携帯カメラによる動画像を用いた低解像度文字の認識手法

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　〒464-8603 愛知県名古屋市千種区不老町
E-mail： yanadume@murase.nuie.nagoya-u.ac.jp

あらまし　本稿では，携帯デジタルカメラにより撮影された文字画像の認識手法を提案する．近年，デジタルビデオカメラやカメラ付き携帯電話といった，動画像が容易に撮影可能なデジタル撮影機器が急激に普及している．しかし，これらの機器で撮影した画像は常に品質の良いものとは限らず，また，手ぶれ等の影響で画像がぶれてしまうことがある．そのため，一枚の画像のみからこのような低品質の文字を認識するのは困難である．提案する手法は入力に動画像の複数フレームの画像を用いて，各画像の情報を部分空間法により統合し，認識精度を向上させるものである．デジタルビデオカメラとカメラ付き携帯電話を用いた実験により本手法の有効性を確認した．

キーワード　文字認識，部分空間法，低解像度，動画像

Recognition of very low-resolution characters from motion images captured by a portable digital camera

Shinsuke Yanadume　　Yoshito Mekada*　　Ichiro Ide　　Hiroshi Murase
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8603 Japan

Abstract　Many kinds of digital devices can easily take motion images such as digital video cameras or camera-equipped cellular phones. If an image is taken with such devices under everyday situations, the resolution is not always high; moreover, hand vibration can cause blurring, making accurate recognition of characters from such poor images difficult. This paper presents a new character recognition algorithm for very low-resolution video data. The proposed method uses multi-frame images to integrate information from each image based on a subspace method. Experimental results using a DV camera and a phone camera show that our method improves recognition accuracy.

Keyword　Character recognition，Subspace method，Low resolution，Motion images

1. は じ め に

近年，デジタルビデオカメラ（以下 DV カメラ）やカメラ付き携帯電話といった携帯デジタル撮影機器が比較的安価に手に入るようになり，これらの機器を日常で携帯する機会が増加している．もし，これらの機器で撮影した画像からの文字の自動認識が可能となれば，掲示板中のテキスト認識や，雑誌などに書かれた URL を携帯電話で撮影し入力するシステムといった，マンマシンインタフェースのための有用な技術になると考えられる．

これまでにも多くの文字認識手法が提案されてきた[1]．しかし，携帯デジタル撮影機器を用いて一度にドキュメント全体を撮影した場合，1 文字あたりの画像の大きさはとても小さくなってしまい，十分な品質を得ることが困難となる．また，手ぶれやカメラのレンズによる劣化等も撮影文字画像の品質を低下させる原因となる．そのため，単一の画像からこのような低品質の文字画像を認識するのは困難である．

そこで本研究では，動画像を用いて低解像度の文字を認識する手法を提案する．動画像から複数フレームの画像情報を用いることができれば[2]，これらの低品質の文字を認識する際に有用な手段になると考えられる．これまでにも動画像を用いて低解像度の文字を認識する手法は提案されてきた．小佐井らは，移動差分手法によりフレーム間の文字画像の差分を取り，それによって得られたエッジ特徴により文字を認識する手法を提案した[3]．しかし，この手法では，認識に用いるエッジ特徴の抽出のために，撮影された動画像を構成する複数フレームから，文字が特定の方向に移動した画像を選択する必要があった．

そこで本手法では，複数の画像情報の統合に部分空間法[4]～[8]を用いた．この手法では，低解像度の文字や手ぶれによる画像のぶれといった入力を想定した画像を作成し，これらの学習データから部分空間を作成する．そして，いくつかの方向に少しずつ移動させて撮影された複数枚の文字画像と部分空間とのマッチングを行うことにより，従来手法のように文字の移動方向を指定することなく認識することが可能である．福井らは顔画像の認識において，複数フレームの入力画像から作成した部分空間と辞書データである部分空間とのマッチングにより認識を行った[8]．我々は，複数フレームの入力画像と辞書データとの類似度の平均を認識に使用

