

自動ラベル付けによる情景文字画像データベースの構築

塚田 真規[†] 岩村 雅一[†] 黄瀬 浩一[†]

[†] 大阪府立大学大学院工学研究科 〒599-8531 堺市中区学園町 1-1

E-mail: tsukada@m.cs.osakafu-u.ac.jp, {masa,kise}@cs.osakafu-u.ac.jp

あらまし 多様な情景文字画像の認識には、多様なデータが登録された大規模文字データベースが必要となる。大規模データベースの構築には収集された文字画像に対して、文字領域の切り出し、データへのラベル付け、データベースへの登録という人の介入が必要な処理があり、膨大なコストがかかる。本稿では事例ベース文字検出手法と Semi-supervised learning に基づく自動ラベル付け手法の2つの手法を提案し、データベース構築に必要な処理を自動化することで構築に必要なコストを削減する。実験では各手法に対する評価実験と、2つの提案手法を組み合わせる文字領域の切り出しからデータベースへの登録までを自動化した統合システムについても評価した。

キーワード Segmentation by Recognition, Labeling by Recognition, 情景文字認識, Semi-supervised Learning, 大規模データセット

1. はじめに

スマートフォンなどの普及により、いつでも誰でも簡単にカメラが使えるようになった。これに伴い、文字認識を用いたアプリケーションが開発されており、翻訳システムや関連情報へのアクセスなどのサービスを提供している。しかし現在の文字認識技術では情景画像中の文字に対して認識可能な文字に限られているという問題がある。これは認識対象となる情景中文字がフォントや形状が多様であることが原因である。またカメラで撮影することによって射影変換、照明の変化、隠れ、文字の低解像度化など様々な要因が生じ、文字が劣化し、多様化する。この多様さが情景画像中文字の認識を困難にしている。

情景画像中の文字を認識するためには文字の多様性に対処する必要がある。これを解決する方法の一つとして事例ベース認識 [1, 2] がある。事例ベース認識はあらかじめ多様なデータを登録したデータベースからクエリに最も類似したデータを探索することで認識する方法である。この方法で多様なフォントや撮影状況の影響を受けた文字に対応するためには、大規模文字データベースが必要となる。

文字データベースの構築には、図1のように (i) 文字画像の収集、(ii) 文字領域の検出、(iii) 文字画像へのラベル付け、(iv) ラベル付き文字画像のデータベースへの登録、といった処理が必要である。しかしこれらの処理には基本的に人の介入が必要なため、膨大なコストがかかる。現在公開されている大規模文字データベースとして NEOCR [3] や Street View House Numbers Dataset [4] があるが、これらのデータベースの構築には人が介入し、膨大な手間やコストがかかっている。コストを抑えてデータベースを構築するためには各処理を自動化し、人の介入を最小限にする必要がある。

本稿では文字データベースの構築に必要な (ii) から (iv) までの処理を自動化し、人の介入を減らすことで、構築に必要なコ

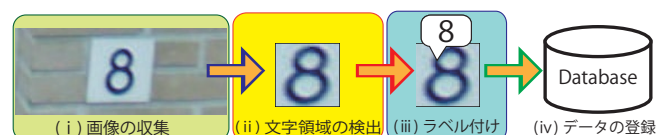


図1: 文字データベースの構築

ストを最小限に抑ええることを目的とする。(ii) を自動化するために事例ベース文字検出手法を、(iii) と (iv) を自動化するために Semi-supervised learning [5, 6] に基づく自動ラベル付け手法を提案する。2つの提案手法によってユーザーが撮影した画像から文字の検出、認識し、ラベル付けすることで順次データベースを拡大していく。これを繰り返すことによって、大規模データベースの構築と多様な文字が認識可能な識別器の作成することが本研究の目的である。

2. 関連手法

本節では提案する事例ベース文字検出手法に関連した手法として他の情景文字検出手法と自動ラベル付け手法に関連した手法として Semi-supervised learning について述べる。

