

令和元年5月22日現在

機関番号：14101

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00236

研究課題名(和文) 確率的補間による文字の3次元回転角度推定の高精度化

研究課題名(英文) Accuracy improvement of the 3D rotation angle estimation of characters by the probabilistic interpolation

研究代表者

若林 哲史 (Wakabayashi, Tetsushi)

三重大学・工学研究科・教授

研究者番号：30240443

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：特徴ベクトルの低次元化と回転角度の確率的補間により、記憶容量を増加させることなく文字の3次元回転角度の精度を高めるとともに、文字候補の組み合わせと車番テンプレートのマッチングを取ることで、様々な角度から撮影された車番の認識を実現した。また、複数の連結成分からなる漢字・仮名で構成される文字列の撮影画像から、組み合わせ最適化により、文字切り出し、認識、回転角度推定を行う手法を提案した。

研究成果の学術的意義や社会的意義

記憶容量の削減と3次元回転角度推定の精度向上により、日本語の3次元回転不変文字認識が可能となり、従来は実現不可能であった日本語の各種看板や道路案内標識を様々な角度から自在に認識することができるようになる。その結果、動画からのリアルタイムの翻訳案内システムやカメラの設置場所を選ばない車番認識システムによる駐車場の入出管理など、カメラベース文字認識システムの発展に大きく寄与する。

研究成果の概要(英文)：We improved the precision of three-dimensional rotation angle estimation of the character without increasing memory usage by reducing feature vector size and employing the probabilistic interpolation of the rotational angle. And we proposed the camera-based license plate recognition for the image photographed from various direction by the matching between the combination of character candidates and the template of license plate. In addition, we proposed the method to detect and recognize kanji character consists of multiple connecting components by combinational optimization.

研究分野：知覚情報処理

キーワード：文字認識 回転角度推定 車番認識 カメラベース認識 確率的補間

様式 C - 19, F - 19 - 1, Z - 19, CK - 19 (共通)

### 1. 研究開始当初の背景

デジタルカメラの性能向上や利用方法の多様化により、カメラを入力機器として利用する文字認識技術に期待が高まっている。過去 20 年間、研究者は従来のスキャナベースの文字認識技術をカメラベースの認識に応用してきた。しかし、スキャナベースとは異なり、カメラで撮影した画像中の文字認識には以下の二つの問題がある。

第 1 に、正面から撮影されていない文字は 3 次元の回転や透視投影の影響で変形するため、認識率が低下する。この問題に対処するため、文字列単位で推定した傾きを補正する手法やアフィン不変な特徴を用いる手法が提案されているが、文字列の傾き検出が困難であったり、複数フォントに対応するために記憶容量と処理時間が増加したりする問題がある。一方で、あらかじめ生成した全ての变形文字と照合を行う手法が提案されているが、照合対象が増加するため、認識に必要な記憶容量と処理時間は膨大なものになる。

第 2 に、複雑な背景上の文字領域を検出するのは困難であるため、ICDAR2013 Robust Reading Competition においても、検出性能は 80% 以下であった。文字検出には、字種や形状を利用する方法、エッジに基づく Stroke Width Transform から得られる連結成分の文字幅を利用する方法、空や看板といった環境コンテキストを利用する手法などが提案されている。

我々の提案する手法は、記憶容量や計算コストの増加を抑えつつ、上記の二つの問題に対処するものである。

### 2. 研究の目的

我々は、これまで培ってきた画像処理技術、文字認識技術を使ってカメラベースの 3 次元回転不変文字認識手法を提案してきた。また、写真中の文字列を検出するには、個々の文字の回転角度を推定して文字印字平面の法線ベクトルを求め、その向きを利用してグループ化することが有効であることも明らかにしてきた。しかし、3 次元の回転角度推定に必要な記憶容量は膨大であるため、字種の多い漢字・仮名を含む日本語を対象とすることが難しかった。本研究では、回転角度推定に必要な記憶容量を削減する手法を提案するとともに、確率的補間を用いて回転角度推定精度を向上させ、カメラベース文字認識の課題を解決する。さらに研究成果を日本語の看板や道路案内標識の認識、撮影方向を選ばない車番認識に適用してその有効性を確認する。

