時に前処理として復元を行わなければならず，効率的ではない．また，サイズの小さな文字画像に対するぶれ除去は困難である．

本稿の手法は復元処理を用いることなく，劣化画像を学習データとすることでカメラ入力に伴う画像の劣化に対処する．

3. 学習データの生成

生成型学習法 [9], [12] は，原画像から多数の学習画像を生成する手法である．様々な手ぶれ画像を撮影により網羅的に収集することは現実的でないため，生成型学習法の導入が効果的である．

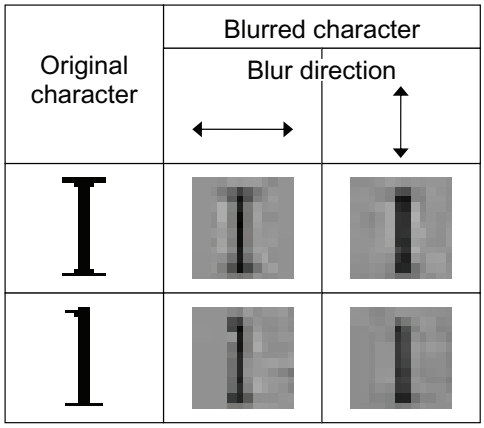


図 1　撮影時のぶれ方向推定値とぶれ画像.

3.1 生成モデル

生成型学習法では，生成時の劣化を制御するための生成パラメータを定義する必要がある．本研究では，以下に示す 4 つのモデルに基づき生成パラメータを定める．

- 光学ぼけモデル

原画像 z_0 に対し，光学ぼけ PSF h_{opt} をたたみ込むことでぼけを含む画像 z_1 を以下の式で生成する．

$$z_1(x,y) = z_0(x,y) * h_{opt}(x,y) \tag{1}$$

認識に用いるカメラの光学ぼけ PSF はあらかじめ推定しておく必要がある．3.2 節にて推定の概要を示す.

- 動きぶれモデル

1 枚のフレームに現れるぶれの方向・強さが一定である場合，動きぶれは容易にモデル化できる [13]．ぶれの強さを表すパラメータ b，およびぶれの方向角を表すパラメータ θ を用い，ぶれる前の画像 z_1 からぶれを加えた画像 z_2 を以下の式で生成する．

$$z_2(x,y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} z_1(x - bt\cos\theta, y - bt\sin\theta)dt \tag{2}$$

この式は，パラメータ b, θ を含む動きぶれ PSF $h_{mot(b,\theta)}(x,y)$ を定義することで，より簡潔に表現することができる．式 (2) に 2 次元フーリエ変換を施すことで，以下のぶれ成分が分離される．

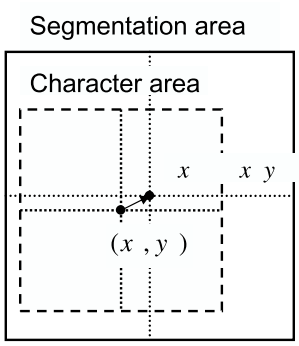
$$H_{mot(b,\theta)}(u,v) = \frac{\sin\left[\pi b(u\cos\theta + v\sin\theta)\right]}{\pi b(u\cos\theta + v\sin\theta)} \tag{3}$$

これに逆フーリエ変換を施し $h_{mot(b,\theta)}(x,y)$ を得る．結果として，式 (2) は以下の式で表される．

$$z_2(x,y) = z_1(x,y) * h_{mot(b,\theta)}(x,y) \tag{4}$$

- 切り出しモデル

文字の切り出し位置ずれ，および切り出し範囲の伸縮をパラメータを用いてシミュレートする．まず文字の原画像に対し，文字全体を含む最小の正方形領域を文字領域（中心 (x_o, y_o), 一辺 l）と定義する．本モデルでは，文字領域に対する切り出し領域の水平方向ずれ Δx，垂直方向ずれ Δy，および文字の拡大率 a の 3 つのパラメータを導入する（図 2）.