2.1 文字検出手法

既存の情景文字検出手法はいくつかの種類があり、文字のテクスチャ情報を用いる手法 [7] や連結成分を求めることで文字領域を推定する手法 [8-11]、またこれらを統合した手法 [12] などがある。本稿では、これらの手法以外の局所特徴量を用いる手法 [2] を用いる。[2] では局所特徴量を用いることで、文字領域の検出とクラスの識別を同時に行い、情景画像に対しても頑健な文字検出・識別を可能としている。提案手法では [2] と同様に局所特徴量を用いて、文字領域を推定する。異なる点としては [2] ではクエリ上の文字とテンプレートを1対1対応で文字領域を検出するのに対して、提案手法ではクエリ画像の特徴

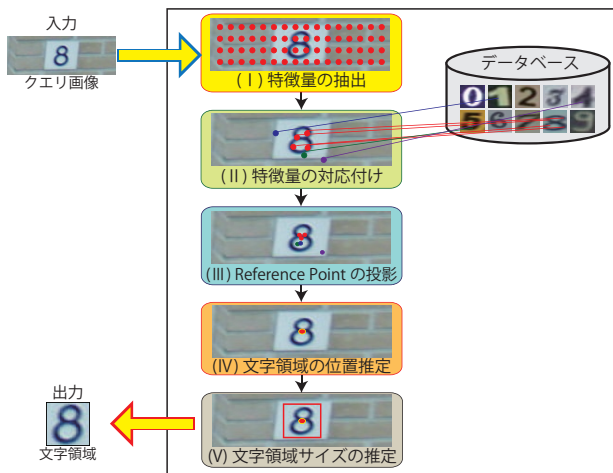


図 2: 事例ベース文字検出手法の概要

量と文字クラスとを対応付け、自動ラベル手法の認識と同じ方式で切り出すことで、ラベルなしデータを学習するたびに切り出し精度が向上することが期待できる。

2.2 Semi-supervised learning

Semi-supervised learning [5, 6] はラベルなしデータを学習に用いることでラベル付けのコストを削減し、識別器の性能の向上を目的としている。Semi-supervised learning の一種である Self-training は少量のラベル付きデータで識別器を学習した後、ラベルなしデータの認識結果をラベルと見なすことで、ラベルなしデータを学習する手法 [13] であり、手書き文字認識でも有効性が確かめられてきた [14]。

提案手法では Self-training の考えに基づき、ラベルなしデータへのラベル付けと学習を行う。Self-training を用いた理由は他の Semi-supervised learning の手法に対して特別な仮定を必要としないからである。本稿では多様な情景文字画像を対象とするため、他の手法の仮定を満たすことは難しい。これに対して特別な仮定を設けない Self-training に基づく手法を提案することで対応する。また Nearest Neighbor 法と組み合わせることによって逐次的な学習が容易である点が挙げられる。Nearest Neighbor 法では学習データをデータベースに登録するだけで良いので、計算コストをかけずに学習することができる。

3. 提案手法

3.1 事例ベース文字検出手法

提案する事例ベース文字検出手法の概要を図 2 に示す。以下では手順の詳細を説明する。

まずクエリ画像から (I) 特微量の抽出を行う。本手法では局所特微量である PCA-SIFT [15] を用いる。局所特微量は照明変化に頑健であり、局所領域から特徴を抽出するため、隠れなどにも対応できる。本手法では低解像度画像であっても一定数以上の局所特微量を抽出するため、一定間隔に特徴点を配置する Dense Sampling によって特徴点を決定する。

クエリ画像から特微量を抽出した後、各特微量に対してデータベースから最近傍点を探索し、(II) 特微量を対応付ける。データベースには文字領域が切り出され、正規化された画像から抽

