

携帯カメラ入力型文字認識におけるぼけやぶれに対処するための 生成型学習法

石田 皓之[†] 高橋 友和[†] 井手 一郎[†] 目加田慶人^{††}
村瀬 洋[†]

A Generative Learning Method for the Recognition of Blurred Characters
Taken by Portable Cameras

Hiroyuki ISHIDA[†], Tomokazu TAKAHASHI[†], Ichiro IDE[†], Yoshito MEKADA^{††},
and Hiroshi MURASE[†]

あらまし 近年、デジタルカメラ、カメラ付携帯電話が普及しており、それらの機器を用いた文字認識技術の実用化が期待されている。しかし、撮影された画像は低解像度なものになりやすく、更にぼけやぶれなどの影響を受けるため、高精度な認識は難しい。我々の提案する生成型学習法は、人工的に劣化を加えて生成した画像を学習データとすることで、特に携帯デジタルカメラで問題となる光学的ぼけ、手ぶれ、低解像度化に対処するものである。従来は、学習データを撮影により収集する手法が用いられてきたが、撮影環境が異なるデータに対応するために様々な条件のもとで学習データを収集する必要がある。それに対し本手法は、学習データを原画像から生成するため効率的である。また生成時のパラメータを変化させることで多種の劣化に対応する学習データが生成できる。本手法を用い、実際の劣化を学習に反映させることで、ぼけやぶれを含む低品質文字の認識率が向上することを実験によって示した。

キーワード 生成型学習法、文字認識、低品質文字、部分空間法

1. ま え が き

デジタルカメラやカメラ機能付携帯電話といったデジタル映像入力機器が近年普及している。それらの機器を用いた文字認識が可能となれば、携帯カメラの文字入力インタフェースを改善させる有用な技術となる [1]。しかし認識対象が低解像度の文字となり、ぼけや手ぶれの影響を受けやすいことから、実用化に至るまでには多くの課題がある。センサの高精度化、メモリの小型大容量化や低価格化に伴い、撮影画像の品質は大きく向上しているが、それでも十分な品質の画像を得られない場合が多い。例えば、利用者が一度に広範囲の文字領域を取得しようと試み、カメラを被写体から遠ざければ、それも低解像度化の原因となる。

柳詰らは、動画像中の複数のフレームを用い、カメラ入力の低解像度文字をロバストに認識する手法を考案した [2], [3]。この手法では、認識時と同じカメラを用い、全カテゴリーの学習データを様々な撮影距離で撮影し収集する必要がある。しかし、収集には多くの手間を要するという問題がある。また、カメラの種類の変更やカテゴリー数の増加に対応することが困難である。

このような問題に対し、少数の学習データに摂動を与え、サンプル数を増やすアプローチが古くから提案されてきた [4]。本論文では、低解像度文字の認識のための生成型学習法を提案する。本手法は学習データを現実の劣化特性に則して人工的に生成するため、学習データを撮影によって収集する従来の学習法と比べ効率的である。また生成時のパラメータを制御することにより、劣化状況を考慮した多種の学習データを作成できるため、認識時の多様な撮影条件にも柔軟に対応することが可能である。本手法では、あらかじめカメラ固有の光学的ぼけ関数 (PSF: Point Spread

[†] 名古屋大学大学院情報科学研究科, 名古屋市
Graduate School of Information Science, Nagoya University,
Furo-cho, Chikusa-ku, Nagoya-shi, 464-8603 Japan

^{††} 中京大学生命システム工学部, 豊田市
Life System Science and Technology, Chukyo University, 101
Tokodachi, Kaizu, Toyota-shi, 470-0393 Japan

Function) を推定し、それを用いて学習データを生成する。様々な解像度の入力データに対処するため、劣化の強さが異なる様々な学習データをパラメータの制御により生成し、部分空間法 [5] を用い認識する。この手法は、収集に基づく手法 [2], [3] と比べ、原画像と PSF のみから様々な解像度の学習データを入手できる。

以下、2. では関連する研究の概要を紹介する。3. では撮影による文字の劣化過程の概要を示し、学習データ生成のためのモデル化について述べる。4. では提案する生成型学習法の詳細を述べる。5. では文字認識手法について述べる。6. で提案手法を用いた比較実験について紹介し、7. で結論を述べる。

2. 関連研究

前章でも述べたように、学習画像を撮影により収集する従来の手法 [2], [3] では、収集に手間を要するという問題がある。また、ぼけやぶれなどの多様な劣化要因に対応することが困難である。これに対して、画像復元によりそれらの劣化を除去するための手法が多数提案されており [6] ~ [9], これらによって復元された画像を認識に用いることも考えられる。Mancas-Thillouらは Teager フィルタを用いた超解像処理により、文書画像の画質を復元する手法を考案した [10]。また、ぶれの影響を含む画像を復元する手法もある [11]。しかし、これらの手法で認識に用いる画像を復元する場合は、復元のためのパラメータを認識時に推定しなければならないため、効率的ではない。

本手法は、認識時に入力されるであろう劣化文字画像を擬似的に生成し、それを学習データとして用いることで、現実の多様な劣化に対応する。そのためにここではカメラ固有の PSF をあらかじめ推定しておき、それに基づいて学習データを生成する手法を提案する。本手法では綱島らによって提案されたコンパウンド法 [12] を動画像に対して適用し、雑音成分を抑制することで PSF の推定を行う。