—

l

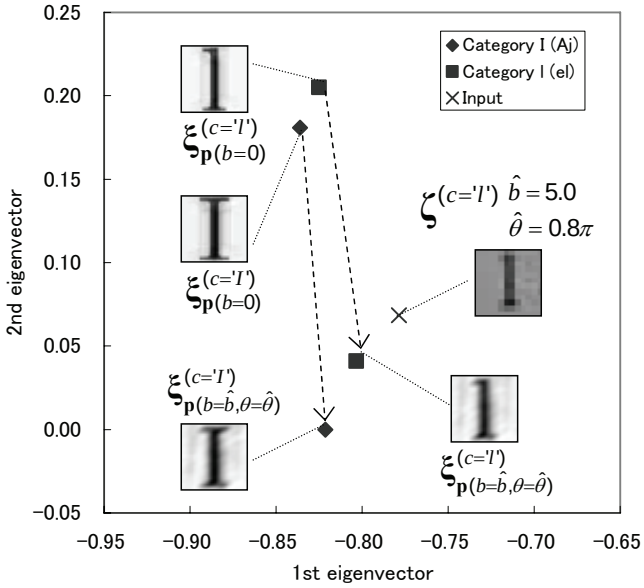


図 7　ぶれ情報の推定により識別精度が改善する例.
認識結果候補：I（◆），l（■）．入力画像：l（×）.
ぶれ情報を推定しなかった場合（誤り）：
$\left\|\boldsymbol{\zeta}^{(c=\text{`}l\text{'})} - \boldsymbol{\xi}^{(c=\text{`}I\text{'})}_{\boldsymbol{p}(b=0)}\right\| < \left\|\boldsymbol{\zeta}^{(c=\text{`}l\text{'})} - \boldsymbol{\xi}^{(c=\text{`}l\text{'})}_{\boldsymbol{p}(b=0)}\right\|$
ぶれ情報を推定した場合（正答）：
$\left\|\boldsymbol{\zeta}^{(c=\text{`}l\text{'})} - \boldsymbol{\xi}^{(c=\text{`}I\text{'})}_{\boldsymbol{p}(b=\hat{b},\theta=\hat{\theta})}\right\| > \left\|\boldsymbol{\zeta}^{(c=\text{`}l\text{'})} - \boldsymbol{\xi}^{(c=\text{`}l\text{'})}_{\boldsymbol{p}(b=\hat{b},\theta=\hat{\theta})}\right\|$

タを $\Xi^{(g)}$ 中に存在する $\boldsymbol{p}$ の成分 b,θ から選び，$b_\alpha, b_\beta, \theta_\alpha, \theta_\beta$ とする．ただし，$b_\alpha \leqq \hat{b}_m < b_\beta, \theta_\alpha \leqq \hat{\theta}_m < \theta_\beta$ とする．点 $\boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b=\hat{b}_m,\theta=\hat{\theta}_m)}$ は線形補間により以下の式で求められる．

$$\begin{aligned}\boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b=\hat{b}_m,\theta=\hat{\theta}_m)} = {} & \frac{b_\beta-\hat{b}_m}{b_\beta-b_\alpha}\frac{\theta_\beta-\hat{\theta}_m}{\theta_\beta-\theta_\alpha}\boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b=b_\alpha,\theta=\theta_\alpha)} \\ & - \frac{b_\beta-\hat{b}_m}{b_\beta-b_\alpha}\frac{\theta_\alpha-\hat{\theta}_m}{\theta_\beta-\theta_\alpha}\boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b=b_\alpha,\theta=\theta_\beta)} \\ & - \frac{b_\alpha-\hat{b}_m}{b_\beta-b_\alpha}\frac{\theta_\beta-\hat{\theta}_m}{\theta_\beta-\theta_\alpha}\boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b=b_\beta,\theta=\theta_\alpha)} \\ & + \frac{b_\alpha-\hat{b}_m}{b_\beta-b_\alpha}\frac{\theta_\alpha-\hat{\theta}_m}{\theta_\beta-\theta_\alpha}\boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b=b_\beta,\theta=\theta_\beta)} \qquad (21)\end{aligned}$$

最終的な認識結果はぶれ情報をもとに，次式で表される．

$$\hat{c}_2 = \arg\min_{c\in\mathcal{G}^{(g)}} \sum_{m=1}^{M} \min_{b=\hat{b}_m,\theta=\hat{\theta}_m} \left\|\boldsymbol{\zeta}_m - \boldsymbol{\xi}^{(c)}_{\boldsymbol{p}(b,\theta)}\right\| \qquad (22)$$

以上に述べた認識第 2 段階の概略を図 7 に示す．この図では，ぶれ情報を推定して利用することで識別精度が向上する例を示している．

6. 実　　験

提案手法の有効性を確めるため実験を行った．実験では 640×480 pixel，30 fps で撮影可能なディジタルカメラ（Panasonic DMC-FX9）を使用した．このカメラにより文字の印刷された紙を動画像として 10 秒間撮影し，試験用データを得た．認識対象は英数字 62 文字（A-Z, a-z, 1-9：Century フォント）とし，これらの文字が各 1 文字づつ印刷された紙を被写体として用いた．紙の上の文字の大きさは平均 5 mm 四方であった．撮影時はカメラと被写体の距離を 30 cm とし，広角（焦点距離 5.8 mm）で撮影した．文字の切り出し時は，個々の文字領域を含む最小の正方形領域を切り出し領域とし，自動で切り出した．このとき，全 62 文字を切り出すことのできなかったフレームは抽出失敗とし，認識対象から除外した．撮影された文字の平均サイズは 11 pixel 四方となった．なお，カメラの手ぶれ補正機能は OFF とした．

