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の2値化と認識

ENDO, Kotaro / 遠藤, 光太郎

(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)
8
(開始ページ / Start Page)
117
(終了ページ / End Page)
122
(発行年 / Year)
2013-03
(URL)
https://doi.org/10.15002/00009584

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の 2 値化と認識

Binarization and Recognition of Color Characters in Scene Images Using K-means Clustering and Support Vector Machines

遠藤　光太郎
Kotaro Endo
法政大学大学院情報科学研究科情報科学専攻
E-mail: kotaro.endo.4h@stu.hosei.ac.jp

Abstract

This paper proposes a new method for binalizing and recognizing color characters in scene images. The proposed method consists of the following five steps. (1) A set of candidate binarized images is generated via every dichotomization of K clusters obtained by K-means clustering in the RGB color space. The total number of candidate binarized images equals 2^K-2. (2) The optimal number of clusters in K-means application is automatically determined based on the sum of within-cluster variances. (3) We calculate mesh and weighted direction code histogram features from every binarized image and feed them into a support vector machine (SVM) to calculate the degree of "character-likeness." (4) The binarized image with the maximum value of "character-likeness" is output as an optimal binarization result. (5) The optimal binarized image is recognized via the modified projection distance. Experimental results made on the public color character image dataset "the Chars 74K" show that the proposed method achieves correctly binarized character image generation and selection rates of 97.6% and 85.6%, and the final character recognition rate of 63.1%, which are superior to those obtained by the conventional techniques.

1. はじめに

近年，デジタルカメラの性能は大きく向上しており，身近な携帯電話でも 1000 万画素を超えるものも登場している．そのため，デジタルカメラで撮影した画像中の文字をコンピュータによって認識する技術，すなわち情景内文字認識技術が実現することで翻訳やナビゲーションなど様々なサービスが可能になる．

情景内文字画像においては，文字部分・背景部分ともに多種多様な色や形状が見られ，照明条件等も安定していない．そのため，文字と背景とを分離する 2 値化処理の精度が，最終的な文字認識率を大きく左右するといえる．主な 2 値化手法としては，明度に着目する局所 2 値化[1]，色相の一定性に注目する色クラスタリング[2]がある．単一カラー文字を対象とした 2 値化の研究では，Yokobayashi 等[3]のカラー空間での大津基準[4]による 2 値化が挙げられる．しかし，これらの従来手法では複数の異なる色相から成る文字領域に対応できない．

そこで本研究では，クラスタ数の自動選択機能を持つ K-means クラスタリングによって 2 値化候補画像を複数生成し，サポートベクターマシン(SVM)を用いて文字・非文字の判定を行い，最適な 2 値化画像を選択することで，文字色や背景など，より複雑な文字画像に対応する．また，得られた 2 値化画像に対して改良投影距離法[5]を用いて高精度な文字認識精度を実現する．

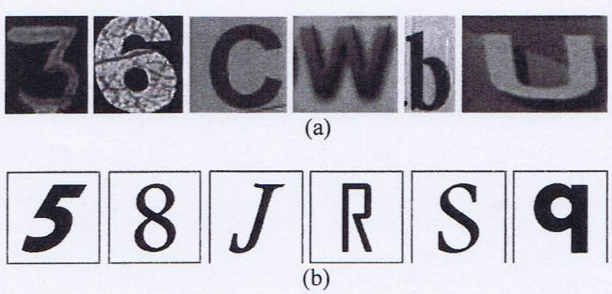
提案手法を情景内文字画像データセット the Chars 74K に適用した評価実験により，正 2 値化画像生成率 97.6%，SVM による正 2 値化画像選択率 85.6%，最終的な文字認識率 63.1%を達成した．

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.** でサポートベクターマシンを用いた 2 値化画像の決定，**5.** でカラー文字 2 値化の実験結果，**6.** で改良投影距離法を用いた文字認識と結果，について述べ，**7.** で先行研究との比較を示し，**8.** で考察を述べて，**9.** でむすびとする．

2. 使用した画像データ

本研究では，Campos らの作成した情景内文字画像データセット“the Chars 74K”[6]を使用した．0～9 までのアラビア数字 10 種類と，A～Z，a～z のアルファベット 52 種類の計 62 種類からなる画像データのうち，Campos らが撮影した 7705 枚の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各文字 1016 サンプルのフォント画像データ“*Fnt*”を用いて実験を行った．

