

携帯デジタルカメラにより撮影された動画からの低解像度文字認識

柳 詰 進介^{†*} 高橋 友和[†] 井手 一郎[†] 目加田慶人^{††}
 村瀬 洋[†]

Recognition of Very Low-Resolution Characters from Motion Images Taken by Portable Digital Cameras

Shinsuke YANADUME^{†*}, Tomokazu TAKAHASHI[†], Ichiro IDE[†], Yoshito MEKADA^{††}, and Hiroshi MURASE[†]

あらまし 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、動画を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。これらの機器により撮影された画像を用いて文字を認識する技術は、マン・マシンインタフェースにおいて有用な要素技術となる。しかし、これらの撮影機器で紙面等に印刷された文書全体を一度に撮影した場合、各文字は非常に低解像度なものになってしまい、また、手ぶれなどの影響で画像がぶれてしまうことがある。そのため、このような低品質の文字を1枚の画像のみから認識するのは困難である。本論文では、動画中の複数フレームの画像を入力に用いて、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多入力型ロバスト部分空間法を提案する。本手法は認識時に、複数枚の入力サンプルの分布情報を用いて、サンプルに対して重みを与えることで、誤認識を誘発する例外サンプルの影響を抑制し、認識精度の向上を図るものである。デジタルビデオカメラとカメラ付き携帯電話を用いた文字認識実験により、動画から得られた連続した複数の入力画像を用いることで、認識率の向上と入力サンプルへの重み付けの有効性を確認した。キーワード 文字認識, 低解像度, 動画, ロバスト性

1. ま え が き

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的に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文字の自動認識が可能となれば、掲示板や書籍等に書かれたテキストの認識や、認識された URL を入力するシステムなどといった、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。

文字認識を目的としてテキストを撮影した場合、撮

影対象の文字列を絞り込んで被写体に対して接写すれば [1]、カメラ付き携帯電話などの撮影解像度のあまり高くない機器でも、ある程度の品質の文字画像が取得できる。しかし、文書全体の文字を認識するには局所的な撮影と認識を繰り返す必要があり、モザイク処理により撮影された複数枚の静止画像を結合して文字を認識する技術はあるものの [2]、この作業は煩雑なものとなる。より使いやすいインタフェースのためには、文書全体を一度に撮影し認識できることが望ましい。従来より多くの文字認識手法が提案されてきた [3]。しかし、携帯デジタル撮影機器を用いて一度に文書全体を撮影した場合、1文字当りの画像の大きさはとても小さくなる。また、手ぶれやカメラのレンズを原因とする劣化等も含め、十分な品質の文字画像を得ることが困難となる。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。ここで、文字を動画として撮影できれば、様々なバリエーションの文字画像が画像列として取得でき

[†] 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, Kaizu-cho, Toyota-shi, 470-0393 Japan

* 現在 (株)リコー

ると考えられる．このように多様な画像情報を効果的に用いることができれば，単一の画像だけからでは認識が困難であった低品質の文字画像に対しても，認識が可能になることが期待される．

そこで本研究では，動画像を用いて低解像度の文字を高精度に認識する手法を提案する．Cheesemanらは複数枚の低解像度の画像から，より解像度の高い画像1枚を生成する技術を提案した[4]．このような高解像度画像の再構成技術は，文字の認識にも用いられてきた[2]．また，単一の低解像度の文字画像から補間処理により高解像度の文字画像を作成し，認識する技術も提案されている[5]．しかし，これらの画像を再構成する技術は多量の計算時間を要するため，実時間での認識が求められるようなシステムへの適用は不向きである．小佐井らは動画像を用いて低解像度の文字を認識する技術として，移動差分手法によりフレーム間の文字画像の差分により得られるエッジ情報を用いて文字を認識する手法を提案した[6]．しかし，この手法では，認識に用いるエッジ情報の抽出のために，撮影された複数のフレームから文字の移動方向を推定する必要がある．

