

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

石田 翔之[†] 柳詰 進介[†] 目加田慶人^{††} 井手 一郎[†] 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 愛知県豊田市貝津町床立 101

E-mail: [†]{hishi,yanadume}@murase.nuie.nagoya-u.ac.jp, ^{††}{mekada,ide,murase}@is.nagoya-u.ac.jp

あらまし 我々は、低解像度文字を認識するため、複数の画像フレームからの情報を部分空間法で統合する手法を開発している。これは学習段階において、文字画像のパターン特徴を近似する部分空間を作成し、これらと入力された複数フレームのパターンとの類似度により文字を認識する手法である。部分空間法で文字の判別分類を行なうには、文字画像パターンの学習が必要である。従来、文字そのものを撮影した画像を用いて学習していたが、異なる環境や撮影装置に対応するためには、さまざまな条件のもとで撮影された画像が必要であった。本報告では様々な条件下での学習用画像を生成する方法について述べる。この学習法では、撮影画像から推定した劣化モデルをもとに学習サンプルを生成し、学習に用いる。実験により、提案する学習法の有効性を示した。

キーワード 文字認識、低解像度、部分空間法、動画像、学習法

Generative Learning Method for the Recognition of Low-Resolution Characters Using the Subspace Method

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

[†] Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan
^{††} School of Life System Science and Technology, Chukyo University
101 Tokodate, Kaizu-cho, Toyota, Aichi, 470-0393 Japan

E-mail: [†]{hishi,yanadume}@murase.nuie.nagoya-u.ac.jp, ^{††}{mekada,ide,murase}@is.nagoya-u.ac.jp

Abstract We are developing a recognition system for low-resolution characters using the Subspace method. This method identifies target characters by comparing the similarity between the characters and each subspace, which approximates the patterns of a character. Learning degraded characters is necessary for the recognition by the Subspace method. Former methods made use of characters captured by a camera, which requires us to collect characters of all categories in various conditions. In this report, we propose a new learning method, which generates the degraded characters for learning by a point spread function estimated beforehand by captured images. We confirmed the usefulness of our method by experiments.

Key words Character Recognition, Low Resolution, Subspace Method, Movie, Learning Method

1. はじめに

近年、デジタルカメラ、カメラ付き携帯電話等のデジタル映像機器が普及している。それらの機器で撮影した画像の認識理解は、画像によるユーザインタフェスの改善や監視システムへの応用が考えられる。カメラに用いられるセンサの高精度化、撮影画像を蓄積するメモリの小型大容量化や低価格化に伴い、画像の品質は大きく向上した。しかし、最新の機器でも十分な品質の画像を得るのは依然として困難な場合が多い。文字は一文字ごとの大きさが小さく、近距離で撮影しなければ高解像度の画像とはならないためである。しかし、カメラの撮影者は、撮影される文字の解像度や適切な撮影距離まで気を遣わないと、認識するのに十分な品質が得られない場合が多い。他にもカメラを持つ手のゆれによる画質の劣化、センサの解像度に伴う劣化などが認識を困難にしている原因として存在する。

我々は、低解像度文字の認識法として複数画像フレームからの情報を部分空間を用いて累積的に利用する手法を開発している。一般に部分空間法 [1][2] は、学習と認識の 2 段階からなり、学習段階では多数の学習サンプルのパターンを近似する固有ベクトルを作成する。認識では学習段階で作成した部分空間に認識対象文字のパターンを射影し、学習パターンとの類似度を求め、その類似度を基準に対象文字をカテゴリに分類する。多くの文字の特徴をより低い次元で表現した部分空間を作成することによって、入力画像にみられる膨大な数のパターンへの対応が可能であることが部分空間法の利点である。

この部分空間法で文字の分類を行なうには、学習サンプルが重要な役割を担う。本報告では、撮影画像から劣化関数 [3] を推定し、それをもとに学習サンプルを生成する手法を提案する。本稿ではカメラから推定した劣化関数を用いて学習サンプルを自動生成することが、部分空間法による低解像度文字の認識に有效であることを示す。