6.1 学 習 処 理

生成型学習を適用するにあたり，光学ぼけ PSF を推定する必要がある．そのため，カメラと被写体の距離を 30 cm に保ち手で静止させた状態で，文字の書かれた文書を動画像として撮影し，それから光学ぼけ PSF を推定した．式 (5) で定義したパラメータベクトル $\boldsymbol{p}$ を様々に変化させ，学習データを生成した．具体的には，劣化パラメータを $d = 0.25, \cdots, 2.00$ の 8 通り，ぶれ強度パラメータを $b = 0.0, \cdots, 5.0$ の 11 通り，ぶれ方向角パラメータを $\theta = 0, \cdots, \frac{11}{12}\pi$ の 12 通り，拡大率パラメータを $a = \frac{13}{16}, \frac{14}{16}, \frac{15}{16}, 1$ の 4 通り，位置ずれパラメータ $\Delta x, \Delta y$ をそれぞれ $-a, -\frac{1}{2}a, 0, \frac{1}{2}a, a$ の 5 通りに変化させ，各カテゴリ毎に合計 105,600 個の学習データ（サイズ 32×32 pixel）を生成した．4.1 節で述べた手順によって固有ベクトルを計算し，上位 10 個を認識の第 1 段階に用いた ($R_1 = 10$).

次に，誤りを起こしやすい字種の組をグループ化した．グループ化用の認識サンプルは，カメラを手で静止させ，被写体間の距離を 30 cm とした状態で 10 秒間撮影した．この認識サンプルに対する認識第 1 段階の混同行列を表 1 に示す．本実験ではグループ化の閾値 τ を 0.1 と設定し，以下の 3 つのグループを認識第 2 段階の対象とした．

group Z : $\mathcal{G}^{(\text{`}Z\text{'})} = \{\text{`}Z\text{'}, \text{`}z\text{'}\}$
group l : $\mathcal{G}^{(\text{`}l\text{'})} = \{\text{`}I\text{'}, \text{`}l\text{'}, \text{`}i\text{'}, \text{`}j\text{'}\}$
group 1 : $\mathcal{G}^{(\text{`}1\text{'})} = \{\text{`}1\text{'}, \text{`}i\text{'}\}$

構成されたグループをもとに，5.2 節で述べた手順によって固有空間を作成した．固有ベクトルは上位 10 個を使用した ($R_2 = 10$)．各グループ $\mathcal{G}^{(g)}$ 毎に，含まれるカテゴリの全学習データを固有空間に投影し，点集合 $\Xi^{(g)}$ を得た．

6.2 撮 影 条 件

試験用データの撮影条件として，以下の 3 通りの条件を定める．

条件 A　カメラを三脚で固定して撮影する．
条件 B　カメラを手で持ち，静止させた状態で撮影する．
条件 C　カメラを手で持ち，揺らしながら撮影する．

試験用データは，各 10 フレームからなる動画像とした ($M = 10$)．条件 A では 300 本の動画像を使用し，条件 B, C では 6 人の被験者によって各 300 本ずつ撮影された動画像を使用し，その平均の認識率を求めた．試験用データの例を図 8 に示す．なお，条件別のぶれ強度推定値（1 フレームあたりの平均変位ピクセル量）の分布を図 9 に示す．

6.3 実 験 結 果

以上の条件のもとで撮影された全カテゴリの文字画像に対し，提案手法を適用した結果を図 10 に示す．比較手法として，認

表 1　グループ化に利用した混同行列．行列中の値は $\rho(g|c)$ を表す．$\rho(g|c) \geqq 0.01$ となるカテゴリのみ示す．

	recognition result g												
input c	I	L	S	W	Z	i	j	l	s	t	w	z	1
I	0.62							0.36					0.01
L		1.00											
S			1.00										
W				1.00									
Z					1.00								
i	0.02					0.19		0.73					0.06
j							0.52	0.36					0.12
l								1.00					
s			0.07						0.93				
t		0.01								0.99			
w				0.01							0.99		
z					0.51							0.49	
1								0.03					0.97