図 1 に各々の画像例を示す．



(a)

(b)

図 1　文字画像データセット“the Chars 74K”．(a) 情景内カラー文字画像“*Img*”．(b) フォント画像“*Fnt*”．

Supervisor: Prof. Toru Wakahara

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 クラスタリングについて述べる．

3.1. 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 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 RGB カラー空間内の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．

ただし，K-means クラスタリングの結果は，K 個のクラスタ中心の初期値の選び方に依存する．そのため，本研究では K-means クラスタリングにマルチスタート探索を適用した．

マルチスタート探索とは，最適化問題の解が一般に初期値に強く依存してしまうため，初期値を複数設定してそれぞれの解を求めておき，それから最適解を選択する手法である．本研究の K-means 法においては 30 個の解の中からクラスタ内分散和の値が最小となるクラスタリング結果を採用した．

3. 2. K クラスタの 2 分割による 2 値化候補文字画像の生成

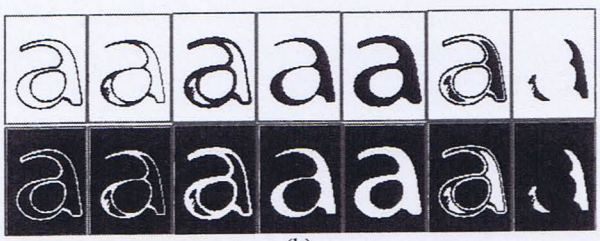
K 枚の分離画像群を網羅的に 2 つのグループに 2 分割して，一方を文字部分(黒)，他方を背景部分(白)として，複数の 2 値化候補文字画像を生成する．

上記 2 分割の全ての組み合わせにより生成される 2 値化候補文字画像の総数 N_{binary} は式(1)の通りである．

$$N_{binary} = \sum_{i=0}^{K-1} {}_K C_i = 2^K - 2 \quad (1)$$

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) 元のカラー画像．(b) 生成された 2 値化候補画像．

3.3. クラスタ数の自動選択

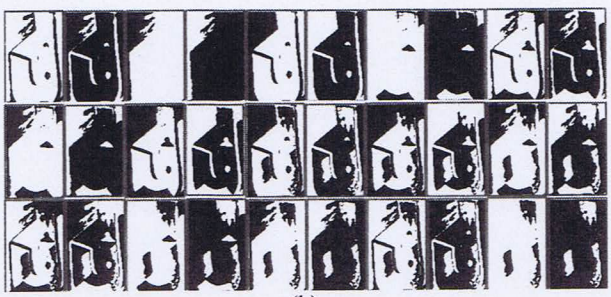
図 3 に示すように，情景内カラー文字画像の文字部分や背景部分の複雑さは，被写体によって大きく異なる．

図 3　様々な情景内カラー文字画像．

そのため，K-means クラスタリングによって 2 値化を行う際に，クラスタ数 K の値を固定すると，元の文字画像の複雑さによって，正しい 2 値化画像の生成率や候補画像の総生成数などに問題が生じる．例えば，K=5 に固定した場合，K=6 以上でなければ正しい 2 値化画像が生成されないような複雑な画像に対応できない．

図 4 に，クラスタ数が足りないため，文字部分と背景部分との分離が困難な画像の例を示す．

(a)



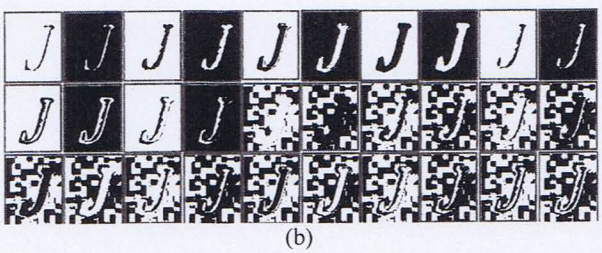
(b)

図 4　文字と背景の分離が困難な画像．(a) 元の画像．(b) 生成された 30 枚の 2 値化候補画像．

一方で K=4 で十分な単純な画像に対し，K=5 に固定すると必要以上の 2 値化候補画像を生成してしまう．

図 5 に，文字部分と背景部分との分離が容易にもかかわらず，必要以上の 2 値化候補画像を生成してしまう例を示す．

(a)