2. 劣化過程

学習サンプルを生成するためには、現実に起こる劣化の特性を知る必要がある。本章では撮影による文字の劣化過程について述べる。

認識対象となる劣化文字画像の例を図 1 に示す。図は印刷文字をデジタルビデオカメラで撮影したものであるが、文字の大きさが小さく、撮影時の条件による劣化を受けているため、計算機による識別が困難になっている。文字が劣化する要因にはさまざまなものがある。文字がカメラのレンズを通してセンサ上に写像されるとき、まずレンズが原因となる光学的なぼけが生じる。また、センサによって光情報をサンプリングする過程で、解像度の低下がおこる(図 2)。このとき動画像中の連続するフレームであっても、手ぶれによって各センサに入力される光量は変化するため、毎回異なる劣化画像が得られる(図 3)。

他にも、露光中にカメラが移動することで発生するぼけや、カメラ内部で行われる手ぶれ補正処理、高解像度化の処理など、さまざまな要因が画像に作用する。

これらの劣化を劣化要因別に分離して考えることは困難であ

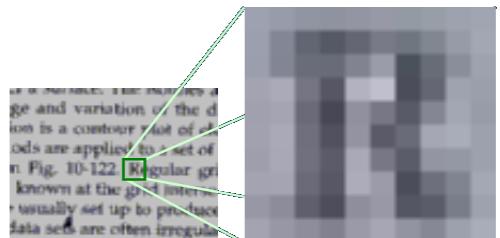


図 1 劣化画像の例



図 2 劣化過程

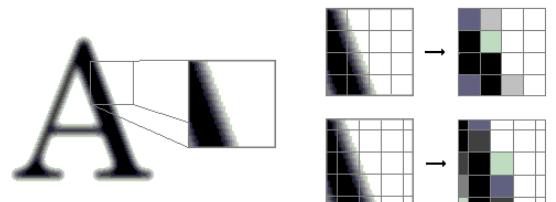


図 3 手ぶれによる画像の変化

るため、本手法では、これらの劣化過程を一体化して扱う。本研究では、2 つの劣化モデルについて検討した。第 1 の劣化モデルは解像度を低くすることによる劣化を生成するモデルである。これを劣化モデル B と呼ぶ。第 2 の劣化モデルは、解像度の低下を含むさまざまな劣化要因によって、原画像が劣化画像となるに至るまでの伝達関数を、一つの劣化関数で表すものである(これを劣化モデル C と呼ぶ)。劣化関数とは、画質劣化の過程における伝達関数を意味するものである。動画像を元に、連続する多数のフレームを利用することで、撮影状況やカメラに対応した劣化関数を推定する。これらの劣化モデル B、および劣化モデル C を用いて劣化のシミュレーションを行う。なお、便宜上原画像をそのまま出力するモデルを劣化モデル A と呼ぶ。

3. 動画像を用いた部分空間法による文字認識

我々は、部分空間法を用い、動画像中の複数フレームを用いて文字を認識する手法の研究を行ってきた [4]。部分空間法は、学習サンプルのパターンを近似する固有ベクトルを作成し、分類に用いる。従来法では、学習サンプルは撮影によって得た文字画像を用いていたが、以下の章で提案する学習法では、推定した劣化関数をもとに様々な劣化画像を計算機上で作成する。従来法も提案手法も、固有ベクトルの計算法としては同じであるが、使用する学習サンプルが異なる。

以下より、部分空間法による認識手法を示す。文字のカテゴリ数を M 、カテゴリ毎に用意する学習サンプルの数を N 、認識に用いる固有ベクトルの数を L とする。このときカテゴリ

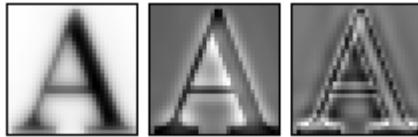


図 4 固有ベクトル (上位 3 個)

m ($m = 0, \dots, M - 1$) 中の n ($n = 0, \dots, N - 1$) 番目の学習サンプル画像を、ラスタスキャン方式でベクトル化し, $\mathbf{x}_{m,n}$ と表す。これらのベクトルを 1 列に並べ、その行列の自己相関行列 \mathbf{X}_m を次式 (1) で求める。

$$\mathbf{X}_m = [\mathbf{x}_{m,0} \ \dots \ \mathbf{x}_{m,N-1}] [\mathbf{x}_{m,0} \ \dots \ \mathbf{x}_{m,N-1}]^t \quad (1)$$

この自己相関行列の固有ベクトル展開を行う。 \mathbf{X}_m の固有値を求め、固有値の大きい順にそれぞれ対応する L ($\leq N$) 個の固有ベクトル $\mathbf{u}_{m,l}$ ($l = 0, \dots, L - 1$) を用いる(図 4)。

