法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-04

情景内カラー文字認識のための最適2値化と変形耐性画像マッチング

Suzuki, Masato / 鈴木, 慎人

(出版者 / Publisher)
法政大学大学院情報科学研究科
(雑誌名 / Journal or Publication Title)
法政大学大学院紀要. 情報科学研究科編
(巻 / Volume)
14
(開始ページ / Start Page)
1
(終了ページ / End Page)
6
(発行年 / Year)
2019-03-31
(URL)
https://doi.org/10.15002/00021943

情景内カラー文字認識のための最適 2 値化と変形耐性画像マッチング

Recognition of Color Characters in Scene Images via Optimal Binarization and Distortion-tolerant Image Matching

鈴木　慎人
Masato Suzuki
法政大学大学院情報科学研究科情報科学専攻
E-mail: 17t0015@cis.k.hosei.ac.jp

Abstract

In this paper, we propose the method of recognition of color characters in scene images. In this method, we apply two key ideas to recognition. The first idea is optimal binarization, which separates the image into the character region and the background region. We use K-means clustering in Lab color space to generate multiple segmented images. After generation, these segmented images are classified to correctly binarized character image or not, by the convolutional neural network. All of the segmented images are classified, one of binarized image is selected as the optimal binarized character image among them. The second idea is distortion-tolerant image matching. Distortion indicates rotation, expansion, shearing, and translation, expressed by an affine transformation. In this matching, the correlation value is calculated between the target image and the template images first. Then, the optimal affine transformation is applied to the template images, to make a higher correlation value. These steps are repeated while the maximum correlation values are gained. Finally, we select k samples which gained higher correlation values and classify the target image using k-NN classification method. Experimental results made on the public color character image dataset "the Chars74K" show that the proposed method achieves 96.6% of correctly binarized character image selection rates and 73.9% of character recognition rates.

1. まえがき

近年，文字認識の研究において手書き文字に対する認識精度は飛躍的に向上している．主に CNN（Convolutional Neural Network）を用いた認識手法が数多く提案されており，手書き数字の公開データセット“MNIST”に対する認識では 99%を超える極めて高い精度が得られている．また，一般の手書き文書においても認識手法が確立されつつあり，実用化例として手書き文書に対して人工知能を組み合わせて認識を行うサービスが存在する．

一方で，情景内カラー文字画像に対しては，安定して高い精度が得られる手法が確立されていない．Shi 等[1]は，CNN によるカラー単文字画像の認識手法を報告している．カラー文字画像をグレースケールに変換して，CNN の入力としている．また，小規模データベースに対する認識精度を向上させるため，認識対象とは別のデータベースを用いた fine-tuning を行っている．

情景内カラー文字画像においては，文字領域・背景領域ともに多種多様な色や形状が見られ，照明条件も安定していない．単純なグレースケール化では文字領域のみを抽出することは難しいため，2 値化によって，背景領域を除去する必要がある．

また，文字認識において CNN を用いる場合，拡大・縮小や回転に対するロバスト性が保証されていない．さらに，fine-tuning を行う際，学習済みモデルの表現力を評価することは難しく，再学習の度合いには任意性がある．

上記の問題を解決するため，本研究では，最適 2 値化および変形耐性画像マッチングによる文字認識手法を提案する．まず，カラー文字画像に対して Lab 色空間で K-means クラスタリングを行い，複数の 2 値化候補画像を生成する．次いで，CNN を用いた 2 クラス分類で，複数の候補画像から最適な 2 値化画像を選択する．最後に，アフィン変換耐性を有する GAT 相関法を用いた *k*-NN 分類による認識を行う．

提案手法を情景内文字画像データセット“The Chars74K”に適用した評価実験により，最適 2 値化画像選択率 96.6%，文字認識率 73.9%を達成した．

以下，**2.** で使用した画像データを紹介する．**3.** で K-means クラスタリングによる 2 値化候補画像の生成，**4.** で CNN を用いた文字・非文字の判定と 2 値化画像の決定，**5.** で変形耐性画像マッチングによる文字認識，**6.** で最適 2 値化の実験結果，**7.** で文字認識の実験結果について述べる．**8.** で考察を述べ，**9.** でむすびとする．