本論文では、認識において基本的には部分空間法 [5] を用いる。単一フレームの低解像度文字から高い認識率を得ることは困難であるため、柳詰らの手法 [3] に基づき動画像中の複数のフレームを累積的に利用して認識した。

3. 撮影による文字の劣化

生成型学習においては、現実の劣化過程を知り、そ

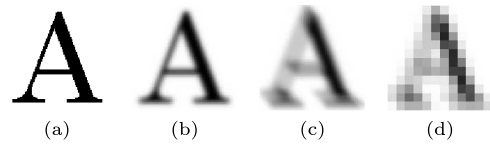


図1 劣化過程 (a) 原画像 (b) 光学ぼけを含む画像 (c) 動きぶれを含む画像 (d) 低解像度画像

Fig.1 Degradation of a character: (a) original image, (b) optically blurred image, (c) motion blurred image, (d) low-resolution image.

れを学習に反映させる必要がある。図1に例示するように、我々は撮影による文字の劣化を以下の三つの要因に分離した。

- レンズ系による光学的ぼけ (図1(b))
- カメラの揺れによる動きぶれ (図1(c))
- 標本化による解像度の低下 (図1(d))

以下で提案する生成型学習法は、これら3種の劣化要因に対応した以下の三つの生成モデルを定義し、それに基づき学習データを生成する。

- (1) 光学的ぼけモデル
- (2) 動きぶれモデル
- (3) 解像度低下モデル

後の実験で、劣化文字に対する各生成モデルの有効性を比較検討する。

4. PSF を用いた生成型学習法

生成型学習法は、学習データとして用いる画像を原画像から人工的に生成する手法である。本章では、前章で述べた生成モデルに従い、現実の劣化特性に基づく学習データを PSF に基づき生成する手法を述べる。本手法を用いた認識までの流れを図2に示す。提案する生成型学習は PSF の推定段階 (Estimation step) と学習データの生成段階 (Generation step) から構成される。以降、4.1にて光学ぼけ PSF, 4.2にて動きぶれ PSF, 4.3にて解像度変換を用いた生成モデルの詳細を述べ、4.4でそれらを統合した生成段階について述べる。

4.1 光学ぼけ PSF

デジタルカメラにより撮影された低解像度文字画像は、ぼけやサンプリングの影響による画像の変動が大きく、単純なテンプレートマッチングによる認識は有効でない。そのため本手法では、カメラの平均的な光学ぼけ特性を表す PSF を用いてぼけをシミュレートする。撮影画像の劣化特性はカメラの種類により様々であり、またガウシアン等の単純なぼかし関数ではそ

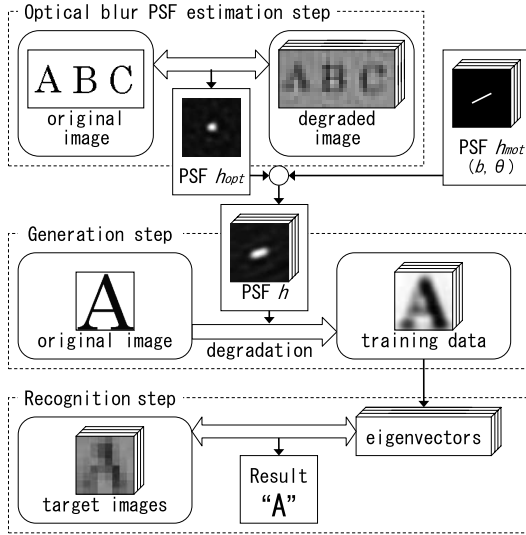


図 2 生成型学習法のフローチャート。
Fig.2 Flow of the generative learning method.

これらの特性を近似することは不十分であると考えられるからである。光学ぼけ PSF と解像度低下モデルにおける解像度パラメータを組み合わせることで、ぼけやサンプリングによる多様な画像変動に対応する学習データを多数生成する。

4.1.1 光学ぼけ PSF 推定のための準備

はじめに、認識に用いるカメラを用いて画像を撮影し、それらから光学ぼけ PSF を推定する必要がある。そのため、推定に用いる図 1(a) のような原画像 (2 値画像) を計算機上で作成し、印刷する。印刷した用紙をカメラで撮影し、劣化画像を得る。文献 [12] の手法を利用するために、一定数のフレーム画像を動画像として撮影する必要がある。光学ぼけ PSF の推定に用いる原画像は様々な方向のエッジ成分を含むものが望ましい [12]。そのため多数の異なる文字が含まれる画像を用いると、良好な推定結果が得やすい。

4.1.2 光学ぼけ PSF の推定

光学ぼけ PSF の推定にはコンパウンド法 [12] を用いる。コンパウンド法は撮影によって得た複数の画像から PSF を推定する手法である。本手法では、計算機上で作成した原画像と、撮影した動画像中の劣化画像を用いて推定する。原画像と劣化画像は縮尺が異なるため、劣化画像に対し原画像を伸縮させながら照合し、位置合せとサイズ合せをする必要がある。