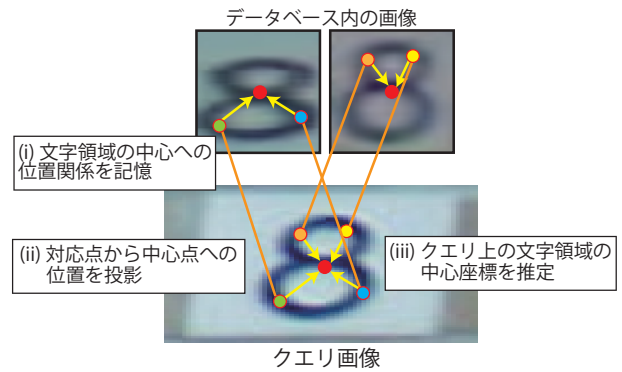


図 3: 文字領域の位置推定

出された特微量が登録されている。最近傍点の探索にはデータベース内に登録されている全ての特微量と距離計算をする必要があり、データ数が増加すると最近傍点の探索時間が膨大になるという問題がある、本手法ではこれを防ぐために近似最近傍探索 [16] を用いて、最近傍点を探索する。抽出された特微量の中には、本来と異なるクラスの特微量と対応付け、結果として検出や認識に悪影響を及ぼす特微量が存在する。このような特微量は最近傍点との距離 d_{nn} と第 2 近傍点との距離 d_{2nn} の差が大きくなると減少する傾向が見られるため、以下の不等式

$$\frac{d_{nn}}{d_{2nn}} < t_d \quad (1)$$

を満たすかを調べ、距離の差が大きい特微量のみを対応付ける。 t_d は閾値である。

特微量の対応付け後、対応付いた特微量を用いて文字領域の位置を推定する。図 3 にその概要を示す。本手法では文字領域の位置推定を (III) Reference Point (RP) [17] の投影により行う。正規化されたデータベース内の画像中心を RP と定義し、対応付いた各特徴点に対して最近傍点の RP をクエリ画像上に投影する。

次にクエリ上に投影された RP の分布を調べて (IV) 文字領域の位置を推定する。具体的な処理は以下の通りである。(i) RP に対して距離 d_c 以下にあり、かつ同じクラスに属する RP の数を調べる、(ii) 近傍にある RP の数が閾値 n_c 以上であるとき、クラスタを生成する、(iii) クラスタ内の各 RP に対して x 座標、 y 座標を調べ、その中央値を文字領域の中心座標とする、この手順によって文字領域の中心座標が求まり、クラスタのクラスを調べることで文字領域のクラスが求まる。

(V) 文字領域のサイズ推定の概要を図 4 に示す。まずクラスタから 2 つの特微量を選択し、クエリ上の x 座標の距離 d'_x と、それらのデータベース中に対応する特微量同士の x 座標の距離 d_x を求める。特微量の対応付けが正しいとき、 d'_x と d_x の比は文字サイズの x 軸方向のスケール比と等しくなる。正規化後の画像サイズを S とすると、検出した領域の x 軸方向の文字領域サイズ S'_x は $S'_x = S \times \frac{d'_x}{d_x}$ で求められる。クラスタ内の特微量の全ての組み合わせについて文字サイズを求め、その中央値を用いる。 y 軸方向についても同様の処理手順を行うことで検出した文字領域のサイズを求めることができる。