2. 情景内カラー文字画像データセット

本研究では，Campos 等の作成した情景内文字画像データセット“The Chars74K”[2]を使用した．0～9 までのアラビア数字 10 カテゴリと，A～Z，a～z のアルファベット 52 カテゴリの計 62 カテゴリからなる画像データのうち，情景内単独カラー文字画像データ“*Img*”と，コンピュータによって生成されたフォント画像データ“*Fnt*”を用いた．*Img* データセットは Campos らが撮影した 12,503 枚の画像から各カテゴリについて学習用 15 枚，テ

Supervisor: Prof. Toru Wakahara

スト用 15 枚をランダムに選出したものであり，*Fnt* データセットは各カテゴリ 1,016 種のフォント画像である．

図 1 に各々の画像例を示す．

(a)

(b)

図1. データセット“The Chars74K”．(a) 情景内単独カラー文字画像“*Img*”．(b) フォント画像“*Fnt*”．

3. K-means クラスタリングによる 2 値化候補画像の生成

本章では，クラスタ数の自動選択機能を持つ K-means クラスタリングを用いた 2 値化候補画像の生成[3]について述べる．

3.1. 色空間の変換

カラー文字画像に対して色空間の変換を行う．本研究では色空間に RGB 色空間および Lab 色空間を用いる．Lab 色空間は輝度と色度を軸に持つ色空間であり，変換は OpenCV を用いて実装した．

3.2. K-means クラスタリング

画像サイズ$M \times N$ のカラー文字画像に対して RGB 色空間または Lab 色空間で K-means クラスタリングを行い，K個のクラスタに分割する．K-means クラスタリングにより得られた各クラスタについて，当該クラスタに属する画素群を元のカラー文字画像に逆投影すると，各々1 枚の分離画像が生成される．これらK枚の分離画像の和集合が元の画像となる．

K枚の分離画像群を網羅的に 2 つのグループに 2 分割して，一方を文字部分(黒)，他方を背景部分(白)として，複数の 2 値化候補画像を生成する．

図 2 に，1 枚のカラー文字画像に対する 2 値化候補画像の生成例を示す．

(a)

(b)

図2. 2 値化候補画像の生成例．(a) 元のカラー文字画像．(b) 生成された 2 値化候補画像．

上記 2 分割の全ての組合せにより生成される 2 値化候補画像の総数N_{binary} は式(1)の通りである．

$$N_{binary} = \sum_{i=1}^{K-1} {}_K C_i = 2^K - 2. \qquad (1)$$

クラスタ数Kの値は，クラスタ内分散和に対し閾値Tを設定することで自動選択する．閾値Tは予備実験より定めた．また，閾値Tの設定により，クラスタ数Kの最大値は 8 となった．そのため，本研究で生成される 2 値化候補画像は最小で 2 枚，最大で 254 枚となる．

3.3. 画像の正規化

得られた全ての 2 値化候補画像について，サイズを統一するため，文字部分の位置と大きさの正規化処理[4]を施す．具体的な手順を以下に記す．

まず，画像内全ての黒画素の重心g，および重心g から各黒画素への平均距離r を算出する．次に，あらかじめ定めた平均距離の正規化値$r_0(= 30.0)$ に従い，重心g 周りの画像の伸縮率$s = r_0/r$ を決定する．そして，黒画素の重心g を$(64, 64)$ に移動し，伸縮率s で大きさを正規化して，128×128 のサイズとする．

4. CNN を用いた文字・非文字の判定と 2 値化画像の決定

CNN（Convolutional Neural Network）は，画像を対象とした深層学習の一手法である．畳込み演算やプーリング処理を行うため位置変化に頑健なモデルであり，近年，画像認識の分野で脚光を浴びている[5]．

ここでは，CNN を，直接に画像認識に用いるのではなく，**3.**で得られた 2 値化候補画像に対する文字・非文字の 2 クラス分類に適用する．

4.1. CNN 学習用データの作成

CNN 学習用データセットを構築するために，*Img* データセットの学習用画像から得られた 2 値化候補画像を目視で「Correct」，「Morphed」，「Incorrect」の 3 種類に分類する．それぞれの定義は以下の通りである．

<u>Correct</u>

文字画像を最もよく表している最適 2 値化画像．CNN 学習用の文字データとする．

<u>Morphed</u>

Correct には選ばれなかった正しい 2 値化画像，または欠けなどによって他の文字に見えてしまう 2 値化画像．CNN 学習用のデータとしては用いずに除外する．