### 3. 研究の方法

#### (1) 記憶容量の削減と高速化

過去の研究成果により、濃度こう配特徴ベクトル(392 次元, 三重大学開発)が 3 次元回転不変文字認識に有効であることがわかっている。そのため、これまでは、3 次元回転角度推定にも 392 次元の濃度こう配特徴ベクトルを用いていた。これは文字画像から生成した濃度こう配画像を縦横  $7 \times 7$  のブロックに分割し、各ブロックで 8 つの方向ごとに強度を加算したものである ( $7 \times 7 \times 8 = 392$  次元)。このとき画像の縦横分割数を  $6 \times 6, 5 \times 5, \dots, 1 \times 1$  と減らすことで特徴ベクトルの次元数を削減する。回転情報が重要であるため、方向量子化数は減らさない。三重大学で開発された識別関数である MQDF を文字認識に用いる場合は、通常特徴ベクトルの次元数の  $1/3 \sim 1/2$  程度の固有ベクトルを用いて内積計算を行う。この固有ベクトル数を削減することで学習辞書の容量を削減する。これらは、記憶容量だけでなく計算量の削減にも繋がる。比較実験により、回転角度推定の精度を落とさない範囲の最適な特徴ベクトル次元数と固有ベクトル数を検討する。さらに、ある字種のある角度におけるサンプルの分布が全てのクラス(字種・回転角度)で等しいと仮定し、最適な線形識別関数によってさらに高速な角度推定を試みる。

#### (2) 確率的補間による回転角度推定の高精度化

回転角度を照合するとき得られる、サンプルとすべての回転角度クラスとの間の識別関数値から、補間対象となる複数の回転角度クラスの事後確率を計算し、回転角度各クラスの法線ベクトルと掛け合わせて平均を取ることで、学習に用いた回転角度の刻みよりも高い精度で文字の回転角度を推定することができる。逆に精度を保ちながら学習する回転角度クラスを減らすことも可能になる。これまで  $5^\circ$  刻みで回転角度クラスを設定していたため、その間の角度の回転文字はどちらか近い方の回転角度として検出されていた。確率的補間を用いることで、中間の角度も推定することができるため、推定誤差の減少が期待できる。どの程度推定誤差が減少するかを調査するとともに、 $10^\circ$  刻み、 $15^\circ$  刻みと回転角度クラスを減少させたときの影響を調べる。

#### (3) 撮影方向を選ばない車番認識

プレート検出処理を行うことなく、画像から直接文字検出・回転文字認識を行うことで、様々な角度から撮影された画像に対して車番認識を実現する手法を提案する。本手法では、初めに入力画像中から文字の可能性のある連結領域を文字候補として検出する。次に、検出された文字候補に対し、文字認識・回転角度推定を行う。さらに、サイズを変更しながら画像中の探索ウィンドウをスライドさせ、各ウィンドウに対して以下の処理を行う。まず、ウィンドウ内に含まれる文字候補をノードとする入力グラフを生成する。グラフにおけるノード特徴量は、推定済みの文字回転角度とする。グラフのエッジ特徴量はエッジに接続され

た文字候補ペアの高さの比,中心間ユークリッド距離と高さの比,中心間を結ぶベクトルの三つとしている.予め車番テンプレートを3次元回転させ,様々な角度の参照グラフ群を生成し,入力グラフの推定車番プレート回転角度により参照グラフを選択する.入力グラフと参照グラフ間でグラフマッチングを行い,エッジ特徴量よりコストを計算する.最後に,全探索ウィンドウで最小コストのグラフのノードを車番として出力する.