* 現在，中京大学生命システム工学部

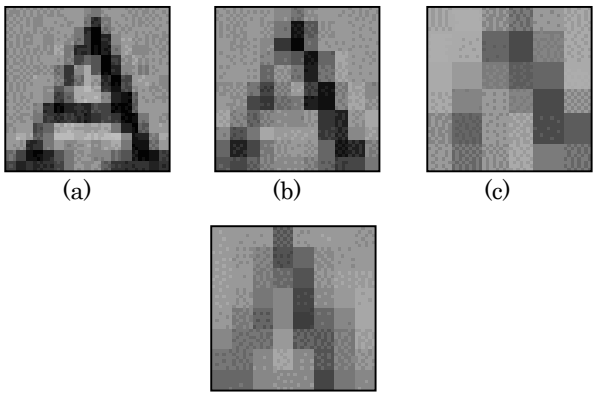


図 1　撮影文字画像‘A’の例.（a），(b)，(c)： DV カメラ（紙とカメラとの距離を変化）(d)：カメラ付き携帯電話

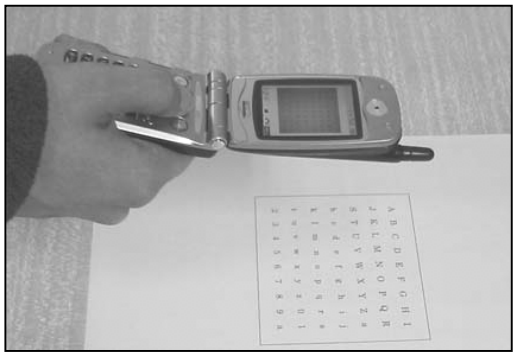


図 2　カメラ付き携帯電話による文字の撮影

することで，複数枚の画像情報を利用する．

本稿では，2 章で携帯デジタルカメラによる撮影文字画像の特徴，3 章で提案するアルゴリズム，4 章で実験結果を記述し，5 章でまとめる．

2.　動画像中の文字画像

2.1　携帯デジタルカメラ

図 1 に携帯デジタルカメラで紙に印刷された文字を撮影したときの文字画像の例を示す．図 1 の(a)，(b)，(c)は DV カメラで紙とカメラとの距離を変化させて撮影したときの文字画像，(d)はカメラ付き携帯電話で撮影したときの文字画像である．一般的に，携帯デジタルカメラを使って一度に全てのドキュメントを撮影すると（図 2），各文字は図 1 の(c)や(d)のような低解像度のものになってしまうため，単一の画像だけからでは，これらの文字を認識することは困難である．動画像から得られた複数フレームの画像情報を利用することで，このような超低品質な文字画像の認識を行うことを我々の目的とする．