<u>Incorrect</u>

文字領域が著しく欠けている，あるいはノイズが多い，正しくない 2 値化画像．CNN 学習用の非文字データとする．

CNN 分類器に入力するために，双一次補間法によってサイズを28×28 に縮小する．

図 3 に，CNN による文字・非文字分類の学習用データの例を示す．

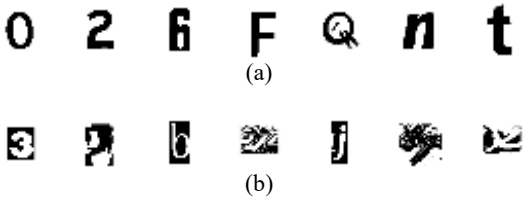

図3. CNN による文字・非文字分類の学習用データ．(a) 文字データ．(b) 非文字データ．

4.2. CNN の学習

CNN の実装には TensorFlow を用いた．

図 4 に，ネットワークの構成を示す．

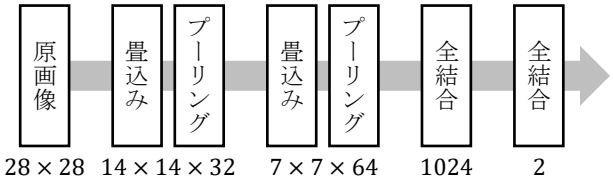

図4. 文字・非文字分類用の CNN 構成.

RGB 色空間を用いた場合の文字データと非文字データの学習サンプル数をそれぞれn_c, n_{nc}と記すと，n_{nc}はn_cの約 12 倍であった．

上記不均衡に対処するため，CNN の学習にあたっては，より少ない学習サンプル数n_cを学習させることを「1 ステップ」と定義することとした．

4.3. 最適 2 値化画像の選択

まず，学習用データに対する文字・非文字分類率が収束するまでステップ単位での学習を行う．

次に，得られた CNN 分類器を用いて *Img* データセットのテスト用画像から得られた全ての 2 値化候補画像に対して文字・非文字の分類を行う．

最後に，全ての 2 値化候補画像についての「文字クラスの出力値」が高い順にソートし，最上位となった画像を最適 2 値化画像として選択する．

5. 変形耐性画像マッチングによる文字認識

最適 2 値化された文字画像について，62 カテゴリ分類による文字認識を行う．

本研究では，変形耐性テンプレートマッチングを適用する．具体的には，アフィン変換耐性を有する GAT 相関法[6]，および 2 次元射影変換耐性を有する GPT 相関法[7]を用いる．いずれも相関ベースのテンプレートマッチング手法である．

5.1. テンプレート画像の前処理

テンプレート画像は，**4.**で「Correct」として選ばれた 2 値化画像，および *Fnt* データセットである．カテゴリごとに 1,031 枚となる．フォント画像については 2 値化したのち，**3.3** と同様の正規化処理を施す．

処理量を削減するために，テンプレート画像は，双一次補間法によってサイズを28×28 に縮小する．

5.2. GAT 相関法の適用

GAT 相関法[6]は，入力画像に対して最適なアフィン変換を施して，テンプレート画像との正規化相互相関値を最大化する画像マッチングの手法である．

入力画像Fの画素(i,j)での濃淡値を$f(i,j)$，テンプレート画像Gの画素(i,j)での濃淡値を$g(i,j)$ と記す．ただし，画像サイズを$M \times N$とする．

まず，次式を満たすように$f(i,j), g(i,j)$を線形変換する．

$$\sum_{j=1}^{N}\sum_{i=1}^{M} f(i,j) = 0, \sum_{j=1}^{N}\sum_{i=1}^{M} g(i,j) = 0,$$
$$\sum_{j=1}^{N}\sum_{i=1}^{M} f(i,j)^2 = 1, \sum_{j=1}^{N}\sum_{i=1}^{M} g(i,j)^2 = 1. \tag{2}$$

この操作により，2 枚の画像F, G の正規化相互相関値は内積$(f,g) = \sum_{j=1}^{N}\sum_{i=1}^{M} f(i,j)g(i,j)$ に等しくなる．