(b)

図 5　文字と背景の分離が容易な画像．(a) 元の画像．(b) 生成された 30 枚の 2 値化候補画像．

そこで本研究では，図 6 のようにクラスタ内分散和に対し閾値 T を設定することでクラスタ数の自動選択を行う．これにより，クラスタリングが容易な文字画像については，最小限の 2 値化候補画像のみを生成し，一方で複雑でクラスタリングが困難な文字画像については，より多くの 2 値化候補画像を生成することで，正 2 値化画像の生成率を高めることが可能となる．

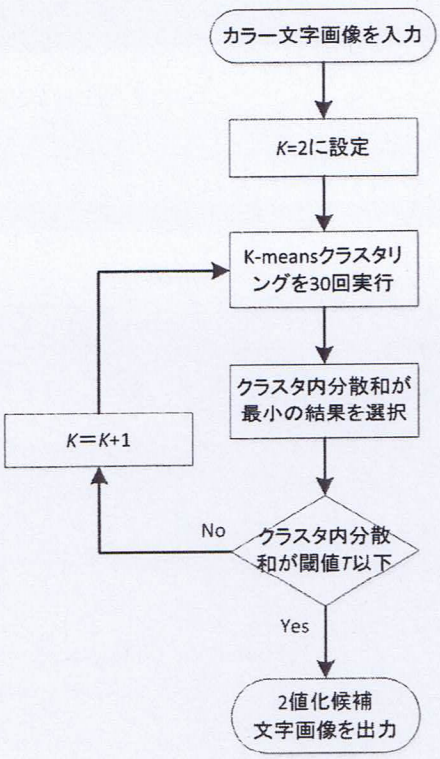


図 6　クラスタ数 K の自動選択．

なお，本研究では閾値 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 値化候補画像は最少で 2 枚，最大で 254 枚となる．

4. サポートベクターマシンを用いた文字・非文字の判定と 2 値化画像の決定

本章ではカラー文字画像から生成された 2 値化候補画像について，サポートベクターマシンを用いて文字か非文字かの判定を行う手法を提案する．

4.1. 画像の正規化

3.で得られた全ての 2 値化候補画像について，サイズを縦 120×横 80 に統一するため，文字部分の位置と大きさの正規化処理を施す．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(=25.0)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(40,60)に移動し伸縮率 s で大きさを正規化して，縦 120×横 80 のサイズとする．

4.2. 2 値化候補画像からの特徴量の抽出

正規化後の 2 値化候補画像から，文字・非文字判定のための特徴量の抽出を行う．ここでは，文字認識技術で従来よく用いられているメッシュ特徴と輪郭方向分布特徴を取り上げる．

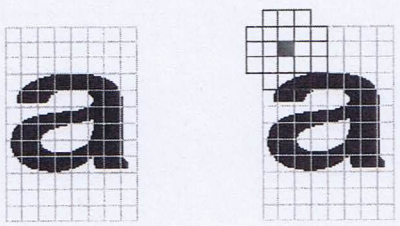
メッシュ特徴

サイズ 120×80 の画像を 96 個の正方ブロック(サイズ 10×10)に分割する．ブロックごとに黒画素数を計算して黒画素比率を算出する．それらを並べた 96 次元ベクトルをメッシュ特徴とする．

輪郭方向分布特徴

サイズ 120×80 の画像を 96 個の正方ブロックに分割する．ブロックごとに局所方向ヒストグラムを算出する．局所方向ヒストグラムに対し，2 次元ガウスフィルタをかけ，96 次元ベクトルへと次元圧縮したものを輪郭方向分布特徴とする．

図 7 に，メッシュ特徴と輪郭方向分布特徴それぞれの概念図を示す．



(a)　(b)

図 7　特徴量概念図．(a) メッシュ特徴．　(b) 輪郭方向分布特徴．

本研究では，上記の特徴に加え，両者を組み合わせた合計 192 次元の混合特徴も用いる．

4.3. サポートベクターマシン

サポートベクターマシン(SVM)は Vapnik らによって提案されたマージン最大化に基づく超平面による 2 クラス分類器である[7].