認識対象の入力文字画像を部分空間に射影することによって類似度を計算し、最大の類似度を与えるカテゴリに分類する。なお、我々の手法では認識対象文字を撮影した動画像から認識を行う。動画像を用いる理由は、1 枚の画像だけからでは識別が困難であっても、動画像中の多数のフレームを用いることによって、劣化文字に見られる文字情報のあいまい性が軽減され、認識率が向上することが期待されるからである。画像復元の手法としては、複数の低解像度画像から高解像度画像を復元するもの [5][6] が存在する。文字認識においても、手ぶれによる微妙な位置変化を含んだ動画像から複数フレームを抽出して用いることで、1 枚の低解像度画像から得られる以上の情報量を得る。

具体的な計算式は次のようになる。 j フレーム目の画像から認識対象の文字を切り出し、学習サンプルと同じ大きさになるように正規化した画像を用意する。それを学習サンプルと同じくラスタスキャンによってベクトル化したものを \mathbf{y}_j とする。 F フレーム分の入力文字画像の情報を累積的に使用するため、カテゴリ m の文字パターンとの類似度を次式で定義する。

$$s_m = \sum_{j=1}^F \sum_{i=0}^{L-1} (\mathbf{u}_{m,i} \cdot \mathbf{y}_j)^2 \quad (2)$$

全てのカテゴリに対して、入力文字との類似度を計算し、最大の類似度を与えるカテゴリを入力文字の属するカテゴリとする。なお F は認識に使用するフレーム数である。 $F = 1$ のとき、单一フレームからの認識、つまり静止画からの認識と等価になる。

4. 生成型学習法

4.1 概 要

生成型学習法は、劣化関数を用いて観測されるであろう劣化画像を多数生成して、これを学習サンプルとして学習する手法である。ここでは 2 種類の劣化モデルを検討した。第 1 の劣化モデル (劣化モデル B) は、画像を解像度変換して、これを学習サンプルとして学習する手法である。劣化モデル B を使用する手法は実験の方法 B で述べることにする。第 2 の劣化モデル

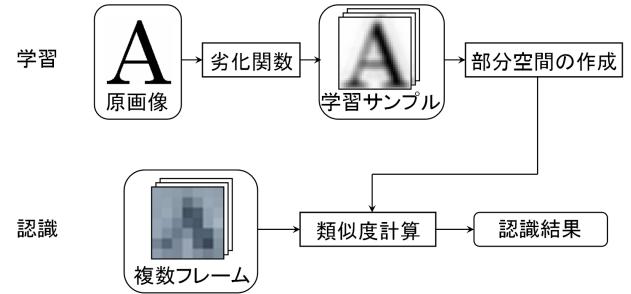


図 5 生成型学習法による認識までの流れ

(劣化モデル C) は、原画像と観測される画像から観測系の伝達関数を推定し、これにより学習サンプルを多数生成する手法である(図 5)。ここでは画像の伝達関数を使用する。その方法について以下、詳細に説明する。

4.2 準 備

本手法では、学習サンプル生成の準備として、はじめに劣化関数推定用の動画像を撮影する。劣化関数の推定に使用する原画像(モノクロ 2 値画像)を計算機上で作成し、印刷する。印刷した用紙を認識に用いるカメラで撮影し、劣化画像を得る。撮影は手でビデオカメラを持ち、被写体との距離を一定に保って行う。

4.3 劣化関数の推定

劣化関数の推定にはコンパウンド法 [7] 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劣化関数を推定する手法である。

本手法では、作成した原画像と、動画中の時間的連続な複数フレームの劣化画像を用いる。劣化画像は原画像とは縮尺が異なるため、拡大および位置合わせを行なう必要がある。画像劣化のモデルは、 f を原画像、 g を劣化画像、 h を劣化の点広がり関数 (PSF : Point Spread Function)、 n を加法雑音として次のように表される。