次に，正規化相互相関値が最大となる最適なアフィン変換$A, \boldsymbol{b}$ を求める．ただし，A は回転・伸縮・せん断を表す2×2行列，$\boldsymbol{b}$ は平行移動を表す 2 次元ベクトルであり，次式で与えられる．

$$A = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}. \tag{3}$$

これより，次の最適化問題を解くことになる．

$$C^{\mathrm{GAT}}(f,g) \equiv \max_{A,\boldsymbol{b}} J_{\mathrm{GAT}}(A,\boldsymbol{b}),$$
$$J_{\mathrm{GAT}}(A,\boldsymbol{b}) = \sum_{j=1}^{N}\sum_{i=1}^{M} f(i,j)g(i',j'), \tag{4}$$
$$i' = a_{00}i + a_{01}j + b_0, j' = a_{10}i + a_{11}j + b_1.$$

ここで，$A, \boldsymbol{b}$の網羅的探索を避けるために，Gauss 型カーネル関数を用いた新たな目的関数$\tilde{J}_{\mathrm{GAT}}(A,\boldsymbol{b})$を導入する．

$$\tilde{J}_{\mathrm{GAT}}(A,\boldsymbol{b}) = \sum_{j=1}^{N}\sum_{i=1}^{M}\sum_{j''=1}^{N}\sum_{i''=1}^{M} G(A,\boldsymbol{b},i,j,i'',j'')f(i,j)g(i'',j''),$$
$$G(A,\boldsymbol{b},i,j,i'',j'') = \exp\left(-\frac{(i''-i')^2 + (j''-j')^2}{2W^2}\right). \tag{5}$$

$\tilde{J}_{\mathrm{GAT}}$ 最大化の必要条件から，$\tilde{J}_{\mathrm{GAT}}$ を$A, \boldsymbol{b}$ の各成分で偏微分した値を 0 とすると，非線形連立方程式が得られる．さらに 0 次近似を適用すると，$A, \boldsymbol{b}$ に関する連立一次方程式が得られる．これは容易に解けるが，$A, \boldsymbol{b}$ の準最適解となっている．このため，準最適解の逐次代入による反復解法を適用する．これにより正規化相互相関値$C^{\mathrm{GAT}}(f,g)$を最大化する最適解$A, \boldsymbol{b}$ が求められる．

5.3. GPT 相関法の適用

GPT 相関法[7]は，GAT 相関法を拡張したもので，入力画像に対して最適な 2 次元射影変換を施して，テンプレート画像との正規化相互相関値$C^{\mathrm{GPT}}(f,g)$を最大化する画像マッチングの手法である．

2 次元射影変換は，式(3)に歪みを表す 2 次元ベクトル$\boldsymbol{c}$が加わり，次式に示す 8 パラメータ：

$$A=\begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \boldsymbol{b}=\begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \boldsymbol{c}=\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \tag{6}$$

で表現される．位置ベクトル$\boldsymbol{x}=(i,j)^{\mathrm{T}}$に対する 2 次元射影変換後のベクトルを$\boldsymbol{x}'$と記すと，$\boldsymbol{x}'$は

$$\boldsymbol{x}'=\frac{A\boldsymbol{x}+\boldsymbol{b}}{1+\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}} \tag{7}$$

で与えられる．

最適な 2 次元射影変換成分を求める目的関数は，GAT 相関法と同様に網羅的探索を避けるために，Gauss 型カーネル関数を用いて次式で定義する．

$$\begin{aligned} &\tilde{J}_{\mathrm{GPT}}(A,\boldsymbol{b},\boldsymbol{c}) \\ &=\sum_{j=1}^{N}\sum_{i=1}^{M}\sum_{j''=1}^{N}\sum_{i''=1}^{M} G(A,\boldsymbol{b},\boldsymbol{c},i,j,i'',j'')f(i,j)g(i'',j''), \\ &G(A,\boldsymbol{b},\boldsymbol{c},i,j,i'',j'') \\ &=\exp\left(-\frac{(i''-i')^2+(j''-j')^2}{2W^2}\right), \\ &\quad i'=\frac{a_{00}i+a_{01}j+b_0}{1+c_0i+c_1j}, j'=\frac{a_{10}i+a_{11}j+b_1}{1+c_0i+c_1j}. \end{aligned} \tag{8}$$