文字サイズを求め、文字領域を決定した後、同じクラスに属

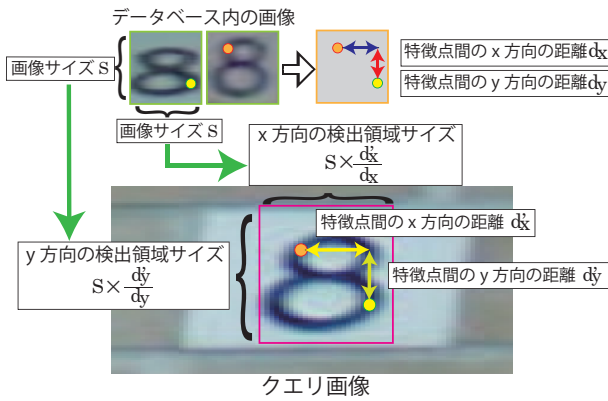


図 4: 文字領域のサイズ推定

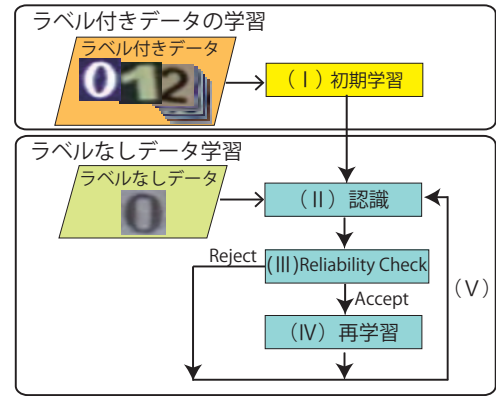


図 5: 自動ラベル付け手法の処理の流れ

し、かつ領域が重なっている場合、それらを含む最小の矩形を文字領域として再定義する。

3.2 自動ラベル付け手法

本手法ではラベルなしデータへのラベル付けと識別器の学習を同時に行い、大規模データベースと多様なデータを認識できる識別器の作成を目的とする。

自動ラベル付け手法の処理の流れを図 5 に示す。処理手順は以下の通りである。(I) ラベル付きデータを用いて初期学習を行い、識別器を作成する、(II) 識別器でラベルなしデータを認識し、クラスを推測する、(III) 推測したクラスラベルの信頼性を調べる、(IV) (III) の結果、クラスラベルが信頼できるとき、その結果をラベルなしデータのラベルと見なし、識別器を再学習する、(V) 別のラベルなしデータに対して (II) ~ (IV) を行う。本手法ではラベル付けと同時に学習を行うため、誤認識結果をラベル付けすると、識別器に悪影響を与える。これを防ぐために Reliability Check により認識結果を調べ、結果が信頼できないデータはラベル付けしない。

3.2.1 認識手法

本手法で局所特徴量と近似最近傍探索を用いた投票処理により文字画像を認識する。特徴点の決定には Dense Sampling を用いる。

画像から特徴量を抽出した後、近似最近傍探索を用いて近傍点を探索する。近似最近傍点を求めた後、そのクラス ID に対して投票する。3.1 節で述べたように抽出された特徴量の中には認識に有効でない特徴量も存在するため、式 (1) を満たす特徴量を選択し、投票に用いる。投票の際には $1/N_i$ の重みをつける。 N_i はデータベースに登録されているクラス i の特徴量数である。各特徴量に対して投票を行った後、得票数が最も大きいクラス ID を認識結果とする。

3.2.2 Reliability Check

Reliability Check ではラベルなしデータの認識結果を調べ、ラベル付け、再学習するかを決定する。本手法では認識結果をラベルなしデータのラベルと見なすため、誤認識時には誤ったクラスラベルが作成される。誤ったラベルを持つデータを学習してしまうと識別器に悪影響を与えてしまうため、正しく認識できたデータのみに対してラベル付けする必要がある。

本手法では認識時に求めた得票値を用いることによって、認識結果が正しいかを調べ、ラベルなしデータが信頼できるかを調べる。ラベルなしデータに対して最大となったクラスの得票値を s_1 、2 番目に大きいクラスの得票値を s_2 としたとき、

$$t_l < \frac{s_2}{s_1} < t_u \quad (2)$$

を満たす認識結果は信頼できると見なし、再学習対象とする。 t_l, t_u は下限値、上限値である。下限値を設けることで背景部分から誤検出された領域を取り除く。背景部分から抽出された領域では得票数が 1 位のクラス以外に対応付いた特徴量が少なかったため、経験的にある一つのクラスにのみ投票が集中している場合は背景領域と見なし、ラベル付けしない。

本手法ではラベル付けされたデータから抽出された特徴量をデータベースに登録する際、認識に寄与した特徴量のみを選択し、データベースに登録する。ラベル付けされたデータから特徴量を選択し、登録する方法を説明する。この方法ではラベル付けされ、再学習対象となったデータから抽出された特徴量に対して式 (1) を満たすか、またデータベース内の対応する特徴量のラベルと認識結果が等しいかを調べる。これらを同時に満たす特徴量のみをデータベースに登録する。

4. 実験

本節では実験により提案手法の性能を調査する。実験 1 では 3.1 節で提案した事例ベース文字検出手法に対する実験、実験 2 では 3.2 節で提案した自動ラベル付け手法に対する実験、実験 3 では事例ベース文字検出手法と自動ラベル付け手法を統合したシステムに対する実験を行う。

4.1 Street View House Numbers Dataset

本実験では情景数字文字画像で構成された Street View House Numbers Dataset [4] を用いる。本論文では解像度や色情報が撮影時の状態で保存されている Full Numbers フォーマットを用いる。画像を文字単位に切り出すときはデータセットと共に公開されている矩形情報を用いて、Full Numbers の画像から文字領域を切り出す。