画像の光学ぼけモデルは、 f を原画像、 g を劣化画像、 h_{opt} を光学ぼけ PSF、 n を加法雑音として次のよ

うに表される。

$$g(x, y) = f(x, y) * h_{opt}(x, y) + n(x, y) \quad (1)$$

ここで $*$ は畳込みを表す。式 (1) に二次元フーリエ変換を施し、次式を得る。

$$H_{opt}(u, v) = \frac{G(u, v)}{F(u, v)} - \frac{N(u, v)}{F(u, v)} \quad (2)$$

雑音成分 $N(u, v)$ は未知であるため、1 枚の劣化画像のみから PSF を求めることは困難である。そこで動画像中の複数のフレーム画像から $H_{opt}(u, v)$ の平均を求めることによって雑音成分を抑制する。 k 個の劣化画像を用いるものとする、平均された $\hat{H}_{opt}(u, v)$ は以下の式で得られる。

$$\hat{H}_{opt}(u, v) = \frac{1}{k} \sum_{i=1}^k \frac{G_i(u, v)}{F(u, v)} - \frac{1}{k} \sum_{i=1}^k \frac{N_i(u, v)}{F(u, v)} \quad (3)$$

劣化画像の枚数 k が十分大きい数であるとき、式 (3) の第 2 項は 0 に収束し、 $\hat{H}_{opt}(u, v)$ は次の式のように近似できる。

$$\hat{H}_{opt}(u, v) \approx \frac{1}{kF(u, v)} \sum_{i=1}^k G_i(u, v) \quad (4)$$

最後に逆フーリエ変換を施すことで、光学ぼけ PSF として $\hat{h}_{opt}(x, y)$ が得られる。

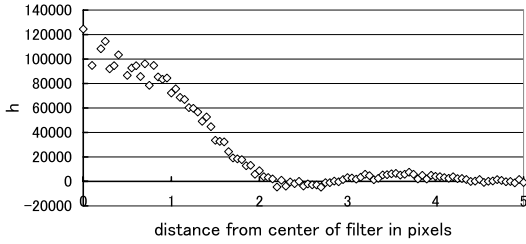
図 3 に本研究で用いる 3 種類のカメラ^(注1)から推定した各々の光学ぼけ PSF を示す。これらは後に述べる実験で推定されたものである。

4.2 動きぶれ PSF

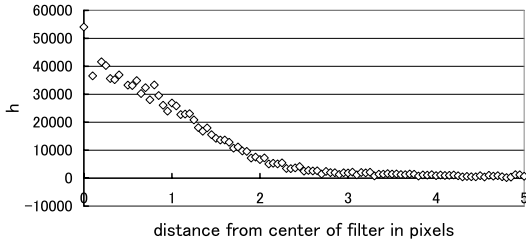
光学ぼけと同様に、動きぶれについても PSF を求め、生成に利用する。動きぶれはパラメータを用いることで容易にモデル化できる [13]。ここではまず 1 枚のフレームに現れる動きぶれの方向、及び速度が一定の場合を考える。したがって、ぶれを加えた画像 z_1 は、ぶれ強度を表すパラメータ b 、ぶれの方向角を表すパラメータ θ を用い、ぶれ処理を行う前の画像 z_0 から以下の式で得られる。

$$z_1(x, y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} z_0(x - bt \cos \theta, y - bt \sin \theta) dt \quad (5)$$

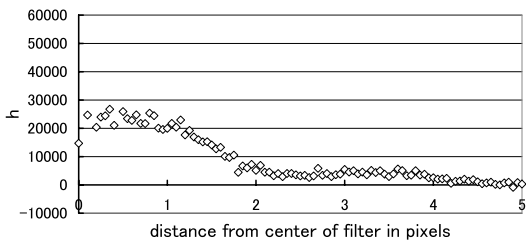
(注1): 本研究で用いたカメラはデジタルビデオカメラ、デジタルカメラ、携帯電話カメラである。型名はそれぞれ、Sony DCR-PC105、Panasonic DMC-FX9-S、NTT ドコモ F901i である。



(a)



(b)



(c)

図 3 光学ぼけ PSF . 中心点 (0, 0) からの距離を横軸とした . (a) デジタルビデオカメラ (b) デジタルカメラ (c) 携帯電話カメラ

Fig. 3 Optical PSF. The x -axis indicates the distance from center of PSF filter. (a) Digital video camera, (b) Digital camera, (c) Camera equipped in cellular phone.

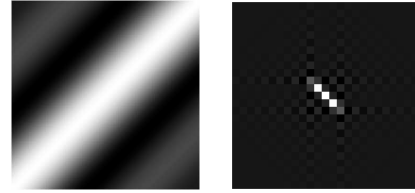
この式に二次元フーリエ変換を施し ,

$$Z_1(u, v) = Z_0(u, v) \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-j2\pi bt(u \cos \theta + v \sin \theta)} dt \quad (6)$$

が得られる . 式 (6) の右辺の第 2 因子がぶれ成分に対応する . この因子をパラメータ b, θ を用いて $H_{mot(b, \theta)}(u, v)$ とし , 積分すると , 以下の式に簡略化される .

$$H_{mot(b, \theta)}(u, v) = \frac{\sin[\pi b(u \cos \theta + v \sin \theta)]}{\pi b(u \cos \theta + v \sin \theta)} \quad (7)$$

この逆フーリエ変換 $h_{mot(b, \theta)}(x, y)$ が動きぶれ PSF である (図 4) .