そこで本研究では複数フレームの画像を直接認識に用い，複数の画像情報の統合に部分空間法[7]～[11]を用いる．部分空間法では，解像度の低下や手ぶれによる画像の変化などの入力を想定した画像から，これらの学習データを近似する部分空間を各文字について作成し，動画像から得られた複数枚の文字画像とこれらの部分空間とをマッチングすることにより，文字を認識する．福井らは顔画像の認識において，相互部分空間法による認識を行った[12]．この手法では複数の入力画像から部分空間を作成し，辞書データとなる部分空間とマッチングするため，認識時の計算コストが高くなる．本研究では，福井らの手法のように入力画像に対して部分空間を計算することなく，複数フレームの入力画像と辞書データとの類似度を積算することで，複数枚の画像情報を利用する．

動画像から得られた複数枚の入力サンプルを認識に用いる場合，カメラのぶれやノイズ等の影響を強く受けたサンプルが存在し，それらの入力サンプルが認識に悪影響を及ぼす可能性がある．そのため，複数フレームの入力サンプルから，認識に有利なサンプル，または誤認識を招く例外サンプルを選定する必要がある．仙田らは，辞書データとの類似度やレンズの特性による劣化の程度を考慮し，認識に最も有利であろう

入力サンプル1枚を用いることで例外サンプルの影響を抑制した[1]．しかしこの手法は，辞書データの性能に強く依存し，また，誤ったカテゴリーの辞書データに非常に近いサンプルが与えられると誤認識が発生する可能性がある．連続して撮影された動画像から得られる複数の入力サンプルに関して，同一カテゴリーに属するサンプルの分布には，その分布中心から離れたサンプルほど例外サンプルである可能性が高く，誤認識を招きやすいという特徴がある．提案する手法では，動画像から得られた複数の入力サンプルの分布中心をロバストに推定し[13]，[14]，その分布中心からの距離に応じて各入力サンプルに重みを与え，認識の指標である類似度に寄与させることで，辞書データに依存しない入力サンプルの選定を実現した．

2. 動画像中の文字画像

ここでは動画像で撮影された文字画像の特徴及び，文字認識において複数枚の入力文字画像を取り扱う際の問題点について述べる．なお，本論文において入力サンプルと入力(文字)画像は同じものを指す．

2.1 携帯デジタルカメラ

図1にカメラ付き携帯電話で紙面に印刷された文字を撮影している様子と撮影された文字画像の例を示す．図1(a)のように，一般的に，携帯デジタルカメラを用いて一度に文書全体を撮影すると，各文字は図1(c)のような低解像度のものになってしまう．そのため，単一の撮影文字画像だけでは，これらの文字を認識することは困難である．

2.2 複数フレームの文字認識への利用

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する場合，手ぶれによりカメラの位置を固定することが困難であり，被写体に対してカメラの

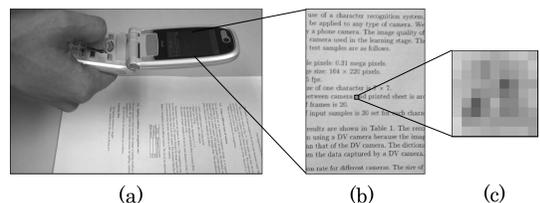


図1 カメラ付き携帯電話による文字の撮影
(a) カメラ付き携帯電話による撮影シーン (b) 撮影文書画像 (c) 切り出された文字画像“a”

Fig. 1 Taking document image with a phone camera.
(a) Taking an image with a phone camera.
(b) Captured document image.
(c) Segmented image of character “a.”

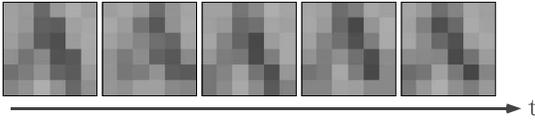


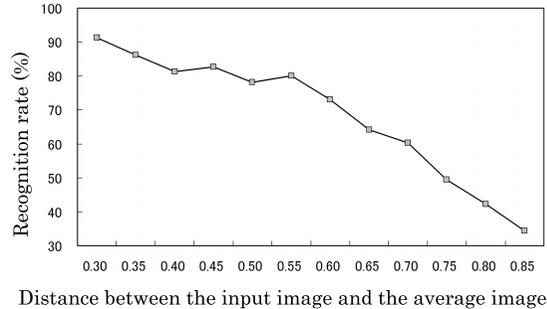
図 2 手ぶれによる文字画像の変化 (文字 “A”)
Fig. 2 Changes in pixel values due to hand motion.