4.2 実験 1: 事例ベース文字検出・認識手法の性能評価

本実験ではデータベース内のデータ数を変化させ、そのときの性能を調べる。ラベル付きデータは train サブセットから選

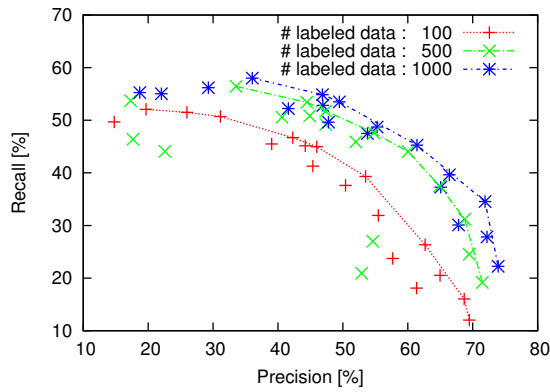


図 6: 実験 1. Recall-Precision の分布

び、文字領域を切り出した後、 96×96 [pixels] に正規化し、特徴点を 200 個抽出した。ラベル付きデータ数を各文字 100, 500, 1,000 と変化させたときの性能を調べる。クエリ画像には extra サブセットから抽出した画像 1,000 枚を用いる。 t_d, d_c, n_c の値はそれぞれ 0.9, 30, 35 とする。

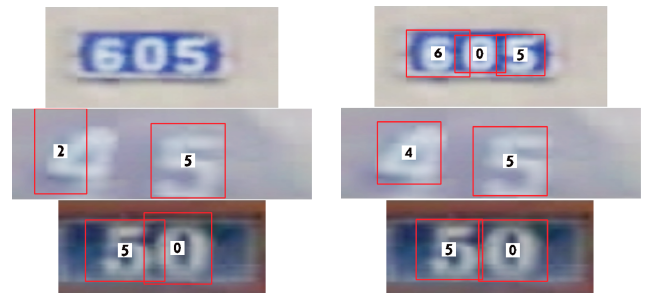
提案手法の性能評価には再現率 (Recall)、適合率 (Precision) を用いる。再現率は画像内に存在する文字領域中、正しく検出できた文字領域数の割合を、適合率は検出した文字領域中、正しく検出された文字領域数の割合を示す。本実験では文字領域の位置推定の用いるパラメータ d_c, n_c を変化させ、それぞれの場合について再現率と適合率を求める。文字が正しく検出できたかを調べる判定法は [18] を用いる。

実験結果を図 6 に示す。線で結ばれた点はパレート最適解である。このグラフではプロット点が右上にあるほど性能が良い。図 6 より、ラベル付きデータ数が増加するほど、プロット点の分布が右上に推移していることがわかる。このことからラベル付きデータ数を増加させることで性能が向上していることがわかる。図 7 はラベル付きデータ数が各文字 100 と 1,000 のときの認識結果の一例である。ラベル付きデータ数が少ないときに検出できなかった文字領域がデータ数を増加させることによって正しく検出できていることがわかる。文字領域のサイズもデータ数を増加することでより正しく求めることができ、図 7 のような低解像度画像であっても文字領域を検出できた。一方、図 8 は誤検出を含む失敗例である。失敗例として最も多かったのが図 8 のようにプレートの縁などを “1” と誤検出するケースであった。

4.3 実験 2: 自動ラベル付け手法の性能評価

実験 2 では 3 つの実験を行う。1 つ目の実験では 3.2 節で述べた投票に用いる特徴量を選択するときの影響と再学習時にデータベースに登録する特徴量を選択するときの影響を調べる。2 つ目の実験では提案手法の Reliability Check の性能について調査する。3 つ目の実験では入力するラベル付きデータ数とラベルなしデータ数の比率を変化させたときの影響を調査する。

本実験ではあらかじめ切り出され、正規化されたデータを対象とする。ラベル付きデータ数は各文字 10 ずつとする。ラベルなしデータ数は各文字 10,000 とし、train と extra サブセッ



(a) 各文字 100 のとき

(b) 各文字 1000 のとき

図 7: 実験 1. ラベル付きデータ数を変化させたときの検出例

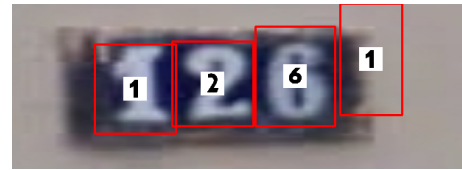


図 8: 実験 1. 検出の失敗例

トより選択する。識別器の性能を評価するために test サブセットを用いて、認識率を求める。本実験では t_d, t_l, t_u をそれぞれ 0.9, 0, 0.2 とする。