$\tilde{J}_{\mathrm{GPT}}$ 最大化の必要条件から，$\tilde{J}_{\mathrm{GPT}}$ を$A,\boldsymbol{b},\boldsymbol{c}$ の各成分で偏微分した値を 0 とすると，非線形連立方程式が得られる．さらに 0 次近似を適用すると，$A,\boldsymbol{b},\boldsymbol{c}$ に関する連立一次方程式が得られる．これを解くと$A,\boldsymbol{b},\boldsymbol{c}$ の準最適解が得られる．準最適解の逐次代入による反復解法により，正規化相互相関値$C^{\mathrm{GPT}}(f,g)$を最大化する最適解$A,\boldsymbol{b},\boldsymbol{c}$ を求める．

6. 最適 2 値化の実験結果

3.および**4.**で述べた文字画像の 2 値化を *Img* データセットのテスト用画像に適用し，正 2 値化画像の生成率，および最適 2 値化画像の選択率を調べた．

表 1 に，色空間ごとの正 2 値化画像の生成率と最適 2 値化画像の選択率を掲げる．ただし，選択率は正 2 値化画像が生成された文字画像について算出した値である．

表 1 より，Lab 色空間を用いた場合，2 値化画像の生成数は RGB 色空間のおよそ 6 分の 1 である．また，最適 2 値化画像の選択率については，Lab 色空間を用いた場合，RGB 色空間よりも 4%ほど高いことが分かる．

図 5 に，CNN による文字・非文字分類における，学習用データに対する分類率の推移を示す．

表1. 正 2 値化画像の生成率と最適 2 値化画像の選択率.

色空間	RGB	Lab
学習用 2 値化画像生成数	14,916	2,732
テスト用 2 値化画像生成数	15,704	2,476
正 2 値化画像の生成率	97.3%	96.9%
最適 2 値化画像の選択率	92.4%	96.6%

図 5 より，Lab 色空間を用いた場合はデータ数が少なく，早期に分類率が 100%に収束しているため，学習を 1,000 ステップで打ち切った．

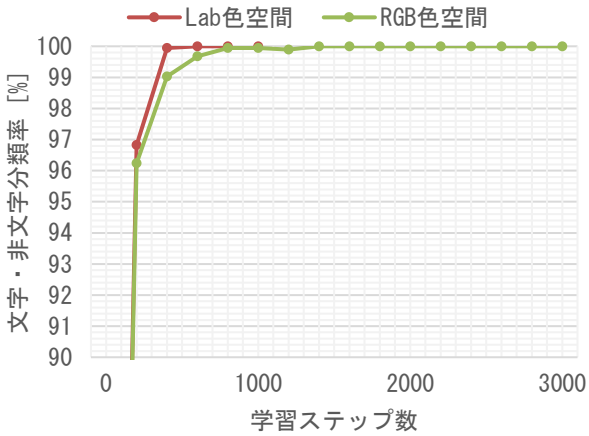

図5. CNN による文字・非文字分類での学習用データに対する分類率の推移.

7. 文字認識の実験結果

最適 2 値化画像として選択された 2 値化画像に対して，**5.**で述べた GAT/GPT 相関法をはじめとした種々の手法による認識実験を行った．

表 2 に，認識率一覧を掲げる．

表に掲げた各手法について以下に説明する．

単純相関法は，入力画像とテンプレートの間でそのまま正規化相互相関値を算出して，最大相関値を持つカテゴリに分類する手法である．

GAT 相関法_loose は，GAT 相関法における収束条件を緩め，逐次代入による反復解法での反復回数の上限を増した手法である．

改良投影距離法[8]は，ベイズアプローチから得られる疑似ベイズ識別関数を変形・近似することによって定義される距離を用いたマッチング手法である．特徴量には 196 次元の輪郭方向分布特徴を用いた[3].

CNN は，図 4 と同じ構成の CNN を用いて，62 カテゴリの認識を行った結果を示す．ただし，学習にはテンプレート画像のみを用いた．

表 2 より，Lab 色空間を用いた GAT 相関法_loose による認識実験で最高認識率 73.2%が得られていることが分かる．

RGB 色空間と Lab 色空間を比べると，CNN による認識における精度は，最適 2 値化画像選択率が高い Lab 色空間の方が高い．一方，GAT 相関法における認識率に差は

あまりない．これは，2 値化画像生成数が多い RGB 色空間においては文字領域が一定以上の割合で含まれている 2 値化画像が多く，選択された 2 値化画像が必ずしも最適でなくとも GAT 相関法によって救済されたテストサンプルが存在することが原因と考えられる．

また，GAT 相関法をはじめとしたテンプレートマッチング手法における認識率が改良投影距離法における認識率よりも高い．これは，学習サンプル数が少ない画像認識における変形耐性テンプレートマッチングの有効性を示している．一方，GPT 相関法における認識率は GAT 相関法よりも低い．これは，歪みへの耐性を付加したことにより過適合を生じたことが原因と考えられる．

表2. 手法ごとの認識率.