わずかな平行移動や傾き、回転などが生じる。そのため、これらの機器で撮影した文字の動画像は、同じ文字であっても、各フレームで微妙に異なる。これらの複数枚の画像情報を文字の判別のために適切に利用可能ならば、単一の画像だけからでは認識が困難な低解像度の文字の認識が可能になると考えられる。図 2 に DV カメラによって撮影された文字 “A” のフレーム画像列を示す。この図は各フレーム中の文字画像が、同一のカテゴリに属するものであっても、微妙に画素値が変化している様子を表している。ここで示したような画像間の微妙な画素値の違いは、認識精度の向上のための大きな手助けになると考えられる。

しかし、動画像として撮影された文字を認識に用いる場合、あるフレームにおいて誤認識を強く誘発する例外サンプルが入力されると、それらの例外サンプルは認識精度を低下させる大きな要因となる。多数の入力サンプルを認識段階において累積的に用いることで、例外サンプルの影響をある程度軽減できると考えられるが、すべての入力サンプルを平等に扱う方法では、依然、その影響は無視できない。このような誤認識を強く誘発する例外サンプルに対して、よりロバストな認識手法が必要である。

3. 入力サンプルの選定

動画像や個別に撮影した複数の画像の使用により、複数の入力サンプルがオブジェクトの認識に使用可能な場合、入力サンプルの集合から認識に有利なサンプルを選定することが非常に重要である。辞書データとの類似度を指標として認識に使用する入力サンプルを選択する手法も考えられるが、この場合、辞書データの性能に強く依存する問題がある。一方、入力サンプル集合がもつ情報のみから、サンプルの選定が可能ならば、辞書データに依存することなく、入力サンプルの選定が可能となる。このように辞書データに独立な入力サンプル選定法は、認識対象とするカテゴリーの増加や変更に対してロバストなため、有用であると考える。



Distance between the input image and the average image

図 3 平均文字画像からの距離に対する各サンプルの認識率

Fig. 3 Recognition rates for the Euclidean distance between the average image and the input image.

そこで本論文では、複数の入力サンプルから、入力サンプルの分布の平均モデルである平均文字画像を推定し、推定された平均文字画像を基準として入力サンプルに重みを付け、辞書データとの類似度を計算する手法を提案する。以下、3.1 で平均文字画像からの入力文字画像の距離に対する認識率の関係を、3.2 で特異な入力サンプルへの対応について述べる。

3.1 平均から外れた文字画像に関する予備実験

図 3 に、ある連続した入力サンプルから作成した平均文字画像と各入力文字画像とのユークリッド距離と、その距離をもつ一つの入力サンプルを部分空間法により認識したときの認識率の関係を示す。入力サンプルの分布中心である平均文字画像から離れるに従って、認識率が減少することが分かる。この結果により、前述したように、分布の中心から外れた入力サンプルほど、例外的なノイズを受けた、誤認識を誘発する可能性が高い例外サンプルであることが確認できた。

3.2 特異な入力サンプルへの対応

2.2 で述べたように、複数の入力サンプルを均等に類似度に寄与させる手法では、特異なサンプルの入力は、最終的に累積された類似度に影響を及ぼし、認識率の低下を招くことが考えられる。そこで、本研究では複数の入力サンプルを用いた認識にロバスト推定 [13], [14] の考えを導入する。ロバスト推定は、あるサンプルの分布から、その分布を近似するモデルを推定する際に、サンプルの外れ値の影響を軽減させるモデル推定法である。

本手法では、複数の入力サンプルを用いて辞書データとの類似度を計算する際、このロバスト推定の考えに基づき例外サンプルの影響を軽減する新たな文字認

識手法を構築する．3.1において，入力サンプルの分布中心（平均文字画像）から離れたサンプルほど，認識誤りが多いことを確認した．そこで本手法では，入力サンプルの集合から推定した平均文字画像から離れたサンプルほど認識に悪影響を及ぼす例外サンプルとみなし，類似度計算の際に各入力サンプルに平均文字画像からの距離に基づく重みを付与することで，例外サンプルの影響を抑える．本手法の特徴は，1．動画画像中の複数の連続した入力サンプルを累積的に認識に用いることによる多数の画像情報の利用，2．携帯デジタルカメラで撮影された動画画像の性質とロバスト推定の考えを効率的に統合した例外サンプルの影響の軽減，の二つである．本論文では，この手法を多入力型ロバスト部分空間法と呼ぶ．次章で提案手法の詳しいアルゴリズムについて説明する．

