

携帯カメラで撮影した低解像度画像からの文字認識

村瀬 洋, 石田皓之, 柳詰進介, 井手一郎, 目加田慶人(名古屋大学)

Character recognition from low resolution images captured by portable digital cameras
Hiroshi Murase, Hiroyuki Ishida, Shinsuke Yanadume, Ichiro Ide, Yoshito Mekada (Nagoya University)

1. まえがき

近年, カメラ付き携帯電話, DVカメラ, デジタルカメラなど, 画像・映像・音などのマルチメディア情報を入力できる装置を日常的に携帯する機会が増えている. このような入力装置を用いて我々の周囲のメディア情報を認識する研究が着目されている. 例えば携帯電話から入力した音響信号から音楽の曲名を認識[2]したり, 携帯カメラで印刷されたURL情報を認識する研究なども進んでいる. しかしながら, 従来, 携帯カメラから文字を認識する場合には画像を近距離から撮影し, 十分な画像品質が得られるようにすることが求められてきた. 例えば文書全体を遠方から撮影すると, 1文字のサイズは小さくなり, 低解像度の文字となる. カメラの分解能は年々増加しているものの, 一度により沢山の情報を入力したいという要求は常にある.

本論文では, 1枚の画像からは認識できないような低解像度の文字を, 精度良く認識する手法を提案する. 本手法では低解像度文字の認識のために以下の2つのアイデアを導入した. それは, (1) 1枚の画像だけでは精度良く認識できないような低解像度画像を複数のフレーム用い, 部分空間法により統合することにより認識精度を向上させること, (2) 低解像度になる過程をモデル化し, モデルを用いて学習サンプルを多数生成することで低解像度の文字の学習を手軽に行えるようにすることである.

本論文では, 動画像中の複数フレームを用いた文字認識法, 生成型の学習法について述べることにする.

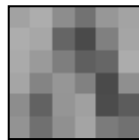


図1. 携帯電話での文字入力 図2. 文字の例“ A ”

2. 複数フレームを用いた低解像度文字認識

2.1 観測される画像データ

低解像度の文字画像とは, ある文字に着目したときに, その文字に割り当てられる画像の画素数が少ない場合を指す. これは, 多数の文字が印刷されている紙面をカメラにより一度で撮影する際(図1)に発生する. しかし, 1枚だけでは精度良く認識できない文字画像(図2)でも, 複数フレームからの画像を統合することにより認識精度が向上する[3]. カメラを手で持ち動画像で撮影する場合など,

手の微妙な揺れなどで入力される画像は様々に変化するが, ここで得られる多様な情報を認識に利用する手法である.

複数のフレームからの低解像度画像が与えられる場合, 2つのアプローチが考えられる. 第1の方法は超解像などの手法により複数の低解像度入力画像から解像度の高い画像を作成し, それを用いて認識する手法である. しかし一般に低解像度画像から解像度の高い画像を復元することは逆問題となり, 容易ではない. この逆問題は低解像度になるモデルが単純な場合にしか解けない. 特に携帯カメラなどにより, 低解像度で撮影された画像は, 単純な低解像度化だけでなく, 光学的なボケや歪み, 照明の変動など様々な要因がある. 第2の方法は, 低解像度化された学習パターンを多数生成し, それと照合する手法である. 本手法では, 生成型の学習を用いる. 一般に様々な変動をもったパターンを生成することは第1の手法に比較してそれほど困難ではない. また認識するには, 必ずしも, もとの解像度の高い画像を作成する必要もない.

2.2 部分空間法による低解像度認識

本手法では, 低解像度の文字を認識するために部分空間法の枠組みを用いることにする. 部分空間法は, 多数の学習サンプルを少数の固有ベクトルで近似し, この固有ベクトルがつくる部分空間と入力画像とを照合する手法である. 本手法のように多数の学習サンプルを生成し, それらと入力画像とを短時間で精度良く照合するには適した手法である. 特に本手法で生成される学習サンプルは互いに相関が強いために, 少ない部分空間の次元で精度良く表現することが可能である. 学習サンプルを生成し, そこから部分空間を作成する手法は, 古くは変動吸収特性核[1]を用いた認識などでも用いられた考え方である. 以下, 本手法の手順を述べる. 処理は, 学習段階と, 認識段階から構成される.

(1) 学習段階

文字カテゴリ数を C , 各カテゴリ毎の学習サンプル数を n , 認識に用いる部分空間の次元を L とする. 各学習サンプルは画素を要素とするベクトルで表現され, 前処理として正規化(大きさ1で画素全体での平均値が0となるような正規化)されているものとする. これらの n 個のベクトルの自己相関行列を計算し, その固有値展開により固有ベクトル (L 個の大きな固有値に対応したもの) \mathbf{e}_i ($i=1, \dots, L$) を計算する.