4.3.1 特徴選択による影響の調査

実験結果を表 1 に示す。初期学習率はラベル付きデータのみに学習した後の認識率、再学習後認識率はラベルなしデータを用いて学習した後の認識率、ラベル付けデータ数は式 (2) を満たしラベル付けされたラベルなしデータ数、ラベル付けデータ精度はラベル付けされたデータのうち正しくラベル付けされたデータの割合である。表中の手法の ‘a. 選択なし’ はラベルなしデータから抽出された全ての特徴量を投票、再学習に用いたとき、‘b. 投票時選択’ は投票に用いる特徴量を式 (1) で選択したとき、‘c. 学習時選択’ はデータベースに登録する特徴量を式 (1) と最近傍点のラベルを調べることで選択したとき、‘d. 投票・学習時選択’ は投票時と再学習時に用いる特徴量を選択する方法である。

まず投票時の特徴選択による影響を調べる。投票に用いる特徴量を選択した b と d は全ての特徴量を用いた a と c に比べて再学習後の認識率が向上した。b と d において初期認識率が低いのはデータベース内の学習が十分に進んでいないため類似した特徴量が見つからず、式 (1) を満たす特徴量が少なかったためである。その結果、投票に用いられる特徴量数が減少し、誤認識が起こり易くなった。その後ラベルなしデータの学習することで類似した特徴量を見つけやすくなり、有効な投票が増えたことで b と d の認識率は a と c に比べて大きく向上した。この結果より、認識時に投票に用いる特徴量を選択することは認識率の向上において有効であることがわかった。

データベースに登録する特徴量を選択したときの影響について調べる。再学習時に特徴選択を行わなかった a, b と特徴選択を行った c, d を比較すると、ラベル付けデータ数、ラベル

表 1: 実験 2-1 . 自動ラベル付け手法の実験結果

手法	a. 選択なし	b. 投票時選択	c. 学習時選択	d. 投票・学習時選択
初期認識率 [%]	46.3	44.8	46.3	44.8
再学習後認識率 [%]	42.7	50.8	50.3	64.7
ラベル付けデータ数	41,360	46,834	30,624	33,531
ラベル付けデータ精度 [%]	63.7	68.2	81.4	89.0

表 2: 実験 3 . 統合システムの実験結果

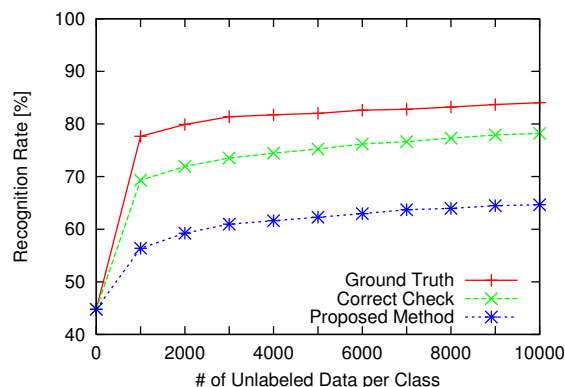
初期認識率 [%]	77.5
再学習後認識率 [%]	78.8
ラベル付けデータ数	4,037
ラベル付けデータ精度 [%]	73.0

付けデータ精度に関して大きな差が生じた。特徴選択を行わないときに、多くのラベルなしデータをラベル付けできた。一方、特徴選択を行ったときはラベル付けされたデータ数は少なくなったが、ラベル付けデータ精度は高くなった。これは特徴を選択することでデータベースに登録される特徴量数が減少してしまうが、正しくラベル付けされ、認識に有効な特徴量をデータベースに登録できていることがわかる。