4. 辞書データの作成

辞書データの作成は，学習データの作成，部分空間の作成の2段階からなる．

4.1 学習データの作成

提案する手法では，認識対象となる入力画像の撮影状況にある程度制約し，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．過度に厳しい制約は環境の変化に対する認識のロバスト性が低下し，緩い制約では学習結果のクラス内分散が大きくなるため，認識精度の低下を招く．ユーザが携帯カメラを手で持ちながら文字の認識を想定して動画画像を撮影する場合，シャッターボタンを押す際の大きなぶれは想定されるものの，撮影中の極端なカメラの傾き，平行移動，回転は起きないと考えられる．そこで本論文では，文字とレンズがほぼ正対した状態で撮影し，回転やレンズの傾きによる文字画像のひずみが少ないという制約のもとで実験を行った．また，認識対象の文字は以下のようなものを用いた．

- 印刷文字
- 対象カテゴリー
 - アルファベット（大文字・小文字）
 - アラビア数字
- 文字サイズ：6×6 pixel 以上

学習データは，携帯カメラで撮影された印刷文字の画像からなる．本手法では，学習データに動画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は知ることができない．そ

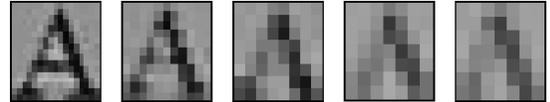


図4 学習データの例（文字“A”）
Fig.4 Excerpt from the training data “A.”

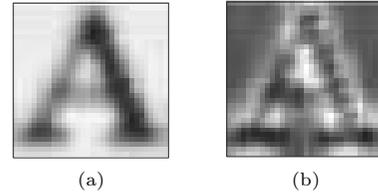


図5 画像化した文字“A”の固有ベクトル
(a) 第1固有ベクトル (b) 第2固有ベクトル
Fig.5 Picturized eigenvector of “A.” (a) The first eigenvector. (b) The second eigenvector.

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，線形補間でリサンプリングすることにより画像の大きさを正規化したものを学習データとして用意した [15]．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，手ぶれなどにより画素値が微妙に異なる複数枚の画像を使用することにより，文字がもつ多くの画像特徴を学習するためのデータとなる．

4.2 部分空間の作成

学習データの集合から既存の部分空間法 [7] を用い，辞書データとなる部分空間を作成する．ある学習画像 i に対して，ラスタスキャンにより次元のベクトルを作成し，更に正準化（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

$$\mathbf{x}_i = [x_1, x_2, \dots, x_N]^T \quad (1)$$

で表す．ここで N は画像の画素数である．正準化を行った k 枚の学習データを並べた行列の自己相関行列を求め，この自己相関行列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．以上の処理を各カテゴリーに対して行う．ここでカテゴリー c の固有ベクトルの集合を以下のようにおく．

$$\{\mathbf{e}_1^{(c)}, \mathbf{e}_2^{(c)}, \dots, \mathbf{e}_R^{(c)}\} \quad (2)$$

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された固有ベクトルを画像化した例である．この図から，異なる解像度の文字画像を学習したことによる文字のぼけや，手ぶれによるエッジの移動が確認できる．

5. 多入力型ロバスト部分空間法

本論文で提案する多入力型ロバスト部分空間法による認識は、複数の入力サンプルを同時に認識に用いることで各サンプルがもつ画像情報を累積的に利用し、また、携帯デジタルカメラで撮影された文字動画像の性質とロバスト推定の考えを効率的に統合した新しい手法である。本手法は、複数フレームの入力画像からの平均文字画像の推定、平均文字画像からの距離に基づく重み付きの部分空間法による認識の二つからなる。平均文字画像の推定はこの処理の後に続く重み付き部分空間法のための重要な処理である。平均文字画像は各フレームの入力サンプルの画素値を平均することで作成するが、このとき全入力サンプルを均等に扱おうと、例外サンプルが存在した際、その影響が大きくなる可能性がある。そこで、平均文字画像の推定を一種のモデル推定ととらえ、例外サンプルの存在に対してロバストな推定を行う。認識では、ここで推定された平均文字画像を基準として、入力サンプルに重みを付けて、類似度を計算することで、例外サンプルの存在が類似度計算に与える影響を抑制する。