$$g(x, y) = f(x, y) * h(x, y) + n(x, y) \quad (3)$$

このモデルでは、原画像と劣化関数のたたみ込み (*) によって劣化画像が生成されるとしている。雑音は加法的であることを仮定している。

ここで求めるのは劣化関数 h であるため、まず次のように(3) 式に 2 次元フーリエ変換を施す。

$$H(u, v) = \frac{G(u, v)}{F(u, v)} - \frac{N(u, v)}{F(u, v)} \quad (4)$$

ここで雑音成分 N は未知である。撮影により得た劣化画像は雑音成分を含むため、1 枚の画像のみから適当な劣化関数を求ることは困難である。そこで本手法では動画像中の複数の画像について $H(u, v)$ の平均を求ることによって雑音を抑制する。 k 枚の劣化画像を用いるものとすると、 $H(u, v)$ は次のようになる。

$$H(u, v) = \frac{1}{k} \sum_{i=1}^k \frac{G_i(u, v)}{F(u, v)} - \frac{1}{k} \sum_{i=1}^k \frac{N_i(u, v)}{F(u, v)} \quad (5)$$

ここで、雑音成分は画像間で無相関であると仮定し、

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \frac{N_i(u, v)}{F(u, v)} = 0 \quad (6)$$

とする。以上より、劣化画像の枚数 k が十分大きい数であるとき、 $H(u, v)$ は次の式に近似できる。

$$H(u, v) = \frac{1}{F(u, v)} \frac{1}{k} \sum_{i=1}^k G_i(u, v) \quad (7)$$

最後に逆フーリエ変換を施して $h(x, y)$ を得、これを PSF として実験に用いる。

図 6, 7, 8 に、デジタルビデオカメラ、デジタルカメラ、携帯電話カメラから求めた劣化関数を示す。グラフは、推定した劣化関数 h を、中心点 $(0, 0)$ からの距離を横軸にしてプロットしたものである。

4.4 学習サンプルの生成

求めた劣化関数をもとに、劣化学習サンプル画像を作成する。学習サンプルの作成には $h(x, y)$ をそのまま劣化のない文字画像にたたみこむ方法も考えられるが、本手法ではさまざまな強さの劣化に対応するため、劣化強度 (D 段階) という指標を導入し、一つの劣化関数、一つの文字画像から、さまざまな劣化の強さの画像を作成する。

そこで、図 9 のフィルタサイズ $(R_h + 1) \times (R_h + 1)$ の劣化関数フィルタを用いる。

計算量削減のため、劣化画像の生成に必要な範囲 $[-\frac{R_h}{2}, \frac{R_h}{2}]$ のみを用いる。このフィルタを用い、原文字画像を空間フィルタリングすることによって人工的な劣化画像が作られる。ただし通常のフィルタ処理とは異なり、対象画像の画素単位で積和演算を行うのではない。原文字画像上に生成する学習サンプル画像の画素に対応する格子点を配置し、それらの点を中心とした範囲に、劣化強度に合わせて伸張したフィルタを適用し、フィルタの重みと、その位置に対応する原文字画像中の濃度値を取得し、積和演算を行って学習サンプル画像の濃度値を決定する。

各文字カテゴリにつき、劣化強度を段階的に変化させながら、劣化の強さがさまざまな劣化画像を作成する。その具体式を次に示す。

$$g_{(d)}(p, q) = \sum_{j=-R_h/2}^{R_h/2} \sum_{i=-R_h/2}^{R_h/2} h(p-i, q-j) \times f(x(p, q) + d \frac{R_F}{R_G} \frac{i}{R_h}, y(p, q) + d \frac{R_F}{R_G} \frac{j}{R_h}) \quad (8)$$