本研究では，the Chars 74K の *Img* データセットから K-means クラスタリングによって生成された 2 値化候補画像を，目視で「文字」と「非文字」の 2 種類に分類し，それぞれについて **4.2.**で述べた特徴量を抽出する．この特徴量を SVM の学習に用いることで SVM による「文字らしさ」の判定を行う．なお，本研究では SVM の実装に SVM^{light}[8]を用いた．

SVM の学習に用いたデータは以下の通りである．

SVM 学習用の文字データ

1. K-means クラスタリングによって生成された正しい 2 値化画像．
2. the Chars 74K の *Fnt* データセット(各文字 1016 サンプル)．

SVM 学習用の非文字データ

K-means クラスタリングによって生成された正しくない 2 値化画像．

図 8 に，SVM 学習用データの例を示す．

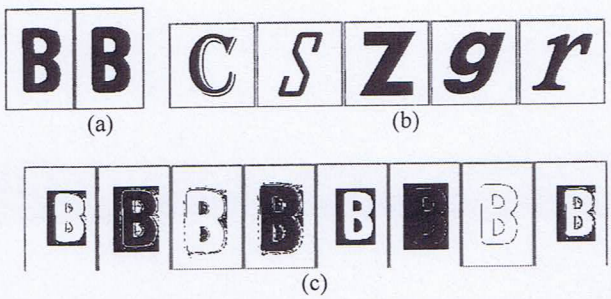


図 8　SVM 学習用の例．　(a) 文字画像(正しい 2 値化候補画像)．(b) 文字画像(*Fnt* データ)．(c) 非文字画像(正しくない 2 値化候補画像)．

ただし，2 値化処理の結果，他カテゴリに見えてしまう 2 値化候補画像が生成された場合，その画像については文字・非文字どちらの学習からも除外した．

図 9 に，学習から除外した 2 値化候補画像の例を示す．

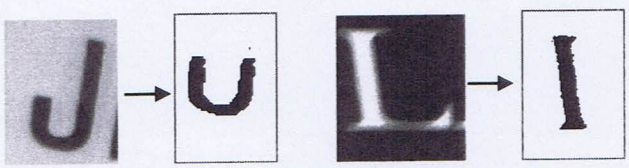


図 9　SVM 学習から除外した 2 値化候補画像．

4.4. サポートベクターマシンによる文字・非文字の判定と 2 値化画像の決定

K-means クラスタリングによって生成された全ての 2 値化画像に対し，前節の手法で学習した SVM を適用する．元のカラー画像 1 枚から生成した (2^K-2)枚の 2 値化候補画像を，SVM によって得られた「文字らしさ」順にソートする．そして最上位になった画像を最終的な 2 値化画像として選択する．

5. カラー文字 2 値化の実験結果

5.1. 正 2 値化画像の生成率

3.で述べた K-means による 2 値化で，クラスタ内分散和の値に応じて K の値を自動選択させた場合と，K=5 で固定した場合それぞれについて，生成された 2 値化候補画像中に正しい 2 値化画像が含まれているかどうか調べる．

表 1 に，2 つの手法の正 2 値化画像生成率，および平均の画像生成数を示す．

表 1 より，クラスタ数 K を変動させることにより，正しい 2 値化画像の生成率が向上し，一方で過剰な画像の生成を防ぐことが可能であることがわかる．

表 1　正 2 値化画像生成率と平均の画像生成数．

クラスタ数 K	正 2 値化画像の生成率(%)	平均の画像生成数 (枚)
自動選択	97.61	14
5 で固定	93.10	30

5.2. 正 2 値化画像の選択率

3.で生成した 2 値化候補画像について，SVM による文字･非文字判定の実験を 5-fold cross validation で行う．元のカラー画像を 5 分割し，それぞれ 2 値化候補画像を生成する．

表 2 に，学習とテストへのデータ分割の方法を示す．

表 2　5-fold cross validation によるデータの分割．

実験	学習データ	テストデータ
1	(B, C, D, E) + *Fnt*	A
2	(A, C, D, E) + *Fnt*	B
3	(A, B, D, E) + *Fnt*	C
4	(A, B, C, E) + *Fnt*	D
5	(A, B, C, D) + *Fnt*	E