(a) $H_{mot}(b, \theta)$

(b) $h_{mot}(b, \theta)$

図 4 動きぶれ PSF ($b = 5, \theta = 0.25\pi$)
Fig. 4 Motion blur PSF ($b = 5, \theta = 0.25\pi$).

4.3 解像度変換

以上で述べた光学ぼけ PSF や動きぶれ PSF などの劣化関数を文字の原画像にフィルタとして適用させることで人工的に劣化させた画像が生成される . このとき , 同時に解像度変換も扱うことができる . 劣化関数 (h とする) をそのまま原画像に畳み込む方法も考えられるが , 本手法では解像度パラメータ d を導入することで , 劣化の強さが様々な学習データを作成する . d は劣化関数フィルタの伸縮率に等しい . 生成された学習データは , $d = 1$ の場合には h をそのまま畳み込んだ画像に等しく , $d = 0$ の場合には原画像に等しい . 学習データ x を , 文字の原画像 f_c から以下の式で生成する .

$$x(p, q) = \sum_{i, j} h(i, j) f_c(p - di, q - dj) \quad (8)$$

4.4 学習データの生成

学習データは , 劣化関数 h を用い , パラメータ制御によって生成される . 光学ぼけと動きぶれの両方の生成モデルを用いる場合 , 劣化関数 h は以下の式で表される .

$$h(x, y) = h_{opt}(x, y) * h_{mot(b, \theta)}(x, y) \quad (9)$$

このように求まった劣化関数を用い , 式 (8) によって画像を生成する . 様々な劣化の度合に対処するため , ぶれ強度パラメータ b , ぶれ方向角パラメータ θ , 解像度パラメータ d を変化させ , 多数の学習データを生成する .

5. 認識手法

本手法における認識は , 基本的には部分空間法 [5] 及び複数フレーム認識 [3] に基づくが , 後処理としてカメラ速度を用いた選別法を導入する . 動きぶれを用いる生成型学習法と , カメラ速度に基づく選別とを併用することで , 認識精度の改善が期待される . 6.2 の

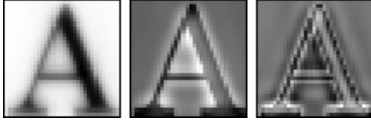


図 5 固有ベクトル (上位 3 個)

Fig. 5 Three eigenvectors corresponding to the top three eigenvalues.

実験にて、カメラ速度選別の有効性を検討する。

5.1 部分空間の作成

部分空間法による文字認識の概要を示す。

あるカテゴリ c のグレースケール画像 N 枚を学習データとして用いるものとする。 n 番目の画像上における各点の濃度値をラスタスキャン方式で並べたベクトルをつくり、平均値が 0 となるように正規化したものを $x_n^{(c)}$ とおく。 N 個の列ベクトル $x_n^{(c)}$ を並べた行列の自己相関行列 $X^{(c)}$ を以下の式で求める。

$$X^{(c)} = \begin{bmatrix} x_1^{(c)} & \cdots & x_N^{(c)} \end{bmatrix} \begin{bmatrix} x_1^{(c)} & \cdots & x_N^{(c)} \end{bmatrix}^t \quad (10)$$

この行列を固有値展開し、固有値の大きい順に、対応する R 個の固有ベクトル $e_r^{(c)}$ を求め、それらを認識に用いる。図 5 に固有ベクトルの例を示す。

5.2 複数フレームからの文字認識

部分空間法は、対象の入力文字画像と部分空間との類似度を計算し、認識する。このとき、動画中の複数フレームの類似度を加算し、その和が最大となるカテゴリに分類することで認識率が向上することが報告されている [3]。複数フレーム認識は、単一フレームからでは文字の識別に必要な情報量が得られないときに有効である。特にカメラが揺れた状態で撮影した場合は、各フレームの画像に変化が現れるため、複数フレームを利用することで認識率の向上が期待できる。

フレーム j における入力文字画像に対し、サイズ及び画素値の正規化を行い、ラスタスキャンによりベクトルにしたものを y_j とする。 F 枚のフレームを用いる場合、入力文字の認識結果 \hat{c} は R 個の固有ベクトル $e_r^{(c)}$ との内積を用い、以下の式で決定される。

$$\hat{c} = \arg \max \sum_{j=1}^F \sum_{r=1}^R (e_r^{(c)} \cdot y_j)^2 \quad (11)$$

5.3 カメラ速度選別を用いた認識

カメラをもつ手の揺れの度合をカメラ速度として定量化する。文字を抽出する際に、画像中の文字位置 (w_j, h_j) の変化量から位置ずれベクトル

$\Delta_j = (w_j - w_{j-1}, h_j - h_{j-1})$ が得られる。複数フレーム認識に用いるフレーム群に対し、位置ずれ量の平均を以下の式によって求め、これをカメラ速度 v_b [pixel/frame] と定める。

$$v_b = \frac{1}{F-1} \sum_{j=2}^F \|\Delta_j\| \quad (12)$$

このカメラ速度に基づき、認識に用いるフレーム群を選別することで、認識精度の向上を図る。

6. 実験

提案手法の有効性を確かめるため実験を行った。本実験の認識段階においては、動画中の連続する 10 フレームを用い、抽出された各文字につき式 (11) により認識結果を求めた ($F = 10$)。