2.2　動画像における文字画像の変化

カメラ付き携帯電話や DV カメラのようなハンディカメラで撮影した場合，手ぶれによりカメラの位置を固定すること

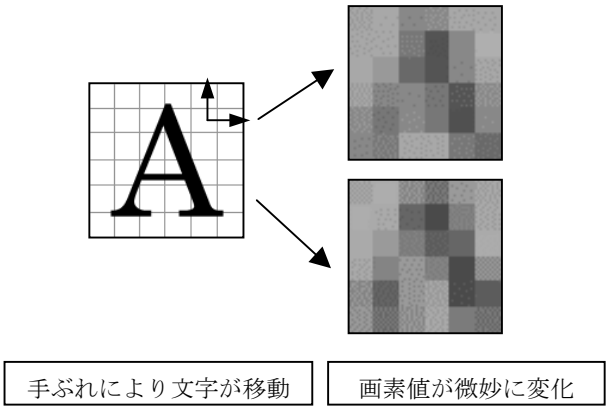


図 3　手ぶれによる文字画像の変化

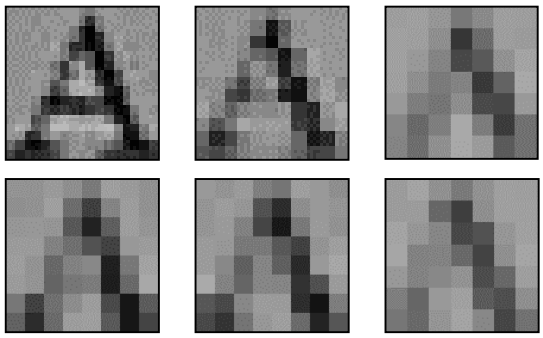


図 4　学習データの例（文字“A”）

が困難となる．そのため，これらの機器で撮影した文字の動画像は微妙に異なるたくさんの画像から構成されていることがわかる．もし，これらの複数枚の画像情報を適切に利用することができれば，単一の画像だけからでは認識が不可能な低解像度の文字に対しても，認識が可能になると考えられる．図 3 に DV カメラによって撮影された文字“A”の二つのフレーム画像を示す．従来の文字認識手法を用いた場合，単一の画像だけからでは，このような低解像度の文字を認識することは困難である．しかし，この 2 枚の画像の情報をうまく用いることができれば，画像間の微妙な画素値の違いは，認識の精度を向上させるための大きな手助けになると考えられる．

3.　アルゴリズム

本手法は，学習データの作成，部分空間の作成，入力データからの文字の認識の 3 段階からなる．学習データの作成は，高い認識率を達成するための重要な段階である．本手法では，各カテゴリで入力を想定した様々なパターンの文字画像を作成して学習データとすることにより，認識率の向上を目指す．部分空間の作成では，作成した学習データを用いて固有ベクトルを計算し，辞書データとする．認識では，動画像から得られた複数枚の画像を入力とすることにより，手ぶれによる画像情報の変化を利用する．

3.1　学習データの作成

提案する手法では，認識対象の特性をある程度限定し，それと似た状況で学習データを収集することが重要である．あまりに厳しい制約を科すと環境の変化に対するロバスト性が低下し，ゆるい制約の場合には学習結果のクラス内分散が大きくなるため，精度の低下を引き起こす．カメラを手で持ちながら撮影するという条件での文字認識を考えるため，文字とレンズが平行に近い状態で撮影し，文字の回転やレンズの傾きによる歪みが少ないという制約条件で実験を行った．また，認識対象の文字は以下のようなものとした．

- 印刷文字
- アルファベット（大文字・小文字），アラビア数字
- 1 文字の画像サイズが 6×6 pixel 以上

我々は，学習データに動画像から得られた複数枚の画像を使用することで，手ぶれによる文字画像の変化を学習した．また一般に，入力される文字画像の大きさを事前に知ることはできない．そこで，文字が印刷された紙とカメラとの距離を変化させて撮影することで，様々な大きさの文字画像を作成し，大きさを正規化したものを学習データとした．図 4 に本手法で使用する学習データの例を示す．この例のように，低解像度，かつ同程度の解像度の文字でも画素値が微妙に異なる複数枚の画像の使用により，手ぶれによる文字画像の変化や低解像度の文字画像を学習するためのデータとなる．

3.2　部分空間の作成

部分空間法は，学習データの集合からそれらの分布を近似する部分空間と未知入力データとの類似度を用いて認識を行う手法である．ある学習画像 i に対して，これを 1 次元のベクトルにラスタスキャンし，さらに正準化（平均値 0 の単位ベクトルへの正規化）を行ったベクトルを

$$\boldsymbol{x_i} = [x_1, x_2, \cdots, x_N]^T$$

で表す．ここで Nは画像の画素数である．次に，前節で生成した k 枚の学習データを適当に並べた行列を

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k]$$

で表し，それに対して自己相関行列

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}$$

を求める．各カテゴリに対して，この $\boldsymbol{Q}$ の固有値を大きい順に R個計算し，これに対応する固有ベクトルを求める．ここでカテゴリ cの固有ベクトルの集合を以下のようにおく．

$$\{\boldsymbol{e}_1^{(c)}, \boldsymbol{e}_2^{(c)}, \cdots, \boldsymbol{e}_R^{(c)}\}$$

本手法ではこれらの固有ベクトルの集合を辞書データとする．図 5 は実際に作成した学習データから計算し，画像化した固有ベクトルの例である．この図から，複数の解像度で学習したことによる文字のボケや，手ぶれによるエッジの移動が確認できる．