5.1 平均文字画像の推定

本手法では、入力サンプルの分布中心、つまり平均文字画像からの距離に基づき入力サンプルに重みを付与して、平均文字画像の更新を繰り返すことで、平均文字画像をロバストに推定する。

具体的な処理手順は以下のとおりである：

Step1(初期平均文字画像ベクトルの作成)： 4.2 と同じ正準化処理を行った M 個の入力画像の集合 $\{a_1, a_2, \dots, a_M\}$ を用い、以下の式により初期平均文字画像 $\bar{a}^{(0)}$ を作成する。

$$\bar{a}^{(0)} = \frac{1}{M} \sum_{m=1}^M a_m \quad (3)$$

Step2(平均文字画像の更新)： 作成した平均文字画像を基準として、以下の式により平均文字画像を更新する。

$$\bar{a}^{(t)} = \frac{1}{\hat{M}^{(t)}} \sum_{m=1}^M \omega_m^{(t)} a_m \quad (4)$$

ここで、 $\bar{a}^{(t)}$ は t 回目のループで計算された平均文字画像であり、 $\omega_m^{(t)}$ は入力画像 a_m に対する重み、 $\hat{M}^{(t)}$ は重みの和を 1 とするための正規化項である。

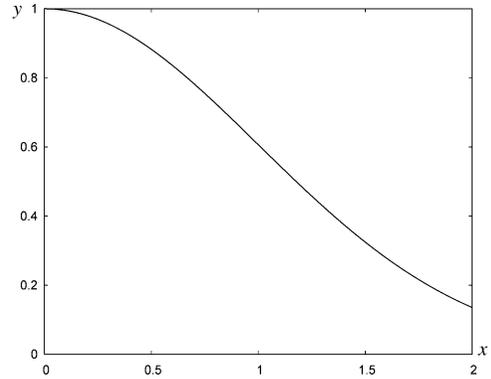


図 6 ガウス関数の例 . $\exp(-x^2/2.0)$
Fig.6 An example of a Gaussian function.
 $\exp(-x^2/2.0)$

Step3(終了判定)： 以下の条件を満たすまで、Step2を繰り返す。

$$t \geq T \quad \text{または} \quad \|\bar{a}^{(t)} - \bar{a}^{(t-1)}\| < D. \quad (5)$$

T は繰返し回数の上限、 D は推定された平均文字画像の変化量の収束を計るしきい値である。

図 3 に示したように、入力サンプルの分布の中心から離れた例外サンプルの影響を抑制するために、重み関数は、平均文字画像からのユークリッド距離が小さいもの、つまり入力サンプルの分布中心である平均文字画像に近い入力サンプルほど大きな重みを与え、平均文字画像から遠いものほど小さな重みを与える関数である必要がある。そこで本手法では重み関数に以下のようなガウス関数を用いる(図 6)。

$$\omega_m^{(t)} = \exp\left(\frac{-\|\bar{a}^{(t-1)} - a_m\|^2}{\sigma}\right) \quad (6)$$

このような重み関数を平均文字画像の計算時に各入力サンプルに付与することで、相対的に入力サンプルの分布中心に近いサンプルほど、推定される平均文字画像への寄与を大きくし、例外サンプルの影響を抑制する。また、パラメータ σ は実験的に決定する。ここで最終的に推定された平均文字画像を \bar{a} とする。

5.2 重み付きの部分空間法による認識

ここでは、平均文字画像を基準とした、重み付きの部分空間法について説明する。

平均文字画像の推定に使用したものと同一入力画像列 $\{a_1, a_2, \dots, a_M\}$ を用い、平均文字画像からの距離に基づく重み付き類似度を計算する。5.1 の処理で

推定された平均文字画像 \bar{a} を用い、入力画像列とカテゴリ c の辞書データとの類似度を以下のように定義する。

$$L_{robust}^{(c)}(\mathbf{a}) = \frac{1}{\hat{M}} \sum_{m=1}^M \sum_{r=1}^R \theta_m(\mathbf{a}_m, \mathbf{e}_r^{(c)})^2 \quad (7)$$

ただし $\theta_m = \exp\left(\frac{-\|\bar{\mathbf{a}} - \mathbf{a}_m\|^2}{v}\right)$