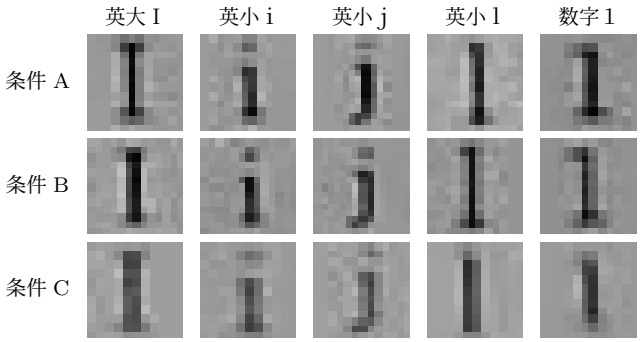


図 8　試験用データの例．

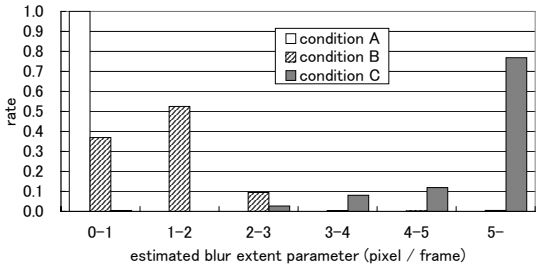


図 9　ぶれ強度推定値の分布．

識第 1 段階のみの認識率と，手ぶれ情報を推定せず，ぶれを与えていない学習データ（$b = 0$）とのみ照合した場合の認識率を示す．ぶれの影響が少ない条件 A, B においても，認識第 1 段階のみを用いる手法と比べ認識率が向上している．なお，複数のフレームを用いて認識した場合，カメラを固定した場合（条件 A）と比べ，適度な画像変化がある場合（条件 B）の方が認識精度が高いことが確かめられている [15] が，本実験においても同様の結果が得られた．

同様に，第 2 段階が適用され（$|\mathcal{G}^{(\hat{c}_1)}| \geqq 2$），かつそのことにより正答が得られる可能性がある（$c \in \mathcal{G}^{(\hat{c}_1)}$）文字のみを対象とした認識率を評価した．ここで，文字の正答カテゴリを c, 第 1 段階の認識結果を $\hat{c}_1$，第 2 段階の認識結果を $\hat{c}_2$ と表す．以下の認識率 P_{i}, P_{ii} を評価した．

- 第 1 段階終了時における認識率

$$P_{\mathrm{i}}\Big(\hat{c}_1 \equiv c \,\Big|\, (|\mathcal{G}^{(\hat{c}_1)}| \geqq 2) \wedge (c \in \mathcal{G}^{(\hat{c}_1)})\Big)$$

- 第 2 段階適用後の認識率

$$P_{\mathrm{ii}}\Big(\hat{c}_2 \equiv c \,\Big|\, (|\mathcal{G}^{(\hat{c}_1)}| \geqq 2) \wedge (c \in \mathcal{G}^{(\hat{c}_1)})\Big)$$

この結果を図 11 に示す．図 10 と同様に，手ぶれ情報を推定しなかった場合の認識率と併せて示す．

6.4 考　　察

実験結果より，認識 2 段階目における再分類が有効であることが明らかとなった．手ぶれの影響が大きい条件 C（ぶれ強度推定値 3[pixel/frame] 以上）では，ぶれ情報の利用が識別精度改善のために有効であった．しかし，それ以外の条件下では，認識を 2 段階に分けたことの影響が大きい．

グループ化された文字のみを対象とした分析結果（図 11）によると，P_{i} に対し P_{ii} は大きく改善しており，第 1 段階では正しく分類できていなかった多数の文字が，認識第 2 段階により正しいカテゴリに分類されたことを表している．認識第 2 段階において良好な再分類結果を得るためには，適切なグループ化がされている必要がある．P_{ii} はグループ化時の閾値 τ と関連があるため，実際の P_{ii} を想定し，$\tau \leqq 1 - P_{\mathrm{ii}}$ となる閾値設定が有効であると考えられる．様々な τ に対し，条件 B における認識第 2 段階適用前後の認識率を調べた結果を図 12 に示す．横軸はグループ化された文字を対象とした認識率とし，縦軸は全カテゴリに対する認識率とする．P_{ii} はほぼ 80% であり，いずれの閾値の場合も認識率の改善がみられたが，$\tau < 0.05$ では効果が低下している．これは，グループ化すべきでない文字種までグループ化してしまったためである．

7. おわりに

本稿では，携帯型カメラで撮影された文字の認識率向上のために，ぶれ情報を認識に利用する手法を提案した．誤りやすい字種を識別するため，2 段階の認識を導入し，推定されたぶれ

100%