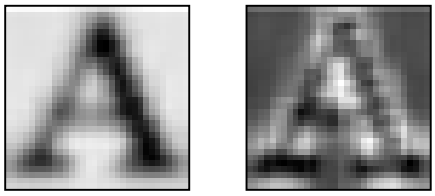


図 5　画像化した文字“A”の固有ベクトル．
左：第 1 固有ベクトル　右：第 2 固有ベクトル

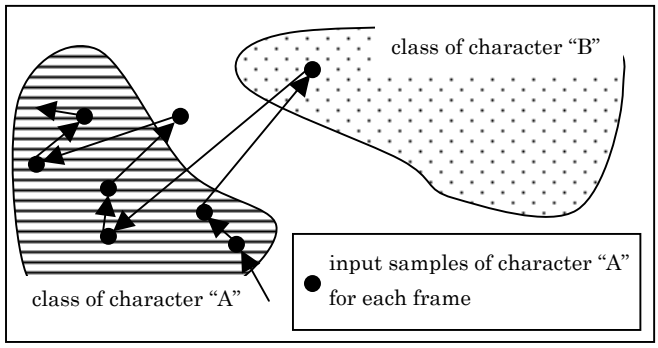


図 6　複数フレーム入力の効果

3.3　認識方法

動画像から得られた入力画像列から文字の切り出し，大きさの正規化，および部分空間の作成段階と同様の正準化処理を行った画像ベクトルの集合を

$$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_M\}$$

とする．ここで $\boldsymbol{a}_m$は入力画像ベクトル，Mは入力フレーム数を表す．そして，入力画像列とカテゴリ cとの類似度を

$$L^{(c)}(\boldsymbol{a}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{r=1}^{R}(\boldsymbol{a}_m, \boldsymbol{e}_r^{(c)})^2$$

と定義する．ただし$(\boldsymbol{x}, \boldsymbol{y})$ はベクトルの内積を表す．この類似度を，作成した辞書データの全てのカテゴリに対して計算し，その値が最大となるカテゴリを認識結果とする．これにより，図 6 のように，あるフレームで正解とは異なるクラスにより近い入力が与えられたとしても，複数フレームに対する平均類似度を用いることで，正解を出力することが期待できる．

4.　実験と考察

本手法の有効性を確認するために，携帯デジタルカメラ（DV カメラ，カメラ付き携帯電話）で撮影した文字画像を用いて，学習データの作成と認識実験を行った．学習した文字は Century フォントの数字とアルファベットの大文字と小文字（計 62 カテゴリ）であり，これらの文字を紙に印刷したうえで、撮影した．使用したフォントを図 7 に示す．ここで得られた動画像からフレーム毎に文字を切り出し，それぞれ別の日に撮影したものを学習データと認識実験に用いた．また，切り出された文字画像は，類似度の計算時におけるベクトルの次元数統一のために，全て同じ大きさに正規化

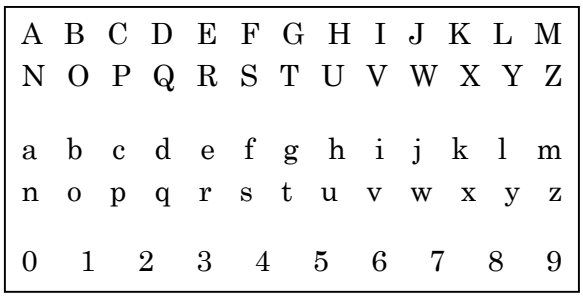


図 7　使用フォント（Century）

する．この際，解像度の高い文字画像の入力に対しても，認識結果に影響を及ぼさない程度で，かつ類似度計算における計算量を考慮した文字サイズに正規化する必要がある．以下の実験では学習データ，評価データにおける全ての文字画像の正規化文字サイズを 32×32 pixel とした．

類似度の計算において使用する固有ベクトルの数（次元数）の選択は，認識結果に大きな影響を及ぼすと考えられる．そこで予備実験として，辞書データに使用する固有ベクトルの次元数に対する認識率の変化を調べた．実験に使用したデータは以下の通りである．

学習データ

- DV カメラで撮影
- 文字が印刷された紙とカメラとの距離を変化させることで，文字サイズを調整 （最大距離 70cm）．
- 平均文字サイズ（**size1**：16×16，**size2**：11×11，**size3**：8×8，**size4**：7×7，**size5**：6×6 pixel）
- 各カテゴリに複数フレーム使用（各文字サイズ：10 フレーム， 計：50 フレーム）