g : 生成する学習サンプル画像

R_G : 学習サンプル画像の大きさ

d : 劣化強度

f : 原文字画像

R_F : 原文字画像の大きさ

x : 原文字画像中における格子点の水平位置

y : 原文字画像中における格子点の垂直位置

h : 劣化関数

R_h : 劣化関数フィルタの大きさ

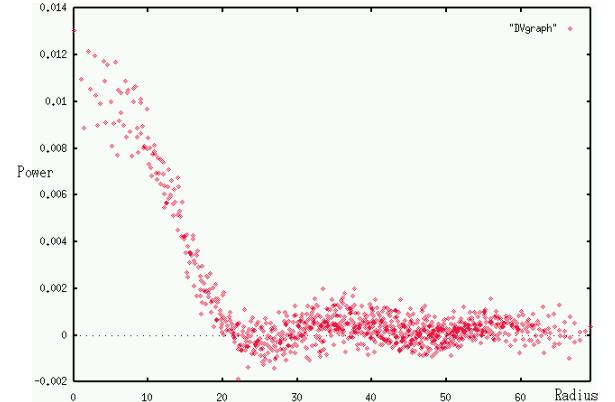


図 6 デジタルビデオカメラの劣化関数

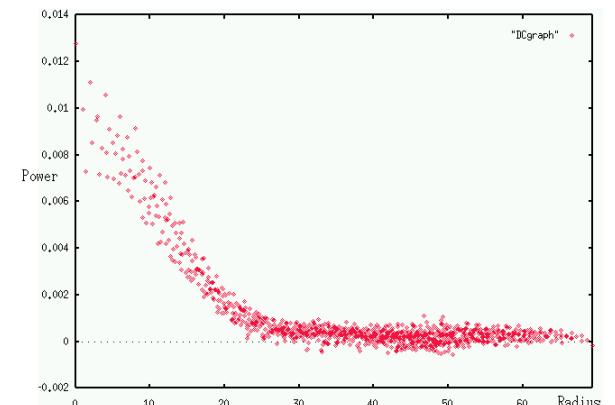


図 7 デジタルカメラの劣化関数

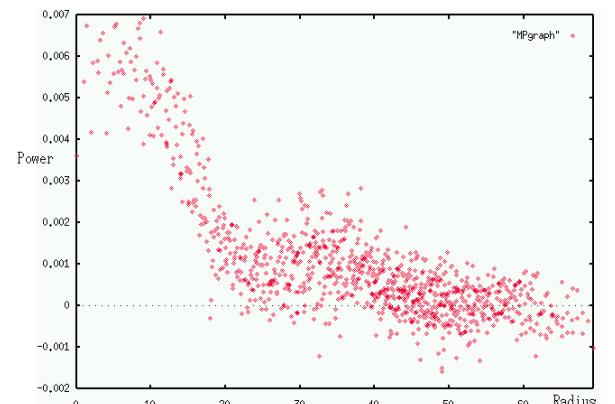


図 8 動画機能つき携帯電話のカメラの劣化関数

$h(-\frac{R_h}{2}, -\frac{R_h}{2})$	\dots	$h(0, -\frac{R_h}{2})$	\dots	$h(\frac{R_h}{2}, -\frac{R_h}{2})$
\vdots		\vdots		\vdots
$h(-\frac{R_h}{2}, 0)$	\dots	$h(0, 0)$	\dots	$h(\frac{R_h}{2}, 0)$
\vdots		\vdots		\vdots
$h(-\frac{R_h}{2}, \frac{R_h}{2})$	\dots	$h(0, \frac{R_h}{2})$	\dots	$h(\frac{R_h}{2}, \frac{R_h}{2})$

図 9 劣化関数フィルタ

格子点の数は生成される学習サンプル画像の画素数に等しい。これを全ての文字 M 個について行い、合計 $M \times D$ 個の学習用サンプル画像を作成する。

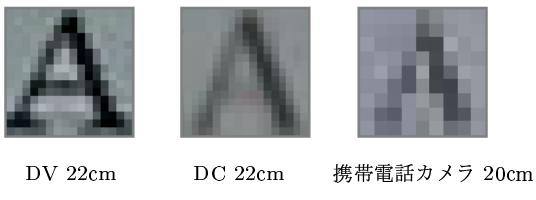


図 10 認識対象の文字画像

表 1 文字サイズ (pixel) 対応表 (DV, DC)

距離	22cm	35cm	50cm	59cm	71cm
DV	16 × 16	10 × 10	7 × 7	6 × 6	5 × 5
DC	17 × 17	13 × 13	10 × 10	9 × 9	8 × 8

表 2 文字サイズ (pixel) 対応表 (携帯電話)

距離	20cm	32cm
携帯電話カメラ	7 × 7	5 × 5

5. 実験

5.1 実験方法

提案する学習法の有効性を示すために、比較実験を行った。この実験では劣化関数を用いて学習した場合と、用いなかった場合の認識結果を比較した。

デジタルビデオカメラ (DV), デジタルカメラ (DC), 携帯電話カメラを用いてアルファベットの大文字・小文字・数字 (A ~ Z, a ~ z, 0 ~ 9) の 62 文字 (フォント: Century) を含んだ画像を印刷し、その紙を撮影した。認識対象の文字画像の一部を図 10 に示す。