6.1 低解像度とぼけに関する実験

本実験では、カメラの種類と被写体間距離を様々な変化させ、本手法で定義した生成モデル (1) 光学ぼけモデル、(3) 解像度低下モデルの有効性を検討した。比較のため、以下の学習法を定義する。

(手法 A) 生成モデルを用いない単純照合

サイズ 32×32 pixel に正規化した原画像 1 枚を照合に用いる。この手法で認識するときは、入力画像と原画像の内積の 2 乗を類似度とする。他手法と同様に 10 フレームの類似度を累積して用いる ($F = 10$)。

(手法 B) 解像度低下モデルのみに基づく生成型学習法

光学ぼけ PSF を用いず、原画像を 25 通り (サイズ 8~32) に解像度変換した画像を学習データとする ($N = 25$)。その後、ニアレストネイバ法 (手法 B1) 及びバイリニア補間法 (手法 B2) によりサイズを 32×32 pixel に正規化し、部分空間を作成する。固有ベクトル上位 10 個を認識に用いる。

(手法 C) 光学ぼけ関数としてガウシアンを仮定した生成型学習法

原画像 (32×32 pixel) 1 枚に対し、ガウス関数を積み込んだ画像を学習データとする。ガウスパラメータを $\sigma = 0.5, \dots, 10.0$ の 20 段階に変化させ、学習データを生成する ($N = 20$)。部分空間を作成し、固有ベクトル上位 10 個を認識に用いる。

(手法 D) 光学ぼけモデルと解像度低下モデルに基づく生成型学習法

3 種類のカメラから推定した各々の光学ぼけ PSF h_{opt} (図 3) を用い学習データを生成する。PSF 推定

表 1 PSF 推定時の被写体間距離

Table 1 Camera distance for PSF estimation.

Camera	Distance
DV camera	70 cm
Digital camera	70 cm
Phone camera	32 cm



Method A



Method B1



Method B2



Method C



Method D (Digital camera)

図 6 学習データの例。

Fig. 6 Training data for each learning method.

用として、試験データとは別の日に表 1 に示す撮影距離で撮影した 60 フレームの画像を使用した。生成時は、劣化パラメータを $d = 0.05, \dots, 1.00$ の 20 段階に変化させ、学習データ (32×32 pixel) を生成する ($N = 20$)。部分空間を作成し、固有ベクトル上位 10 個を認識に用いる。

生成された学習データの例を図 6 に示す。

6.1.1 試験用データの撮影

実験のため、表 2 に示す性能をもつ 3 種類のカメラで試験用データを撮影した。撮影者は PSF 推定時、試験用データ撮影時ともに同一の人物が行った。撮影はカメラを手で静止させ、距離を数段階に変えながら行った。撮影距離と撮影された文字の平均サイズの関係を表 3、表 4 に示す。50 回同様の実験を行い、62

表 2 実験に用いたカメラの性能

Table 2 Specification of the cameras.

Camera	Resolution	Frame rate
DV camera	720×480	30 fps
Digital camera	640×480	30 fps
Phone camera	162×220	7.5 fps

表 3 文字サイズ (デジタルビデオカメラ, デジタルカメラ)

Table 3 Size of characters (DV camera, digital camera).

Distance	22 cm	35 cm	50 cm	60 cm	70 cm
DV camera	16×16	10×10	7×7	6×6	5×5
Digital camera	17×17	13×13	10×10	9×9	8×8

表 4 文字サイズ (携帯電話カメラ)

Table 4 Size of characters (Phone camera).

Distance	20 cm	32 cm
Phone camera	7×7	5×5

文字種 \times 50 回の認識結果に対し、認識率を求めた。

認識対象は英数字 62 文字 (A ~ Z, a ~ z, 0 ~ 9: Century フォント) であり、それらが印刷された紙を広角で動画像として撮影した。文字の大きさは 1 文字当り最大 6 mm 四方であった。撮影された動画像からフレームごとに文字を自動で切り出し、試験用データとして用いた。切り出す文字領域は個々の文字領域を含む最小の正方形領域とした。このとき、全 62 文字を切り出すことができなかったフレームは抽出失敗とみなし、認識対象から除外した。なお、切り出した試験用データはニアレストネイバ法により 32×32 pixel にサイズを正規化した。

6.1.2 実験結果

3 種類のカメラで撮影した文字 (図 7) に対して認識実験を行った結果を図 8 に示す。

対象が低解像度の場合に、手法 D が他と比べて高い認識率を示した。とりわけ、光学ぼけの影響の大きい文字画像 (図 7(b): デジタルカメラ) に対して顕著な改善が見られた。このことにより、光学ぼけ PSF を推定し学習データの生成に用いることは低解像度文字の認識にとって有効であることが明らかになった。

解像度低下モデルのみを用いる手法 B は、解像度の比較的高い文字に対しては手法 D と同等の精度を示した。しかしデジタルカメラによって撮影された低解像度文字に対しては正規化手法によらず低い認識率にとどまった。また、手法 C, D の認識率の比較により、近~中距離におけるカメラのぼけ特性を近似するためには、ガウシアンよりも光学ぼけ PSF が適する