評価データ

- DV カメラで撮影
- 学習データとは異なる日に撮影した画像
- 撮影距離（平均文字サイズ）
  - **size5**：70cm （約 6×6 pixel）
- 入力フレーム数：20 フレーム
- 各文字 32×32 pixel に大きさを正規化
- サンプル数:各文字 30 セット

結果を図 8 に示す．実験結果から，固有ベクトルの次元数を増やすことで認識率が向上し，また次元数が 3 付近で向上が飽和していることがわかる．この結果から，辞書データに使用する固有ベクトルの次元数を，認識精度と計算時間を考慮して，今後の実験では次元数 5 に固定することにする．

4.1　入力フレーム数による認識率の変化

低解像度文字の認識に対する本手法の有効性を確認するために，認識に用いる入力画像のフレーム数と文字サイズを変化させて実験を行った．実験に使用したデータは以下の通

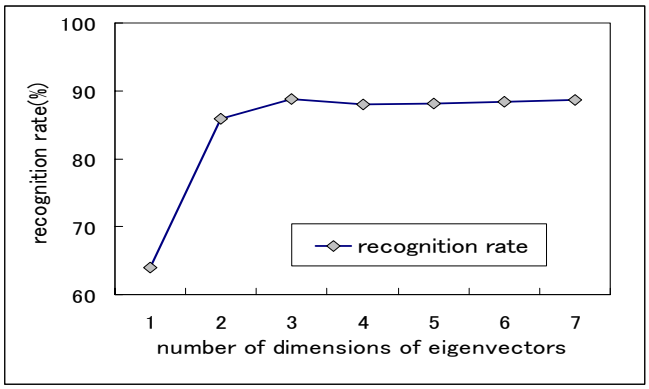


図 8　固有ベクトルの次元数に対する認識率の変化

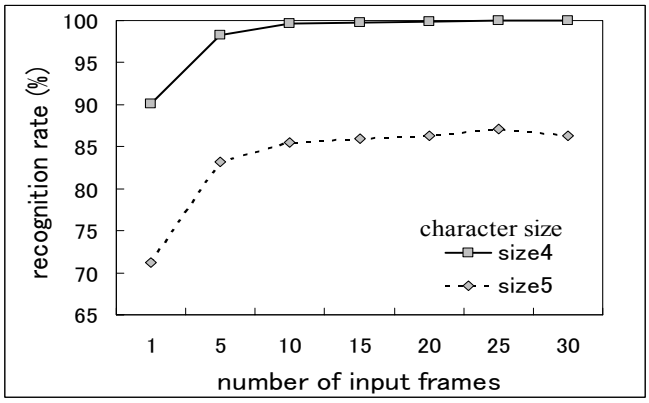


図 9　入力フレーム数による認識率の変化．（**size3** 以上の文字サイズの画像に対しては，ほぼ 100%の認識率）

りである．また学習データは前記の予備実験に使用したものと同一のものである．

辞書データ

- 固有ベクトル 5 個

入力データ

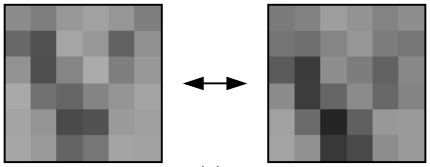
- DV カメラで撮影
- 学習データとは異なる日に撮影した画像
- 撮影距離（平均文字サイズ）
  - **size4**：60cm （約 7×7 pixel）
  - **size5**：70cm （約 6×6 pixel）
- 各文字 32×32 pixel に大きさを正規化
- サンプル数:各文字 30 セット

4.1.1.　フレーム数 vs. 認識率

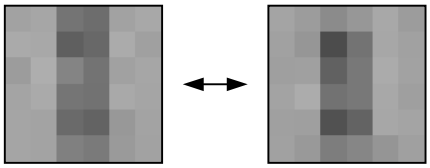
入力フレーム数に対する認識率の変化を図 9 に示す．実験結果から，入力フレーム数を増やすことで認識率が向上していることがわかる．これは，本手法が複数の画像情報を累積的に利用できているためと考えられる．また入力フレーム数が 15 付近で認識率の増加が停滞していることがわかる．このことから，ある程度の数のフレームを入力することで，十分な画像情報が取得できていることがわかる．平均文字サイズが 7×7 pixel の入力に関しては，ほぼ 100%の認識率とな