(4) 印刷漢字・仮名の3次元回転不変文字認識

我々がこれまでに提案してきた3次元回転不変文字認識では,連結成分ごとに文字を認識するため,1文字が1つの連結成分で構成される英数字のみを対象としていた.しかし,漢字・仮名には,一つの文字が複数の連結成分から構成されている字種が多く,1文字単位での切り出し,認識,回転角度推定が困難な場合がある.そこで,抽出した連結成分を組み合わせて文字候補を生成し,生成した文字候補に対して識別関数によりクラスと識別関数値を求め,識別関数値を用いて最適な文字集合の組み合わせを決定する.図1に文字候補生成例を示す.

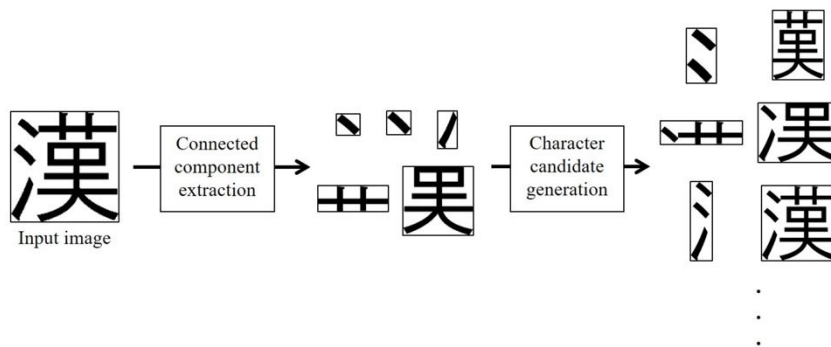


図1 文字候補生成例

4. 研究成果

(1) 記憶容量の削減と高速化

記憶容量を削減するために,1)領域分割数の削減による特徴ベクトルの低次元化,2)識別に用いる固有ベクトル数の削減,3)識別関数の変更の3点について,推定精度と記憶容量に関する実験を行った.1)においては,領域分割数を7x7から2x2にすることで,特徴ベクトルの次元数を392次元から32次元まで落としても認識率はほとんど低下せず学習辞書の記憶容量を1.7GBから140MBに押さえることができた.2)においてはMQDFで用いる固有ベクトル数はk=1で十分であり記憶容量はさらに19.2MBにまで減少した.3)においては線形識別関数を用いることで若干精度は劣るものの記憶容量をさらに6.6MBまで削減することができた.これにより字種数が多い漢字・仮名の3次元回転角度推定が可能となった.

(2) 確率的補間による回転角度推定の高精度化

回転角度推定において,確率的補間ありと無しの場合の回転角度推定精度を比較した.回転角度の間隔を15度,10度,5度の3パターンとした学習辞書を構成した.特徴ベクトルと識別関数は,32次元の濃度こう配特徴ベクトルと線形識別関数を用いる.実験結果を表1に示す.回転角度の間隔を小さくするにつれて推定精度が向上していることがわかるが,確率的補間なしではそれに伴い記憶容量が増加している.確率的補間ありの場合,記憶容量を増加させることなく,推定精度が向上していることがわかる.

表1 確率的補間による法線ベクトルの平均推定誤差(度)

	5度間隔	10度間隔	15度間隔
補間なし	7.77	9.01	10.91
補間あり	6.81	7.53	8.95
記憶容量(MB)	44.0	6.6	1.9

(3) 撮影方向を選ばない車番認識

様々な方向から撮影された日本の車番を含む 1920×1080 画素の情景画像 108 枚を実験データとして認識実験を行った結果、文字単位の認識成功率 99.5%、プレート単位の認識成功率 96.3%という高い車番認識精度を得た。また、サイズが 320×240 画素から 640×480 画素の 2049 画像からなる台湾の公開データセットを用いた認識実験では、文字単位の認識成功率が 92.8%と、他手法を上回る車番認識精度を達成し、低解像度画像に対する有効性も示された。本手法は、車番テンプレートを変更するだけで様々な国や地域の車番に対応可能な 3 次元回転不変車番認識を実現している。図 2 に車番認識の成功例を示す。