SVM による文字･非文字の判定を行った結果，1 位に正しい 2 値化画像が選ばれているかどうか，すなわち正 2 値化画像の選択率を調べる．

表 3 に，メッシュ特徴，輪郭方向分布特徴，そして両者を合わせた混合特徴の計 3 種の特徴量の正 2 値化画像の選択率を示す．

表 3　SVM による正 2 値化画像の選択率．

特徴量	正 2 値化画像選択率(%)
メッシュ特徴(96 次元)	63.82
輪郭方向分布特徴(96 次元)	83.07
混合特徴(192 次元)	85.63

表 3 より，メッシュ特徴と輪郭方向分布特徴の混合特徴が最も選択率が高いことが分かる．そのため，今後の処理においては全て 192 次元の混合特徴を用いる．

5.3. 累積正 2 値化率

図 10 に，正しい 2 値化画像が含まれていたカラー文字画像に対し，4.で述べた手法を適用した結果の上位 20 位までの累積正 2 値化率を示す．

図 10 より，輪郭方向分布特徴および混合特徴については上位 3 位までを選択することで，95%以上の確率で正しい 2 値化画像を選べていることが分かる．また，3 位以降では両者の累積正 2 値化率は殆ど変わらないことも分かる．

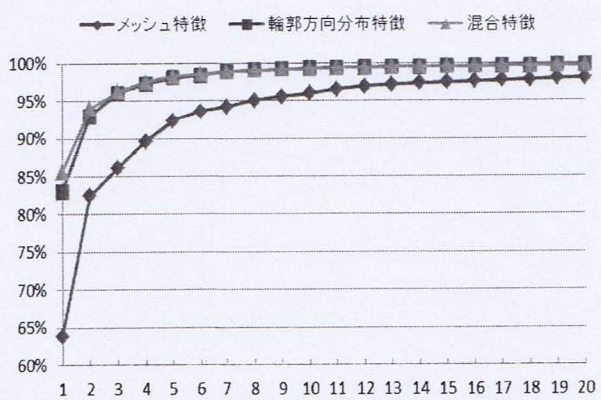


図 10　カラー文字の累積正 2 値化率．

6. 改良投影距離法を用いた文字認識

本章では，改良投影距離法とその基礎である擬似ベイズ識別関数について述べる．また，**4.**で選択した 2 値化画像に対して改良投影距離法による文字認識を行った結果を示す．

6.1. 擬似ベイズ識別関数

擬似ベイズ識別関数とは，ベイズアプローチから得られる最適識別関数を，計算量・記憶容量を削減するために変形して得られる識別関数であり，式(2)により定義される．

ここで，N は各クラスの学習標本の大きさ，N_0は特徴ベクトル X の事前分布の共分散行列の信頼度を表す定数，$P(\omega)$はクラスωの事前確率であり，λ_i，Φ_iはそれぞれ X の共分散行列の第 i 固有値と第 i 固有ベクトル，k は識別に用いる固有ベクトルの数，M はクラスの平均ベクトルである．

$$g(X) = (N + N_0 + 1)\ln\{1 + \frac{1}{N_0\sigma^2}[\|X - M\|^2 - \sum_{i=1}^{k}\frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i + \alpha\sigma^2}\left[\Phi_i^T(X - M)\right]^2]\} + \sum_{i=1}^{k}\ln\left((1-\alpha)\lambda_i + \alpha\sigma^2\right) - 2\ln P(\omega) \quad (2)$$

ただし，$\alpha = \frac{N_0}{N+N_0}$

6.2. 改良投影距離法

前節で定義した擬似ベイズ識別関数をさらに近似することを考える．すべてのクラスにおいて，共分散行列の行列式$P(\omega)$，N，N_0が等しいと仮定すると，式(2)の第 2 項，第 3 項は定数項となり，また第 1 項の係数$(N + N_0 + 1)$はクラス間で共通となる．この結果，次式(3)で定義される改良投影距離によって，式(2)の擬似ベイズ識別関数を近似することができる．

$$g(X) = \|X - M\|^2 - \sum_{i=1}^{k}\frac{(1-\alpha)\lambda_i}{(1-\alpha)\lambda_i + \alpha\sigma^2}\left\{\Phi_i^T(X - M)\right\}^2 \quad (3)$$