表 1　認識誤りの多かった文字（文字サイズ：**size5**．入力フレーム数：20．各カテゴリの入力サンプル数：30）

入力文字	認識結果	発生回数	認識率（%）
S	s	30	0
V	v	30	0
Z	z	30	0
l	1	30	0
0	o	30	0
8	s	30	0
I	1	23	23
	I	7	
X	x	22	26
	X	8	



(1)



(2)

図 10　認識誤りの多かった文字の例．
（1）左：“V”，右：“v”　（2）左：“l”，右：“1”

った．この結果，本手法は低解像度文字の入力に対して有効であることが示された．なお，**size3** 以上の文字サイズの入力に対しては少数の入力フレーム数に対してもほぼ 100%の認識率となった．

4.1.2. 認識誤りの多かった文字

次に，今回の実験において認識誤りの多かった文字の例を表 1 に示す．これは，入力の平均文字サイズを 6×6 pixel，入力フレーム数を 20 枚とした時に認識誤りが多かった文字の例である．この結果から，アルファベットの大文字を，同種の小文字と間違えている場合が多く見られることがわかる．“V”や“S”等では，大文字と小文字の形状が本質的に同一であり，それらの文字間の違いが微少であるため，認識を誤ってしまったものと考えられる（図 10（1））．しかし，アルファベットの大文字と小文字の間には意味の違いはなく，実際の文字認識システムの構築を考えた場合，単語中や文中でのその文字の位置関係を用いることで，この問題は解消すると考えられる．

また，“l”や“8”のように，全く別の種類の文字として認識されてしまった場合もいくつか見られた．これらの文字に関しても，元々文字の形状が類似していることから，今回の実験で用いたもののように，文字が低解像度の画像である場合，両者の間の違いが非常に微少になってしまう事が原因であると考えられる（図 10（2））．

表 2　撮影機器の違いによる認識率の変化．（平均文字サイズ：7×7 pixel．DV カメラにより学習）

撮影機器	認識率(%)
DV カメラ	99.9
カメラ付携帯電話	94.4

4.2　撮影機器の違いによる認識率の変化

現在，デジタルカメラやカメラ付き携帯電話など，個人が所有する携帯デジタル撮影機器の種類は数多く存在する．そのため，実際に携帯デジタル撮影機器で文字を撮影して，認識しようとする場合，必ずしも学習で使用した撮影機器と同じ機器で撮影された画像が入力されるとは限らない．そこで我々は，学習データと入力データ間での撮影機器の違いから生じる画像の劣化の違いが認識に与える影響を調べるために，学習データとして DV カメラで撮影した画像を，入力データとしてカメラ付き携帯電話で撮影した画像を用いて認識実験を行った．一般的に，カメラ付き携帯電話で撮影した画像は DV カメラで撮影したものと比べて，品質が低いものとなってしまう．

学習データと辞書データは前節と同じものを使用した．実験に使用したカメラ付き携帯電話と入力データの仕様は以下の通りである．

カメラ付き携帯電話の仕様

- CCD 有効画素数：31 万画素
- 撮影画像サイズ：162×220 pixel
- フレームレート：7.5 fps

入力データ

- 撮影距離：20 cm
- 平均文字サイズ：7×7 pixel
- 各文字を 32×32 pixel に正規化
- 入力フレーム数：20 フレーム
- サンプル数：各文字 30 セット

結果を表 2 に示す．比較のために，同程度の文字サイズで DV カメラを用いて撮影した画像を入力としたときの認識率も同時に示す．実験結果を見ると，DV カメラで撮影した画像を入力とした場合に比べて，カメラ付き携帯電話で撮影した場合の認識率が低下していることがわかる．これは，カメラ付き携帯電話で撮影した画像は DV カメラで撮影したものと比べて，品質が低く，かつ劣化の過程も異なるものであることが原因であると考えられる．また，入力画像は，学習デ**ータ**

表 3 照明条件の違いによる認識率の変化（文字サイズ：**size5** ）.

照明条件	認識率(%)
明るい	86.3
普通	82.6
暗い	86.4

とは異なる機器で撮影された画像であるため，カメラの把持の容易さやレンズの特性の違いも影響していることも考えられる.