図 2 車番認識の成功例

(4) 印刷漢字・仮名の 3 次元回転不変文字認識

学習用データとして、活字文書画像から切り出した日本語の印刷文字データを用いた。このデータセットは、ひらがな、カタカナ、JIS 第 1 水準に含まれる常用漢字の全 3112 字種を含み、各字種に対して解像度やフォント、大きさが異なる 280 サンプルの画像からなる。評価用データは独自で作成したもので、図 3 に示されるような複数の漢字駅名が含まれる画像を正面からカメラで撮影し、計算機上で 3 次元の回転処理を施した。回転の範囲を x 軸まわり、y 軸まわりに -45 度～45 度、z 軸まわりに -30 度～30 度、刻み幅を 15 度として、回転画像を生成し、全 2450 枚の画像を作成した。実験の結果、画像から文字が正しく抽出できたかどうかの適合率、再現率から求めた F 値が、透視投影による歪みが無い場合に 80.63%、透視投影による歪みがある場合に 77.39%となった。図 4 に提案手法による認識成功例を示す。

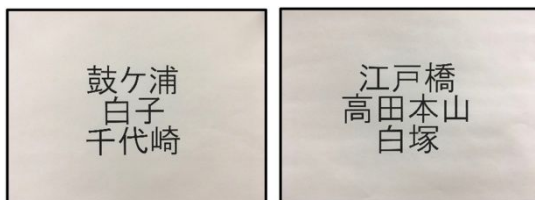


図 3 評価用データの例

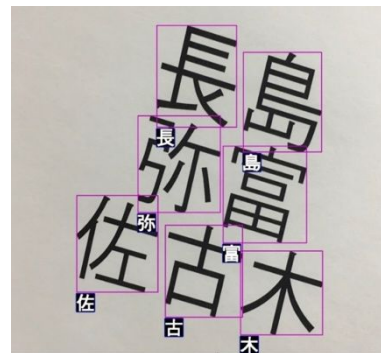


図 4 3 次元回転文字認識成功例

(5) 画像中のメーカーロゴの切り出しと認識

関連研究として、学習サンプルが少ないときの撮影画像内のメーカーロゴの自動抽出・認識手法も提案した。MSER を用いて初期領域を抽出し、反復的に領域をグループ化することでロゴを検出し、2 段階の SVM によりロゴ認識を行った。制約無しに撮影された IEICE-PRMU shoe logo dataset に対する認識実験では、ロゴ検出と認識において有望な結果を得た。図 5 に靴ロゴの検出と認識の例を示す。



図 5 靴ロゴの認識例

(6) 撮影条件に依存しないドットマトリクス文字の抽出

関連研究として、機械学習の生成モデルの一つである条件付き対立ネットワーク (Conditional Generative Adversarial Networks (cGAN)) を用いるドット文字抽出手法を提案した。提案手法では、cGAN を用いて画像中のドット文字を消去する画像変換を学習する。この学習モデルにより、入力画像に対してドット文字を消去する処理を行い、変換前の画像との差分をとることでドット文字のみを抽出する。評価実験では、先行研究で使用していたデータセットに新たな画像を加え、先行研究及び提案手法での文字抽出、文字認識結果を比較した。認識実験の結果、先行研究ではデータセット全体の 440 枚の画像に対して文字認識の F-measure が 56%であったが、提案手法では 62%に向上した。また、先行研究で用いていたデータセット 250 枚に対しての F 値も 65%から 71%に向上し、様々な撮影条件で撮影した画像に対するドット文字の抽出、認識をより高い精度で行えることが明らかになった。図 6 にドット文字検出例を示す。





図 6 ドット文字検出例

## 5. 主な発表論文等

### 〔雑誌論文〕(計 1 件)

Kazunori Aoki, Wataru Ohyama, Tetsushi Wakabayashi “Extraction and Recognition of Shoe Logos with a Wide Variety of Appearance using Two-Stage Classifiers” IEICE Transactions on Information and Systems, Vol.E101-D, No.5, pp.1325-1332 (May 2018) 査読有.