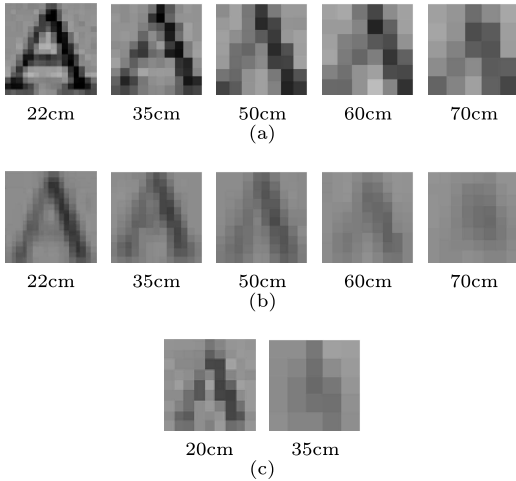


図 7 試験用データの例 (撮影距離別) (a) デジタルビデオカメラ (b) デジタルカメラ (c) 携帯電話カメラ
 Fig. 7 Test data according to the distance. (a) Digital video camera, (b) Digital camera, (c) Camera equipped in cellular phone.

という結果が得られた。

6.2 手ぶれに関する実験

生成型学習における動きぶれモデル、及びカメラ速度選別の有効性を検討する。比較のため、以下の二つの学習法を定義する。

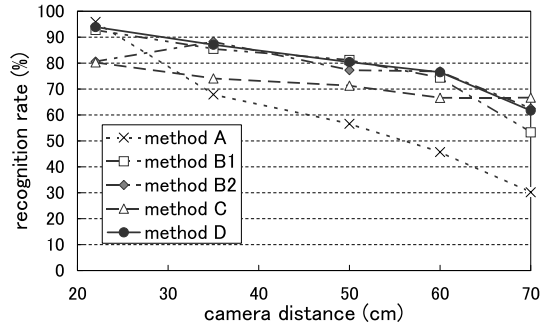
(手法 A) 光学ぼけ PSF を用いた生成型学習法

推定した光学ぼけ PSF h_{opt} を用い、劣化パラメータ d を $d = \frac{1}{240} \times 1, \dots, \frac{1}{240} \times 480$ の 480 通りに変化させ、学習データ (32 × 32 pixel) を生成する。部分空間を作成し、固有ベクトル上位 10 個を認識に用いる。

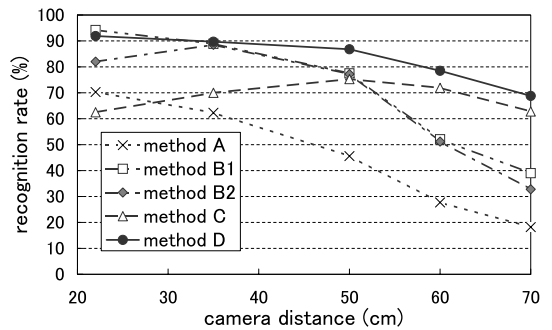
(手法 B) 光学ぼけ PSF と動きぶれ PSF を用いた生成型学習法

推定した光学ぼけ PSF h_{opt} と動きぶれ PSF $h_{mot}(b, \theta)$ を畳み込み、生成に用いる。劣化パラメータ d を $d = 0.25, \dots, 2.00$ の 8 通りに、ぶれ強度パラメータ b を $b = 0.0, \dots, 2.0$ の 5 通りに、ぶれ方向角パラメータ θ を $\theta = 0, \dots, \frac{11}{12}\pi$ の 12 通りに変化させて手法 A と同じ枚数の計 480 個の学習データを生成する (図 9)。部分空間を作成し、固有ベクトル上位 10 個を認識に用いる。

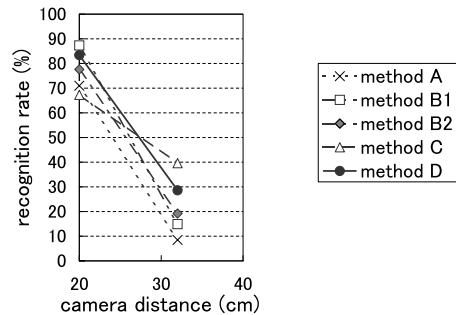
本実験においては、640 × 480 画素、30 fps で撮影が可能なデジタルカメラを用いた。光学ぼけ PSF h_{opt} は、カメラと被写体間の距離を 35 cm に保ち、カメラを静止させた状態で撮影した動画画像から推定した。



(a)



(b)



(c)

図 8 認識結果 (a) デジタルビデオカメラ (b) デジタルカメラ (c) 携帯電話カメラ

Fig. 8 Recognition results. (a) Digital video camera, (b) Digital camera, (c) Camera equipped in cellular phone.

6.2.1 試験用データの撮影

本研究の詳細を知らない 6 人の被験者に、試験用データとして用いる動画画像の撮影を依頼した。被写体は 6.1 で述べた実験と同じものを使用した。撮影時間は各人につき 20 秒とし、被写体との距離は撮影者の任意とした。カメラは片手でもつものとし、なるべく

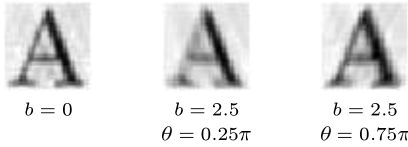


図 9 手法 B の学習データの例
Fig. 9 Training data of method B.