認識手法	認識率(%)	
	RGB 色空間	Lab 色空間
単純相関法	64.5	64.3
GAT 相関法	72.6	72.6
GAT 相関法_loose	73.2	73.3
GPT 相関法	71.5	71.9
改良投影距離法	63.5	63.9
CNN	70.9	71.9

単純相関法，GAT 相関法，GAT 相関法_loose，および GPT 相関法では，k-NN 分類を用いた．表 2 における認識率は，$k = 10$ としたときの値である．

最高認識率が得られた，Lab 色空間を用いた GAT 相関法_loose による認識実験において，k-NN 分類におけるkの値を$k = 1,3,5,7,9,10$ としたときの認識率を調べた．

表 3 に，k の値ごとの認識率を示す．

表 3 より，$k = 3$ のときに，最高認識率 73.9%が得られていることがわかる．

表3. GAT 相関法_loose での k-NN 分類における認識率とkの関係.

kの値	認識率(%)
1	73.7
3	73.9
5	73.1
7	72.8
9	72.6
10	73.3

表 2 に掲げた認識手法の一部については，テンプレートに *Fnt* データセットを含めずに，*Img* データセットの学習用画像のみを用いた場合についての認識率を調べた．テンプレート画像はカテゴリごとに 15 枚となる．

表 4 に，学習用画像のみをテンプレートとしたときの認識率を示す．表中，DirectCNN_gray, DirectCNN_color は，図 4 と同じ構成の CNN を用いて，文字画像そのものに対して直接に文字認識を行った場合を指す．DirectCNN_gray はグレースケールに変換した 1 チャネル文字画像，DirectCNN_color は 3 チャネル文字画像をそれぞれ入力とする．

表 4 より，テンプレートにフォント画像も用いた場合と比べると認識率が若干劣るものの，テンプレートが少数であっても一定の認識率が得られていることが分かる．一方，2 値化せずに CNN に直接，文字画像を入力した場合の認識率はあまり高くない．

RGB 色空間と Lab 色空間を比べると，最適 2 値化画像選択率が高い Lab 色空間の方が，認識率が低い．また，収束条件を緩めた GAT 相関法_loose の方が，GAT 相関法よりも認識率が低い．これより，テンプレートが不十分であり，安定した認識が行えていないと考えられる．

表4. 学習用画像のみをテンプレートとした認識率.

認識手法	認識率(%)	
	RGB 色空間	Lab 色空間
単純相関法	50.5	49.6
GAT 相関法	66.3	64.2
GAT 相関法_loose	65.2	64.2
CNN	62.6	61.5
DirectCNN_gray	41.3	
DirectCNN_color	31.6	35.3

8. 考察

8.1. 誤認識の要因分析

表 2 より，Lab 色空間を用いた GAT 相関法_loose による認識実験，および Lab 色空間を用いた CNN による認識実験において，誤認識した文字画像についての要因分析を行った．

表 5 に，失敗要因と該当した画像数を掲げる．表中，「原因不明」は，最適 2 値化画像が鮮明であるにもかかわらず誤認識したことを示す．

図 6 に，誤認識の「原因不明」の例を示す．

CNN による認識において「原因不明」に分類された画像は全て，GAT 相関法_loose では正しく認識されていることから，GAT 相関法が有する変形耐性が示されたといえる．

表5. 認識に失敗した要因別の画像数.

要因	画像数	
	GAT 相関法_loose	CNN
類似カテゴリへの誤分類	136	105
複数の文字が存在する	47	69
文字領域の欠損やノイズ	19	26
背景領域が最適 2 値化画像として選択された	16	16
文字が輪郭によって表現されている	13	11
文字が横向きである	9	8
文字領域の潰れ	3	4
アノテーションの間違い	5	5
原因不明	0	17
計	248	261

図6. 「原因不明」の例.