\hat{M} は重みの和を 1 とする正規化項であり、 (x, y) はベクトルの内積を表す。重み関数には図 3 に示したように、入力サンプルの分布中心に近いものほど大きな重みを与えるためガウス関数を用い、パラメータ v は実験的に決定する。

6. 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) で文字を撮影した動画画像を用いて、学習データの作成と認識実験を行った。対象文字は 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 カテゴリー) であり、これらの文字を紙に印刷し、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画像からフレームごとに文字を切り出し、異なる日に撮影したものをそれぞれ学習データと認識実験に用いた。文字の切出しは人手で行った。紙とカメラの距離を変化させることで、様々な文字サイズの画像を収集した。学習データと評価データの撮影者は同一人物であり、被験者数は 1 名である。また、学習データ、評価データにおける文字画像の正規化サイズを 32×32 pixel とした。大きさの正規化には 4.1 と同様の手法を用いた。

また、本手法における各パラメータの値は実験的に決定した。その値を表 1 に示す。以降の実験において同じ値を使用する。

本手法では、各入力サンプルに重み付けをして類似度を計算するが、単純な手法として、複数のサンプルの平均文字画像を入力として認識に用いる手法も考えられる。そこで予備実験として、平均文字画像を入力として部分空間法で認識した場合と、本手法を比較した。実験では、文字形状が類似したカテゴリ 2 種 (o: 小文字オー, 0: 数字ゼロ) を仮に選び (平均文字サイズ: 6×6 pixel, 入力フレーム数: 15 フレーム), 30 サンプルに対して実験を行った。認識率は、平均文字画像を入力する手法で 91.7% に対し、本手法では 98.3% となった。これは、各入力サンプルに対して平

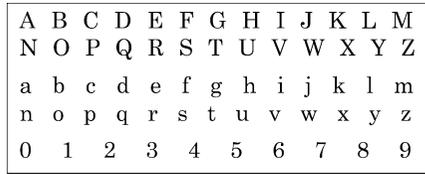


図 7 使用したフォント Century
Fig. 7 Set of target characters (Font: Century).

表 1 本実験における各パラメータの値
Table 1 Values of parameters used for the recognition.

辞書データ	
部分空間の次元数 R	5
平均文字画像推定	
重み関数 ω のパラメータ σ	2.0
ループ回数 T	5
平均文字画像の変化量に対するしきい値 D	0.01
重み付きマルチフレーム部分空間法	
重み関数 θ のパラメータ v	0.2

均からの距離に応じた重みを付加することにより、平均から外れた例外サンプルの影響が抑制されたためと考えられる。平均文字画像を入力する単純な手法に比べ、本手法の有効性が期待できる。

6.1 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文字の認識実験を行った。実験に使用したデータは以下のとおりである。

学習データ

- DV カメラで撮影
 - フレームレート: 30 fps
 - 撮影画像サイズ: 720×480 pixel
- 撮影距離 (平均文字サイズ)
 - size1: 20 cm (16×16 pixel)
 - size2: 30 cm (11×11 pixel)
 - size3: 40 cm (8×8 pixel)
 - size4: 60 cm (7×7 pixel)
 - size5: 70 cm (6×6 pixel)
 - 各文字サイズごとに 10 フレーム ($k = 5 \times 10$)

評価データ

- 学習データ取得の際と同じ 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)
 - 学習データと同様

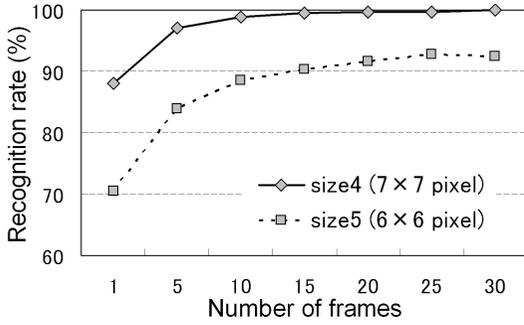


図 8 入力フレーム数による認識率の変化

Fig. 8 Recognition rates vs. number of input frames.