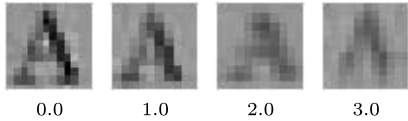


図 10 試験用データの例 (カメラ速度別)
Fig. 10 Test data by camera speed.

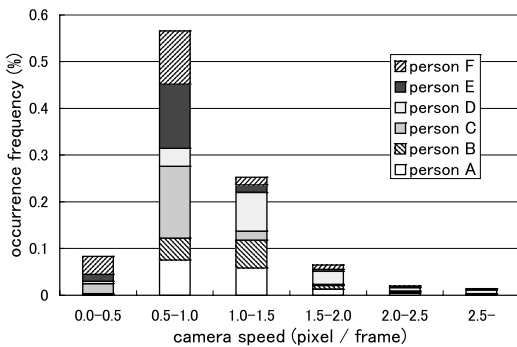


図 11 カメラ速度の分布
Fig. 11 Distribution of the camera speed.

手を静止させた状態で撮影してもらった。なお、デジタルカメラの手ぶれ防止機能は OFF とした。試験用データのサンプルを図 10 に示す。

本実験においてはぶれと認識率の関係を調べるため、式 (12) によりカメラ速度を計算し評価に用いた。

6.2.2 実験結果

揺れの発生割合を調べたグラフを図 11 に示す。被験者が意識してカメラを静止させた場合でも、0.5 [pixel/frame] 未満のカメラ速度の割合はわずが 8.3%であった。カメラ速度の割合に大きな個人差はなく、どの被験者の撮影した画像でも 0.5 [pixel/frame] 以上の揺れが定期的に発生していることが分かる。

文字の平均切出しサイズが 11 × 11 [pixel] となったフレーム群のみを対象に、カメラ速度に対する認識率の変動を手法別に比較した結果を図 12 に示す。結果によると、カメラ速度 1.0 ~ 2.5 [pixel/frame] 付近で認識率が最も高くなっている。低速時に認識率が低かった理由としては、フレーム間の画像変化が少なく、複数フレーム認識 [3] の効果が低下したことが挙げら

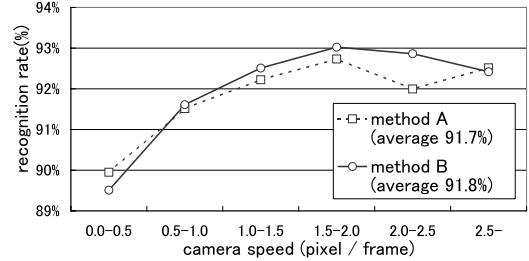


図 12 カメラ速度別の認識率 (サイズ 11 × 11 pixel) 図中 average は全区間の平均を表す
Fig. 12 Recognition rate depending on the camera speed.

れる。また、高速時の認識率低下は、ぶれによる過大な画像劣化によるものと考えられる。

手法 A, B を比較すると、動きぶれモデルを学習した手法 B は、低速時に認識率が低下した一方、カメラ速度 1.0 ~ 2.5 [pixel/frame] における認識率のピークが高くなっている。これらのカメラ速度を観測したフレーム群を選別して認識に用いることで、認識率の向上が期待できる。手法 B を用いて比較したところ、全被験者により撮影された試験用データの平均認識率は 91.8%であったが、カメラ速度 1.0 ~ 2.5 [pixel/frame] のフレームを選別した場合には 92.6%となり、精度改善が見られた。以上より、カメラ速度に基づく選別は有効であり、動きぶれモデルを用いた生成型学習法と併用することで高い効果が得られることが明らかになった。一方、観測されたカメラ速度に対し、学習データ生成時のパラメータをいかに設定するのが最適かは、今後の検討課題としたい。

7. むすび

本論文では、ぼけやぶれを含む低解像度文字認識のための生成型学習法を提案した。本学習法では、撮影によって得られた実データではなく、原画像から人工的に生成した画像を学習データとして用いる。現実の劣化要因に対応するため、光学ぼけモデル、動きぶれモデル、解像度低下モデルを定義した。光学ぼけモデルにおいては、認識に用いるカメラから PSF を推定し、それを生成に用いた。生成時のパラメータを制御することで、多様な条件を想定した学習が行えるようになる。同数の原画像を用いた生成手法の比較実験により、各々のモデルの有効性を示した。

今後の課題としては、ゆがみ、回転などに対する生成モデルの検討が挙げられる。そのためには、PSF に

限らず、より発展した生成手段を生成型学習法に導入することが望まれる。

謝辞 日ごろより熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する。本研究の一部は日本学術振興会科学研究費補助金、21世紀COEプログラム「社会情報基盤のための音声・映像の知的統合」による。