ここで，αは[0，1)の実数パラメータであり，識別率が最大となるように最適化する．σ^2は特徴要素の全カテゴリでの平均分散値である．

また，識別に用いる固有ベクトル数 kは，固有値の累積寄与率を用いて定めることとした．すなわち，次式で定義される固有値の累積寄与率η

$$\eta = \sum_{i=1}^{k}\lambda_i / \sum_{j=1}^{N}\lambda_j \quad (4)$$

が一定値βを超える最小の次元数を k とする．なお，本研究では $\beta = 0.95$として kの値を決定した．

6.3. 文字認識実験結果

The Chars 74K の *Img* データセットに **3.**および **4.**で述べた手法を適用し，得られた 7705 枚の 2 値化画像について，改良投影距離法による文字認識実験を行う．なお，学習には the Chars 74K の *Fnt* データセットを用いた．なお，予備実験より改良投影距離法のパラメータ$\alpha = 0.5$とした．

また，一部の文字カテゴリは，単独では他カテゴリとの区別は困難である．それらについて混同を認めた場合の認識率も調べる．混同を認めたのは以下の文字カテゴリである．

大文字と小文字の区別が困難なカテゴリ
{C, c}, {I, i}, {K, k}, {O, o}, {S, s}, {U, u}, {V, v}, {W, w}, {X, x}, {Z, z}

他の文字との区別が困難なカテゴリ
{0, {O, o}}, {1, {I, i}}, {1, l}, {{I, i}, l}

表 4 に，改良投影距離法による文字認識の平均認識率，および混同を認めた場合の平均認識率を示す．

表 4　改良投影距離法による文字認識率．

	認識率(%)
全ての文字を区別する	63.10
一部の文字を区別しない	76.08

また，混同を認めることで認識率が 20%以上向上した文字カテゴリについて，その認識率の変動を図 11 に示す．

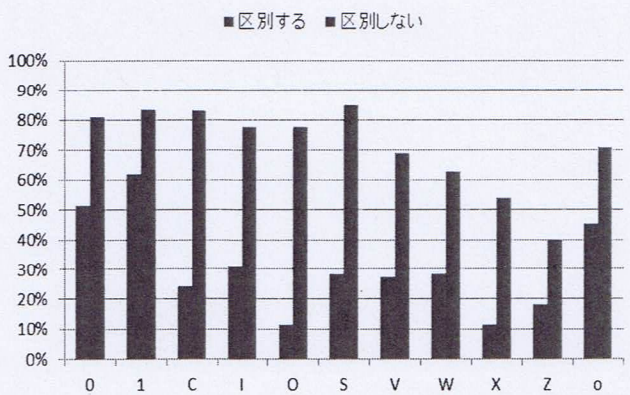


図 11　認識率が大きく向上した文字カテゴリ．

これらの文字カテゴリについては，単独の文字認識では限界があるため，文字列認識における前後の文字との比較や言語知識との照合によって，正確な認識が可能になると思われる．

7. 先行研究との比較

情景内文字画像データセット the Chars 74K は公開データベースであり，これを使用した先行研究がある．

表 5 に，the Chars 74K を用いた先行研究で報告された文字認識率と提案手法での達成値との比較を示す．

表 5 より，提案手法での認識率は必ずしも優位にあるとは言えないが，先行研究と本研究では学習・テストデータの扱いが異なり，一概に比較することは出来ない．

表 5　the Chars74K における文字認識率の比較．

手法	特徴値	認識率(%)
提案手法	メッシュ特徴+ 輪郭方向分布特徴	63.1
Campos ら[6]	Geometric Blur	54.3
Neumann ら[9]	MSER	71.6

8. 考察

本研究では，クラスタ数の自動選択機能を持つ K-means クラスタリングによって，図 12 に示すような，様々な複雑さを持つ情景内カラー文字画像について，正しい 2 値化画像を生成することに成功した．

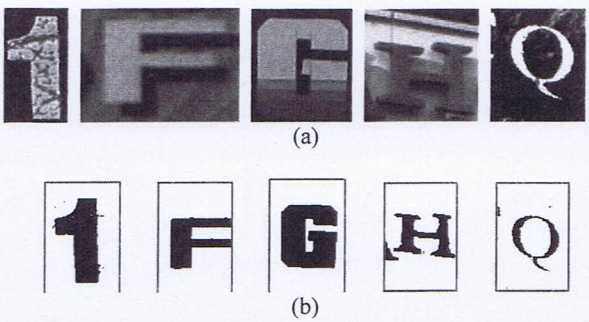


図 12　本手法により正 2 値化画像が成功した例．(a) 元のカラー画像．　(b) 生成された 2 値化画像．

しかし一方で，1 部の情景内カラー文字画像については，1 枚も正しい 2 値化画像が生成されなかった．図 13 にその元のカラー画像の例を示す．これらの画像については，文字部分の色が一定でなく，かつその一部が背景色と酷似しているので，RGB 空間情報のみではクラスタ数を増やしても分離が困難だったためと考えられる．