4.3　照明条件の違いによる認識率の変化

多くのコンピュータビジョンシステムにおいて，照明の変動は深刻な問題である．そこで，本手法において，文字撮影時の照明条件の違いが認識率の変化に及ぼす影響を確認するために，ある照明条件下で撮影した画像を学習データとし，照明条件を変えて撮影した画像を入力したときの認識率を調査した．学習と認識に使用した撮影機器は DV カメラである．学習データは“明るい照明”下で撮影された画像であり，学習データと辞書データにおける残りの条件は 4.1 節のものと同一である．入力データの仕様を以下に示す.

入力データ

- 学習データとは別の日に撮影した画像
- 文字サイズ：**size5**
- 照明条件：“明るい”，“普通”，“暗い”の 3 種類
- 入力フレーム数：20 フレーム
- サンプル数：各文字 30 セット

結果を表 3 に示す．実験結果から，認識率は撮影時の照明条件に依存しないことがわかる．これは文字を認識する際，入力文字画像を平均値が 0 の単位ベクトルに正規化することで，明るさの変動を吸収できているためであると考えられる．このことから，本手法は照明の変動をある程度吸収可能であると言える.

5.　む す び

本研究では動画像を用いた部分空間法による低解像度文字の認識手法を提案した。携帯デジタル撮影機器で多くのドキュメントを一度に撮影した場合，各撮影文字画像は非常に低解像度なものになってしまい，1 枚の画像だけからでは認識が困難である．本手法では，動画像から切り出された複数枚の画像を入力とすることにより，手ぶれ等による文字画像の微妙な変化を含めて学習した．実験により，学習時と同じような条件で撮影された画像に対しては，1 枚の画像だけからでは認識が困難な場合でも，複数フレームの画像を使用することで，認識が可能となることを確認した．また本手法は，入力画像に対して明るさを正規化することにより，撮影時の明るさの変動をある程度吸収できることが，実験により確認できた．実際の文字認識システムを構築した場合，学習で使用した撮影機器と入力に用いる撮影機器との違いが認識率に影響を及ぼすと考えられる．実験により，DV カメラにより学習して作成した辞書データを用いて，カメラ付き携帯電話で撮影した入力文字画像を認識した場合，認識率の低下が見られたことから，学習の際に各カメラがもつ固有の画像化過程をモデル化する必要性が確認できた.

今後の課題としては，日本語や異なるフォントなど，より多くの文字種への対応や，撮影機器ごとのカメラの撮像モデルに対応した学習データの作成などが考えられる.

謝 辞

日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏，特に撮影にご協力頂いた石田皓之氏に感謝する．本研究の一部は文部科学省 21 世紀 COE プログラムおよび科研費補助金（No.16300054）による.

参考文献

[1] S. Mori, K. Yamamoto and M. Yasuda, “Research on machine recognition of handprinted characters,” IEEE Trans. PAMI, vol.PAMI-6, no. 4, pp.386-405, July 1984.

[2] P. Cheeseman, B. Kanefsky, R. Hanson and J. Stutuz, “Super-resolved surface reconstruction from multiple images,” Tech. Rep. FA-94-12, NASA Ames Research Center, Artificial Intelligence Branch, October 1994.

[3] 小佐井潤，加藤邦人，山本和彦，“ビデオカメラを用いた低解像度文字認識，”映像情報メディア学会誌 vol.53, no. 6，pp.867～872，June 1999.

[4] E. Oja, “Subspace methods of pattern recognition, ” UK：Research Studies Press, 1983.

[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. Murase, S. K. Nayar, “Visual Learning and Recognition of 3-D Objects from Appearance,” International Journal of Computer Vision, vol. 14, pp.5-24，1995.

[7] 大町真一郎，阿曽弘具，“低品質文字認識におけるつぶれを動的に補正する部分空間法，”信学論(D-II)，vol.J82-D-II，no. 11，pp.1930-1939，November 1999

[8] 福井 和広, 山口 修, 鈴木 薫, 前田 賢一，“制約相互部分空間法を用いた環境変動にロバストな顔画像認識，”信学論(D-II)，vol.J82-D-II. no. 4，pp.613-620，April 1999.