- 入力フレーム数 (M)
 - 1, 5, 10, 15, 20, 25, 30
- サンプル数: 各文字 30 セット

入力フレーム数による認識率の変化を図 8 に示す。ただし, size1 から size3 については認識率がほぼ 100%であった。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分かる。これは, 本手法が複数の入力画像の情報を効果的に利用できているためと考えられる。size5 では入力フレーム数が 15 付近で認識率の上限に達していることが分かる。これは, 用いた辞書データの次元数や入力画像の解像度による制限と考えられる。また, 平均文字サイズが 7×7 pixel の入力に関しては, ほぼ 100%の認識率に達した。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。

6.2 他手法との比較

他手法との比較実験を行った。本論文で提案した多入力型ロバスト部分空間法による認識手法を「提案手法: proposed method」とし, 比較のために, 入力サンプルに均一の重みを付与したものを「手法 1: comparative method ($\theta_m = 1$)」とした。また他手法との比較のために, 辞書データとの類似度から使用するサンプルを選択する手法を用いた。入力サンプル a_m ($m = 1, 2, \dots, M$) とカテゴリ c の部分空間との類似度を以下のように定義する。

$$l^{(c)}(a_m) = \sum_{r=1}^R (a_m, e_r^{(c)})^2 \quad (8)$$

辞書データとの類似度が最大となるフレームの文字画像 1 枚のみを使用する手法を「手法 2: previous method1」とし, そのときの類似度を以下のように定義する。

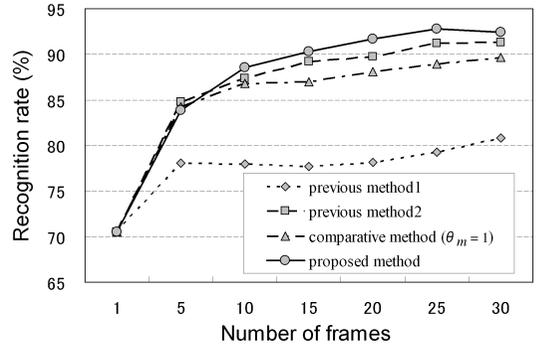


図 9 各手法の認識率の比較

Fig. 9 Comparison of recognition rates by each method.

$$L_{pre_method1}^{(c)}(a) = \max_m l^{(c)}(a_m) \quad (9)$$

また, 手法 2 を複数フレームの入力へと拡張した手法を「手法 3: previous method2」とし, そのときの類似度を以下のように定義する。

$$L_{pre_method2}^{(c)}(a) = \frac{1}{M} \sum_{m=1}^M l^{(c)}(a_m) \quad (10)$$

ただし $a_{m'} \in \{l^{(c)}(a_m) \text{ が大きい } a_m \text{ 上位 } M' \text{ 個}\}$

以上の手法を用い比較実験を行った。なお手法 3 では, 事前に行った実験の結果から M' を入力サンプル数の 8 割とした。

辞書データは 6.1 と同じものを使用し, また評価データには 6.1 のもののうち, size5 (6×6 pixel) のものを使用した。

結果を図 9 に示す。実験結果により, 入力フレーム数を増やすことで, どの手法も認識率の向上が確認できるが, 10 フレーム以上の入力では提案手法が最も高い認識率を示した。これにより, 入力文字画像の情報のみから認識に適切なサンプルを選定する本手法の有効性及び重み付けの効果が示された。

6.3 撮影機器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, 実際にこれらの機器で文字を撮影し, 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, 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た。

この実験では辞書データには 6.1 と同じ DV カメラで撮影した画像から作成したものをを用い、評価データの撮影にはカメラ付き携帯電話を用いた。実験に使用したカメラ付き携帯電話と評価データの仕様は以下のとおりである。

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

評価データ

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

実験の結果、DV カメラで撮影した同程度の文字サイズの入力画像において約 99% の認識率であったものが約 92% となり、辞書データと同じ機器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分かる。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。これらの問題を解決するには、今後、各カメラがもつ固有の画像劣化過程を学習の際に考慮する必要がある。

7. む す び

本論文では、動画像を用いて低解像度文字の認識を高精度で行う多入力型ロバスト部分空間法を提案した。携帯デジタル撮影機器で文書全体を一度に撮影した場合、各文字は非常に低解像度なものとなり、1 枚の画像だけからでは認識が困難である。本手法は、動画像中の連続した複数枚の画像を入力とし、それらの情報を効果的に統合して認識精度の向上を図るものである。その際、入力が想定される動画像の特徴を利用して入力サンプルの統計的な情報から各サンプルに重みを与えることで認識に有利なサンプルと認識に不利なサンプルの選別を行った。実験により、学習時と同条件で撮影された画像に対しては、低品質・低解像度で 1 枚の画像だけからでは認識が困難な場合でも、動画像中の複数フレームの画像を使用することで、高精度な認識が可能となることを確認した。また、他手法との比較実験により本研究で提案した入力サンプルに対