(2) 認識段階

動画像中から選択した複数フレームの画像を \mathbf{x}_j ($j=1, \dots, J$) とする。低品質なために1枚の画像からでは認識できないような画像も、複数フレームからの情報を累積的に利用することにより精度は向上する。相互部分空間法を利用する手法もあるが、ここでは各フレーム毎の類似度をフレーム分で平均した値を類似度とした。

$$s = \frac{1}{J} \sum_{j=1}^J \sum_{i=1}^L (\mathbf{x}_j, \mathbf{e}_i)^2$$

2.3 生成型学習法

学習段階で学習サンプルを収集することは重要であるものの、実際に多数の状況で撮影することは必ずしも容易ではない。ここでは、学習サンプルを効果的に生成することにより収集する手法を提案する[4]。つまり学習サンプルを生成するモデル自体を学習する手法とも言える。撮影により画像が劣化する要因には、光学的なボケ、手ぶれによるボケ、カメラの向きによる歪み、解像度の低下による低品質化などさまざまなものがある。ここではさまざまな要因をとりあえず、ボケの要因と、サンプリング配置の要因に分類する。光学的なボケやシャッター内の手の揺らぎなどは前者の要因である。一方解像度の変化はサンプリング点の密度に関連した量と考えられるため後者の要因と考えられる。

前者の要因は、画像の点広がり関数で表現することにする。この関数は、伝達関数と考えられるので、周波数領域での関数の割り算により計算する。実際は伝達関数自体確率的な関数と考えられるが、ここではとりあえず平均の伝達関数を求め、近似的にその伝達関数の広がり具合を変動させたものを採用する。一方、標準化関数は、様々に解像度を変化させて用いることにする。

3. 実験

3.1 実験方法

提案手法の有効性を示すために、印刷された62字種(英語の大文字と小文字、及び数字; Century フォント)をDVカメラで撮影し、実験に用いた。撮影距離は、20cm~70cm程度で変化させて撮影した。部分空間の次元数は5とした。この次元は事前調査により決定したものである。

(1) 実験1 使用するフレーム数の変化による認識率の変化を調査する実験を行う。学習サンプルとしては異なる距離から実際に撮影したデータ50サンプル/字種を用いた。

(2) 実験2 以下の2種類の手法で学習し、生成型学習法の効果を調査する実験を行う。

- ・生成型学習なし: フォントデータをそのまま学習サンプルに使用,
- ・生成型学習あり: カメラの特性を学習した劣化モデルにより生成した20種類(1字種に対して)の学習サンプル

を学習に利用した。生成の際、解像度変換も考慮している。

3.2 実験結果

(1) 複数フレームを用いた際の認識率

各認識では各文字について動画像から抽出した連続する複数フレームを用いた、フレーム数に関する認識率の変化を図3に示す。使用するフレーム数が増加すれば認識率は向上することがわかる。今回の例では5フレームくらいで認識率は飽和した。

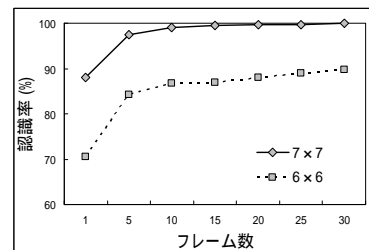


図3. フレーム数を増やした際の認識率の変化
(平均文字サイズが6画素四方7画素四方の場合)

(2) 生成型学習法の効果

2種類の学習手法による認識率の違いを表1に示す。生成型学習法は、解像度の低い入力画像に特に有効であることがわかった。実際に集めたデータを学習サンプルとする手法より認識率は落ちるものの、容易な学習が可能となる。

表1. 生成型学習の効果(複数(10枚)フレームを利用)

平均文字サイズ	12画素	10画素	8画素	6画素
生成型学習あり	91.9%	89.7%	88.8%	78.5%
生成型学習なし	70.3%	62.3%	45.8%	27.7%

4. まとめ

本発表では、1枚の画像からでは認識できないような低品質な文字画像を認識する手法を述べた。これらは近年ますます普及する携帯型のカメラにより撮影した文字の認識に有効である。この問題に対処するために、複数フレームの情報を利用すること、学習時に劣化関数を利用して学習パターンを生成する生成型学習を提案した。実際に撮影した文字データを用いた実験により、本手法により認識精度が向上することを確認した。

謝辞: 本研究の一部は21世紀COE, 科研補助金の支援を得ている。

文献

- [1] 村瀬, 木村, 吉村, 三宅, 信学論, Vol.J64-D, No.3, pp276-283, 1981.
- [2] 黒住, 柏野, 村瀬, 信学論, Vol.J84-DII, No.8, pp.1817-1825, 2001.
- [3] 柳詰, 目加田, 村瀬, 信学全大, Vol.3, pp197, 2004.
- [4] 石田, 柳詰, 目加田, 井手, 村瀬, 信学技法 PRMU2004-7, 2004