図 13　正 2 値化画像が生成されなかった例．

また，最適な 2 値化画像の選択において，本研究では，SVM による「文字らしさ」の評価にメッシュ特徴と輪郭方向分布特徴を合わせた特徴量を用いた．しかし，正しい 2 値化候補画像が生成されているにもかかわらず，明らかに非文字である画像が最終的な 2 値化画像として選択される場合もあった．図 14 にその例を示す．

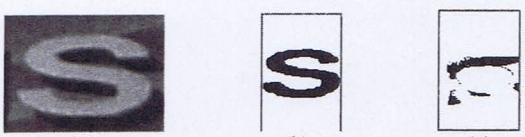


図 14　正しい 2 値化候補画像が選択されなかった例．(a) 元のカラー画像．　(b) 生成された正しい 2 値化候補画像．　(c)最終的に選択された 2 値化画像．

そのため，より適した特徴量を採用することで，正 2 値化画像の選択率を高めることが必要と考える．

更に，図 9 で示したように 1 部の画像は 2 値化候補画像を生成した際，他カテゴリに見える画像を生成してしまうことがある．これらの候補画像については「文字らしさ」によって除外することは困難である．そのため，1 位のみではなく，上位 n 位までの 2 値化候補画像を用いて文字認識を行う．そして，それぞれの認識結果に対して前後の文字と比較・結合を行い，文字列を形成する．こうして得られた文字列が，単語として意味をなすかなど，言語知識を利用した適切な評価基準を設けることで，本来の正 2 値化画像が有効となり，最終的な文字認識率を高めることが可能と考える．

9. むすび

本論文では，クラスタ内分散和に応じたクラスタ数自動選択機能を持つ K-means クラスタリングとサポートベクターマシンによる情景内文字画像の 2 値化，および改良投影距離法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 Chars 74K を用いた評価実験により，正 2 値化画像生成率 97.6%，SVM による正 2 値化画像選択率 85.6%，最終的な文字認識率 63.1%を達成した．

今後の課題は，文字部分の色が背景部分と酷似している画像にも対応可能な 2 値化手法の提案，文字らしさを評価する特徴量の工夫，そして複数の候補画像を用いた文字認識法の開発である．

文　献

[1] 大谷淳, 塩昭夫, “情景画像からの文字パターンの抽出と認識”, 信学論(D), vol. J71-D, no.6, pp. 1037-1047, June 1988.

[2] K. Wang and J. A. Kangus, “Character location in scene images from digital camera,” *Pattern Recognition*, vol. 36, no. 10, pp. 2287–2299, Oct. 2003.

[3] M. Yokobayashi and T. Wakahara. “Binarization and recognition of degraded characters using a maximum separability axis in color space and GAT correlation,” *Proc. of 18th Int. Conf. on Pattern Recognition*, vol. II, pp. 885-888, Hong Kong, Aug. 2006.

[4] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Trans. Systems, Man and Cybernetics*, vol. SMC-9, pp. 62-69, Jan. 1979.

[5] 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.

[6] T. E. de Campos, B. P. Babu, and M. Varma. “Character recognition in natural images,” *Proc. of the International Conference on Computer Vision Theory and Applications*, Lisbon, Portugal, February 2009.

[7] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Second Edition, Springer, 2000.

[8] T. Joachims, “Making large-scale SVM learning practical,” *Advances in Kernel Methods: Support Vector Learning*, B. Schölkopf, C.J.Burges, and A.J.smola (eds.). Chap. 11, MIT Press, 1998.

[9] L. Neumann, and J. Matas, “A method for text localization and recognition in real-world images,” *Proc. of 10th Asian Conference on Computer Vision*, Vol. III, pp. 770-783, Queenstown, New Zealand, Nov. 2010.