表 5 より，誤認識の主な要因は「類似カテゴリへの誤分類」，「複数の文字が存在する」の 2 つであることが分かる．以下に考察を加える．

類似カテゴリへの誤分類

“数字のゼロ”と“大小文字のオー”のように，形状が似ているカテゴリ間で誤分類されたことを示す．単文字認識において避けることは困難であり，文字列が単語として意味を成すかなどの言語知識を利用した認識によって解決できると考える．

複数の文字が存在する

文字画像に別の文字の一部分または全部分が含まれていることを示す．図 7 に例を示す．解決策としては，連結成分によって認識対象以外の文字領域を除去することなどが考えられる．

図7. 「複数の文字が存在する」の例.

8.2. 処理時間

表 2 に掲げた各認識手法の処理時間を調べた．使用した計算機の CPU は 3.60GHz Intel Core i7-4790 である．また，CNN による認識では GPU として NVIDIA GeForce GTX 760 を用いた．

表 6 に，認識処理にかかった時間を掲げる．値は 1 サンプル単位での平均とした．

表6. 手法ごとの 1 サンプルあたりの認識処理時間.

認識手法	処理時間(秒)
単純相関法	2.4
GAT 相関法	270.0
GAT 相関法_loose	1981.0
GPT 相関法	424.8
改良投影距離法	0.1
CNN	0.1

相関ベースの手法においては，各テンプレートとの相関値は個別に計算できるため，8 コアでの並列処理を行ったが，他の手法に比べ時間がかかる．ただ，GAT/GPT 相関法の処理時間短縮については，計算手順の効率化やテーブル参照の活用[9]，パラメータの初期探索の高速化[10]などの報告がある．

9. むすび

本論文では，クラスタ数の自動選択機能を持つ K-means クラスタリングと CNN を用いた情景内カラー文字画像の最適 2 値化，および変形耐性画像マッチングによる文字認識手法を提案した．情景内文字画像データセット“The Chars74K”を用いた評価実験により，正 2 値化画像選択率 96.6%，文字認識率 73.9%を達成した．

提案手法では 2 値化の指標として色空間を用いるため，文字と背景に同じ色が存在すると，両方とも文字領域として抽出されることがある．この場合，文字領域のみを抽出する手法の検討が必要である．また，GAT/GPT 相関法への高速化手法の適用も，今後の課題である．

文　献

[1] C. Shi *et al.*, “Fisher vector for scene character recognition: A comprehensive evaluation,” *Pattern Recognition*, vol. 72, pp. 1-14, 2017.

[2] T. E. de Campos, B. P. Babu, and M. Varma, “Character recognition in natural images,” *In Proc. of the Int. Conf. on Computer Vision Theory and Applications*, Lisbon, Portugal, Feb. 2009.

[3] 遠藤光太郎，若原徹，“クラスタ数の自動選択による K-means クラスタリングとサポートベクターマシンを用いた情景内カラー文字画像の 2 値化と認識”，信学技報, vol. 112, no. 386, pp. 323-328, 2013.

[4] R. G. Casey, “Moment normalization of handprinted characters,” *IBM J. Res. Develop.*, vol. 14, pp. 548-557, 1970.

[5] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, “End-to-end text recognition with convolutional neural networks,” *Proc. of the 21st Int. Conf. on Pattern Recognition*, pp. 3304-3308, Tsukuba, Japan, Nov. 2012.

[6] T. Wakahara, Y. Kimura, and A. Tomono, “Affine-invariant recognition of gray-scale characters using global affine transformation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 4, pp. 384-395, 2001.

[7] T. Wakahara, and Y. Yamashita, “Enhanced GPT correlation for 2D projection transformation invariant template matching,” *Proc. 37th German Conf. on Pattern Recognition*, pp. 435-445, Aachen, Germany, Oct. 2015.

[8] 韓雪仙，若林哲史，木村文隆，三宅康二，“ベイズアプローチによる最適識別系の有限標本効果に関する考察—学習標本の大きさがクラス間で異なる場合—”，信学論(D-II), vol. J82-D-II, no. 4, pp. 621-630, 1999.

[9] Y. Yamashita, and T. Wakahara, “Affine-transformation and 2D-projection invariant *k*-NN classification of handwritten characters via a new matching measure,” *Pattern Recognition*, vol. 52, pp. 459-470, 2016.

[10] 山下幸彦，若原徹，“領域画像マッチングのための初期探索法”，信学技報, vol. 118, no. 219, pp. 119-124, 2018.