する重み付け手法の有効性を確認した。実験結果では、文字サイズが 7 × 7 pixel のような低解像度の文字に対して、約 99.9% の認識率を示した。また、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文字画像を認識した場合の認識率が、学習時に使用したものと同じ DV カメラで撮影した文字画像を認識した場合の認識率と比較して低下したことから、学習の際に各カメラがもつ固有の画像劣化過程をモデル化する必要性を示唆した。今後、異なる操作者に対する有効性の調査、日本語や異なるフォントへの対応、文字列や文への対応も検討する必要がある。

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。本研究の一部は文部科学省 21 世紀 COE プログラム及び科学研究費補助金 (No.16300054) による。

文 献

- [1] S. Senda, K. Nishiyama, T. Asahi, and K. Yamada, "Camera-typing interface for ubiquitous information services," Proc. Conf. Pervasive Computing and Communications 2004, pp.366-369, March 2004.
- [2] 池谷彰彦, 中島 昇, 佐藤智和, 池田 聖, 神原誠之, 横矢直和, 山田敬嗣, "カメラパラメータ推定による紙面を対象とした超解像ビデオモザイクング," 画像の認識理解シンポジウム 2004 論文集, vol.1, pp.505-510, July 2004.
- [3] S. Mori, K. Yamamoto, and M. Yasuda, "Research on machine recognition of handprinted characters," IEEE Trans. Pattern Anal. Mach. Intell., vol.PAMI-6, no.4, pp.386-405, July 1984.
- [4] P. Cheeseman, B. Kanefsky, R. Hanson, and J. Stutz, "Super-resolved surface reconstruction from multiple images," Tech. Rep. FA-94-12, NASA Ames Research Center, Artificial Intelligence Branch, Oct. 1994.
- [5] P.D. Thouin and C.I. Chang, "A method for restoration of low-resolution document images," Int. J. Document Analysis and Recognition, vol.4, no.4, pp.200-210, July 2000.
- [6] 澤口元英, 山本和彦, 加藤邦人, "携帯電話における低解像度文字認識手法の提案," 信学技報, PRMU2002-247, March 2003.
- [7] E. Oja, Subspace methods of pattern recognition, Research Studies, Hertfordshire, UK, 1983.
- [8] 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.
- [9] H. Murase and S.K. Nayar, "Visual learning and recognition of 3-D objects from appearance," Int. J. Computer Vis., vol.14, pp.5-24, 1995.
- [10] 大町真一郎, 阿曾弘具, "低品質文字認識におけるつづれを

- 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, Nov. 1999.
- [11] S. Uchida and H. Sakoe, "Handwritten character recognition using elastic matching based on a class-dependent deformation model," Proc. Int. Conf. Document Analysis and Recognition, vol.1, pp.163-167, Aug. 2003.
- [12] 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.
- [13] R.J. Humber, Robust statistics, John Wiley & Sons, 1981.
- [14] 島井博行, 栗田多喜夫, 梅山伸二, 田中 勝, 三島健稔, "ロバスト統計に基づいた適応的な背景推定法 ; 信学論 (D-II), vol.J86-D-II, no.6, pp.796-806, June 2003.
- [15] M. Sonka, V. Hlavac, and R. Boyle, Image Processing, Analysis, and Machine Vision, 2nd ed., p.67, PWS Publishing, 1999.
- (平成 17 年 3 月 31 日受付, 8 月 4 日再受付)



柳詰 進介

平 15 名大・工・情報卒 . 平 17 同大大学院情報科学研究科メディア科学専攻修士課程了 . 同年株式会社リコー入社 . 現在に至る .



高橋 友和 (正員)

平 9 茨城大・工・情報卒 . 平 12 同大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 フランス情報学・統計システム研究所 (IRISA) 招聘教授 . 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) . 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 本会ソサイエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, ほか受賞 . 情報処理学会, IEEE 各会員 .