### 〔学会発表〕(計 10 件)

溝手 健志, 大山 航, 若林 哲史: 複数の連結成分で構成される 3 次元回転文字の認識: 第 21 回画像の認識・理解シンポジウム (MIRU2018), PS3-47 (2018 年 8 月 5 日~8 日, 札幌コンベンションセンター)

伊藤 冬湖, 大山 航, 若林 哲史, 白井 伸宙: 撮影条件に依存しないドットマトリクス文字抽出: 第 24 回・画像センシングシンポジウム (SSII2018), IS1-31 (2018 年 6 月 13 日-15 日, パシフィコ横浜)

<https://conf.it.atlas.jp/guide/event/ssii2018/subject/IS1-31/advanced>

富岡 永伍, 大山 航, 若林 哲史, 白井 伸宙: カメラベース OCR における連結成分解析を用いた東アジア言語の判定: 第 24 回・画像センシングシンポジウム (SSII2018), IS2-33 (2018 年 6 月 13 日-15 日, パシフィコ横浜)

<https://conf.it.atlas.jp/guide/event/ssii2018/subject/IS2-32/advanced>

溝手 健志, 大山 航, 若林 哲史: 日本語に対応した 3 次元回転不変文字認識: 平成 29 年度電気・電子・情報関係学会東海支部連合大会 (2017 年 9 月 7 日-8 日, 名古屋大学)(F2-2)

青木 一憲, 大山 航, 若林 哲史: 多様な見えの靴ロゴマークに対する自動抽出と認識: 平成 29 年度電気・電子・情報関係学会東海支部連合大会 (2017 年 9 月 7 日-8 日, 名古屋大学)(F2-6)

Wataru Ohyama, Koushi Suzuki, Tetsushi Wakabayashi: “Recognition and Defect Detection of Dot-matrix Text via Variation-model Based Learning”, Proceedings of 13th International Conference on Quality Control by Artificial Vision (QCAV2017), Proc. of SPIE Vol.10338 103380K, pp.1-8 (May 14-16, 2017, Tokyo, Japan)  
DOI: 10.1117/12.2264232

Kazunori Aoki, Wataru Ohyama, Tetsushi Wakabayashi: “Automatic Extraction and Recognition of Shoe Logos with a Wide Variety of Appearance”, Proceedings of 15th IAPR International Conference on Machine Vision Applications (MVA2017), pp.185-188 (May 8-12, 2017, Nagoya, Japan)  
DOI: 10.23919/MVA.2017.7986838

山本秀馬, 大山 航, 若林 哲史: 3 次元回転不変文字認識の印字面形状復元への応用: 信学技報, vol. 116, no. 461, PRMU2016-154, pp. 13-17, (2017 年 2 月, 北海道大学)

Yuki Nakamura, Wataru Ohyama, Tetsushi Wakabayashi, Fumitaka Kimura: “License-Plate Recognition via 3 Dimensional OCR”, Proceedings of 3rd ICET: 2nd International Conference on Information Technology (ICIT2016), ICIT009, pp33-37 (August 6-7, 2016, DITC, Kemaman, Terengganu, Malaysia).

Yuki Nakamura, Wataru Ohyama, Tetsushi Wakabayashi, Fumitaka Kimura :3D-OCR with DoG filter for Detection-free Automatic License-Plate Recognition: 第 19 回画像の認識・理解シンポジウム (MIRU2016), PS1-40 (2016 年 8 月 1 日~4 日, アクトシティ浜松)

〔図書〕(計 0件)

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕

ホームページ等

<http://www.hi.info.mie-u.ac.jp/publications/>

## 6. 研究組織

### (1)研究分担者

研究分担者氏名：大山 航

ローマ字氏名：(OHYAMA, Wataru)

所属研究機関名：九州大学

部局名：システム情報科学研究所

職名：准教授

研究者番号(8桁): 10324550

### (2)研究協力者

研究協力者氏名：

ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。