カメラと被写体との距離をさまざまに変えて認識対象文字を撮影した。得られた動画像を静止画として取り出し、一つ一つの文字を切り出した。切り出しには、入力画像を判別分析法により 2 値化し、雑音除去をした後、1 文字を含む最小の正方形領域として切り出し、 $32 \times 32\text{pixel}$ の大きさに正規化した。なお、撮影時の被写体との距離と、各距離における文字の正規化処理前の大きさとの対応は表 1, 2 のようになった。撮影対象の文字の紙面上での大きさはほぼ 1cm 平方であった。

実験では以下に示す 3 通りの学習法を比較した。(B), (C) が提案の学習法である。

(A) 劣化サンプルを作成せずに、文字画像 ($32 \times 32\text{ pixel}$) 1 枚を学習データとして認識する学習法 (図 11)。

(B) 劣化モデル B: $128 \times 128\text{pixel}$ の文字画像を $8 \times 8 \sim 32 \times 32$ の 25 通りの大きさに縮小した後、 32×32 に再び拡大して 25 個の学習パターンを作成し、認識に用いる学習法。固有ベクトルは上位 10 個を用いる。この学習法では、カメラの劣化特性は考慮されていない (図 12)。

(C) 劣化モデル C: 認識にはそれぞれのカメラの特性を用いる。室内で、適当な距離 (デジタルビデオカメラとデジタル



図 11 手法 A の学習サンプル「A」(全カメラ共通)



図 12 手法 B の学習サンプル「A」の一部 (全カメラ共通)

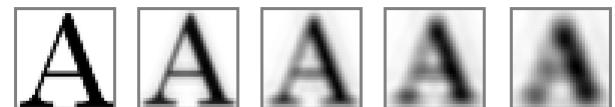


図 13 手法 C の学習サンプル「A」の一部 (デジタルビデオカメラ)

カメラでは 71cm、携帯電話では 20cm とした) の距離から撮影した動画像から劣化関数を求める。原文字画像のサイズを $128 \times 128\text{pixel}$ 、生成する学習サンプルのサイズは、 $32 \times 32\text{pixel}$ とした。つまりフィルタリング処理時の格子点も 32×32 個である。生成する学習サンプル画像は、1 カテゴリにつき 20 段階の劣化強度で劣化させた (図 13)。

5.2 実験結果

デジタルビデオカメラ、デジタルカメラ、携帯電話で撮影した文字に対して認識実験を行なった結果を以下の図 14, 15, 16 に示す。アルファベットの大文字・小文字・数字の 62 文字すべてに対して認識を行い、認識率の平均を求めた。認識は動画像中の連続する 10 フレームを使用した。計 50 回同様の認識実験を行い、全体の認識率の平均を求め、認識結果とした。

5.3 結果の考察

学習法 A は、撮影による文字の劣化を考慮に入れない学習法であり、劣化文字の認識に適切でないばかりか、多数の学習サンプルの劣化パターンを近似するという部分空間法の目的に合わないものである。一方の学習法 B は解像度の低下をシミュレートするものであるが、学習法 C と異なり、その他の劣化要因やカメラの劣化特性を考慮しない。実験によって、文字が低解像度であるほど、学習法 C の有効性が際立つことが明らかとなつた。また、デジタルビデオカメラと比較して、デジタルカメラの方が学習法 C による認識率の向上が大きくみられた。この差は次のように解釈できる。デジタルビデオカメラの撮影画像と比べて、デジタルカメラの撮影画像は、エッジのはっきりしないぼけた画像となっている (図 10)。この違いは、推定された劣化関数のグラフからも見ることができる。デジタルカメラの劣化関数 (図 7) にはデジタルビデオカメラの劣化関数 (図 6) にみられるリップルが存在しない。結果として境界線の不明瞭な文字画像が撮影されている。今回の実験ではデジタルカメラによる撮影画像の方が低品質で認識の難しい画像であるといえるが、そのような悪条件の場合において、劣化モデル C を用いる手法が適していると考えることができる。