文 献

- [1] D. Doermann, J. Liang, and H. Li, "Progress in camera-based document image analysis," Proc. 5th Int. Conf. on Document Analysis and Recognition, pp.606-616, Edinburgh, Scotland, Aug. 2003.
- [2] 柳詰進介, 高橋友和, 井手一郎, 目加田慶人, 村瀬 洋, "携帯デジタルカメラにより撮影された動画画像からの低解像度文字認識," 信学論(D), vol.J89-D, no.2, pp.323-331, Feb. 2006.
- [3] S. Yanadume, Y. Mekada, I. Ide, and H. Murase, "Recognition of very low-resolution characters from motion images," Proc. PCM2004, Lecture Notes in Computer Science, vol.3331, pp.247-254, Springer-Verlag, Dec. 2004.
- [4] 村瀬 洋, "画像認識のための生成型学習法," 情処学論, vol.46, no.SIG15, pp.35-42, Oct. 2005.
- [5] 村瀬 洋, 木村文隆, 吉村ミツ, 三宅康二, "パターン整合法における特性核の改良とその手書き平仮名文字認識への応用," 信学論(D), vol.J64-D, no.3, pp.276-283, March 1981.
- [6] H. Andrew and B. Hunt, Digital image restoration, Prentice-Hall, 1977.
- [7] 橋本正一, 斎藤秀雄, "PSF パラメータ分布を推定するソフトバリアントなぼけ画像の復元法," 信学論(D-II), vol.J77-D-II, no.4, pp.719-728, April 1994.
- [8] J. Hobby and H. Baird, "Degraded character image restoration," Proc. 5th UNLV Symp. on Document Analysis and Information Retrieval, pp.233-245, Las Vegas, USA, April 1996.
- [9] H. Li and D. Doermann, "Text enhancement in digital video using multiple frame integration," Proc. 7th ACM Int. Conf. on Multimedia, pp.19-22, Orlando, Florida, Nov. 1999.
- [10] C. Mancas-Thillou and M. Mirmehdi, "Super-resolution text using the teager filter," Proc. 1st Int. Workshop on Camera-Based Document Analysis and Recognition, pp.10-16, Seoul, South Korea, Aug. 2005.
- [11] Y. Yitzhaky and N. Kopeika, "Identification of blur parameters from motion blurred images," Graphical Models and Image Processing, vol.59, no.5, pp.310-320, Sept. 1997.
- [12] 網島宣浩, 中島真人, "コンバウンド法を用いた PSF の推定とぼけ画像の復元," 信学論(D-II), vol.J81-D-II, no.11, pp.2688-2692, Nov. 1998.
- [13] M. Potmesil, "Modeling motion blur in computer-generated images," Comput. Graph., vol.17, no.3,

pp.389-399, July 1983.

(平成 17 年 11 月 14 日受付, 18 年 3 月 24 日再受付)



石田 皓之 (学生員)

平 14 名大・工・情報卒。現在, 同大学院情報科学研究科メディア科学専攻博士後期課程在学中。



高橋 友和 (正員)

平 9 茨城大・工・情報卒。平 12 同大学院理工学研究科博士前期課程了。平 15 同研究科博士後期課程了。博士(工学)。同年より 2 年間名古屋大学大学院情報科学研究科 COE 研究員。平 17 より日本学術振興会特別研究員, 現在に至る。三次元画像処理並びに三次元モデリングの基礎研究に興味をもつ。画像電子学会会員。



井手 一郎 (正員)

平 6 東大・工・電子卒。平 8 同大学院工学系研究科情報工学専攻修士課程了。平 12 同研究科電気工学専攻博士課程了。博士(工学)。同年国立情報学研究所助手。平 16 より名古屋大学大学院情報科学研究科助教授, 情報・システム研究機構国立情報学研究所客員助教授(連携)兼任。この間, 平 14~16 総合研究大学院大学数物科学研究科助手併任, 平 17, 18 フランス情報学・統計システム研究所(IRISA)招聘助教授。映像メディア処理全般に興味をもっている。平 7 第 51 回情報処理学会全国大会奨励賞受賞。人工知能学会, 情報処理学会, IEEE Computer Society, ACM 各会員。



目加田慶人 (正員)

平 3 名大・工・情報卒。平 8 同大学院博士後期課程了。平 8 宇都宮大学工学部情報工学科助手。平 13 名古屋大学大学院工学研究科情報工学専攻助教授, 平 15 同大学院情報科学研究科助教授, 平 16 より中京大学生命システム工学部教授, 現在に至る。画像処理・パターン認識とその医用応用, 信号処理などの研究に従事。博士(工学)。平 10 日本医用画像工学会論文賞, 2002 FPGA/PLD Design Conference 審査委員特別賞, MIRU2004 優秀論文賞。IEEE, 日本医用画像工学会, コンピュータ支援画像診断学会, 胸部 CT 検診研究会各会員。



村瀬 洋 (正員)

昭 53 名大・工・電気卒．昭 55 同大学院修士課程了．同年日本電信電話公社（現 NTT）入社．平 4 から 1 年間米国コロンビア大客員研究員．平 15 から名古屋大学大学院情報科学研究科教授．現在に至る．文字・図形認識，コンピュータビジョン，マルチメディア認識の研究に従事．工博．昭 60 本会学術奨励賞，平 6 IEEE-CVPR 最優秀論文賞，平 7 情報処理学会山下記念研究賞，平 8 IEEE-ICRA 最優秀ビデオ賞，平 13 高柳記念奨励賞，平 13 本会ソサエティ論文賞，平 14 本会業績賞，平 15 文部科学大臣賞，平 16 IEEE Trans. MM 論文賞，他受賞．情報処理学会，IEEE 各会員．