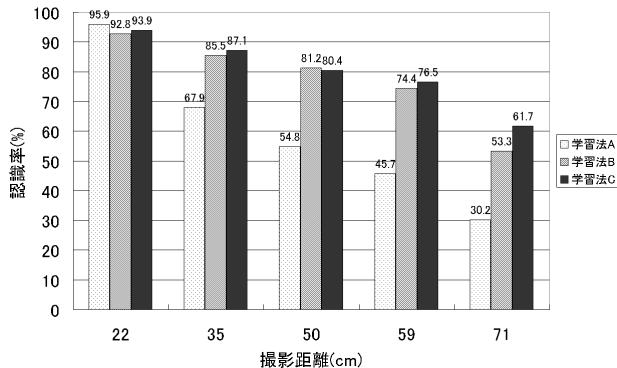


図 14 デジタルビデオカメラ実験結果

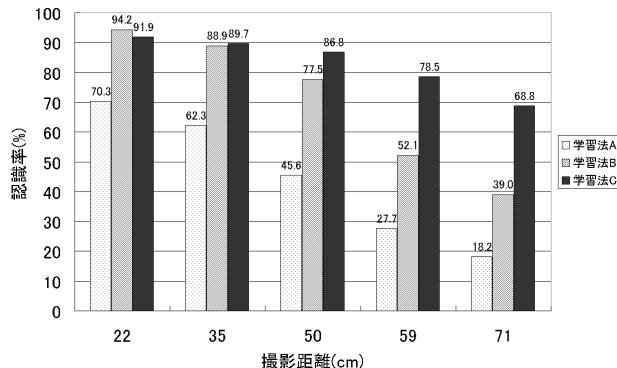


図 15 デジタルカメラ実験結果

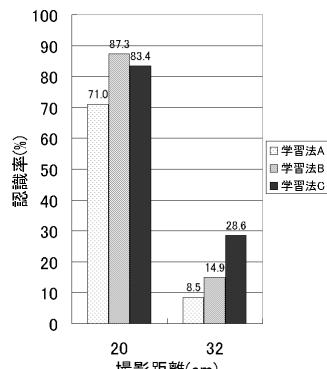


図 16 携帯電話カメラ実験結果

6. ま と め

本論文では、様々な要因を含む劣化画像の自動生成による学習サンプルに基づく低解像度文字認識の検討を行った。生成型学習法では、撮影によって得られた実データではなく、撮影によって得られた画像から劣化を推定することにより、任意のカテゴリに対して様々な劣化強度の学習データを生成することができる。認識手法には部分空間法を用い、自動生成した学習サンプルによる部分空間と認識対象の時系列画像との類似度により文字認識を行った。劣化モデル C では、デジタルビデオカメラ、デジタルカメラ、携帯電話カメラを対象として劣化関数を推定し、劣化画像を生成して認識実験を行った。画像劣化を考えない学習、生成型学習法(劣化モデル B、劣化モデル C)による学習との結果を比較したところ、特に認識対象の空間解像度が低い場合に

劣化モデル C が有効であることを明らかにした。例えばデジタルカメラを用いた認識実験(撮影距離 71cm)においては、画像劣化を考えない手法による認識率 18.2%に対して、劣化モデル C では 68.8%と大幅に向上了。本手法では安定した撮影条件下で劣化関数を推定する必要はあるが、種々の劣化画像を作成できることの利点は十分に示すことができた。今後の課題としては、方向性のあるぶれに対応するために劣化関数を近似し、より柔軟性のある劣化シミュレーションを行うことなどが挙げられる。

文 献

- [1] 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, pp276-283, March, 1981.
- [2] 田村秀行、村瀬洋、松山隆司、山本裕之、"コンピュータ画像処理", オーム社出版局, 2002.
- [3] 橋本正一、斎藤秀雄、"PSF パラメータ分布を推定するシフトパリエントなぼけ画像の復元法", 信学論(D-II), Vol.J77-D-II, no.4, pp719-728, April, 1994.
- [4] 柳詰進介、目加田慶人、村瀬洋、"携帯デジタルカメラによる動画像を用いた低解像度文字の認識", 2004 年度電子通信学会総合全国大会講演論文集, Vol.3, pp197, March, 2004.
- [5] 青木伸、"複数のデジタル画像データによる超解像処理", Ricoh Technical Report, Vol.24, pp19-25, November, 1998.
- [6] 藤本浩司、藤田和弘、吉田靖夫、"画像確率モデルに基づく複数の劣化画像からの復元", 信学論(D-II), Vol.J82-D-II, No.5, pp863-871, May, 1999.
- [7] 綱島宣浩、中島真人、"コンパウンド法を用いた PSF の推定とぼけ画像の復元", 信学論(D-II), Vol.J81-D-II, No.11, pp2688-2692, November, 1998.