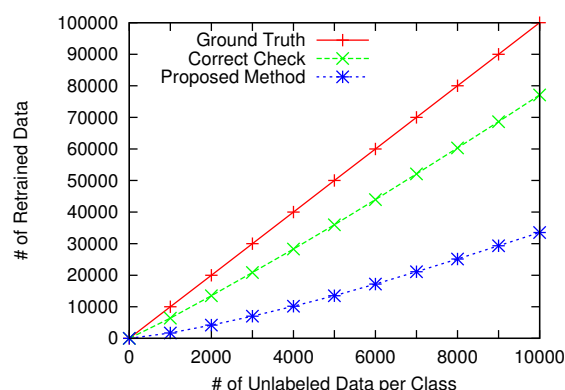
実験結果より提案した投票時と再学習時に特徴量を選択する手法が認識率とラベル付けデータ精度を向上させるために有効であることがわかった。提案手法はラベルなしデータの学習時にデータベースに登録する特徴量を選択するためラベル付けされるデータ数は減ってしまうが、ラベル付けデータ精度を高く保ったままラベル付けが行えることがわかった。データベースの作成には誤ってラベル付けされたデータを少なくする必要があるのであるため、提案手法が最も有効であると言える。

4.3.2 Reliability Check の性能調査

図 9(a) はラベルなしデータ数を各文字 1,000 から 10,000 まで 1,000 ずつ変化させたときの提案手法、全てのラベルなしデータを真のクラスで学習したとき (以下, Ground Truth), 認識結果が真のクラスと等しいときに学習したとき (以下, Correct Check) の認識率である。Correct Check は提案手法と同様に最近傍点のクラスと認識結果が等しく、式 (1) を満たす特徴量のみをデータベースに登録する。正しく認識されたときのみ学習するため、Reliability Check によるラベル付けデータの選択が完全に正しいときの認識率となる。Ground Truth と Correct Check の認識率の差は識別器の性能の差である。またデータベースに登録する特徴量を選択したため、学習時に登録される特徴量が減り、学習が進まなくなったことも認識率に差が生じた原因の一つである。次に Correct Check と提案手法の差であるが、これは Reliability Check の性能差である。これらの手法に認識率の差が出た理由として学習に用いたラベルなしデータのラベル付け精度とデータ数にある。誤ってクラスでラベル付けされたデータを学習に用いることによって認識精度に悪影響を及ぼすため、ラベル付けデータ精度を高くする必要がある。またラベル付けされたデータ数が多いほど、多様な文字がデータベースに登録されるため認識率は向上する。図 9(b) は図 9(a) においてラベル付けされたデータ数を示している。これより提案手法の Reliability Check では学習すべきデータを十分に学習できていないことがわかる。入力するラベルなしデータ数を増やすことによって、学習に用いたデータ数も増加し、認識率がより上昇していることがわかる。このことからラベル付けされたデータ数が識別器の認識精度に大きく影響し



(a) ラベルなしデータ数と認識率の関係



(b) ラベルなしデータ数と学習したデータ数の関係

図 9: 実験 2-2 . ラベルなしデータ数を变化させた時の結果

ていることがわかる。

4.3.3 ラベル付きデータ数とラベルなしデータ数の比率による性能調査

図 10 は入力するラベル付きデータ数とラベルなしデータ数の比率を变化させたときの結果である。横軸はクラス毎に入力したラベル付きデータ数とラベルなしデータ数の和である。このグラフより、入力されたデータ数が同じでも、ラベル付きデータが多い方が認識率が高いことがわかる。これはラベルなしデータに対してラベル付きデータの方が情報量が多いためである。ラベルなしデータを学習に用いることで、ラベル付きデータのみを用いたときに比べて認識精度を上げることができるが、ラベル付きデータより情報量が少なく、学習に用いられるデータ数も限られるため、認識率の上昇の限界が存在する。

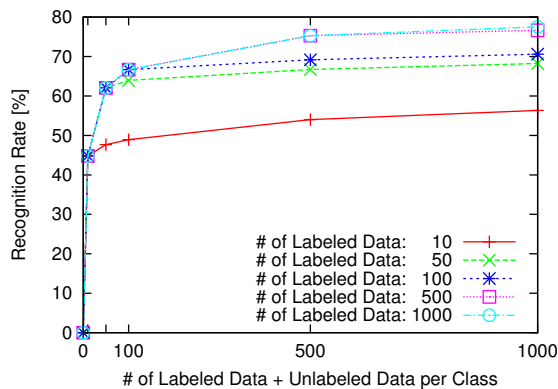


図 10: 実験 2-3. ラベル付きデータ数とラベルなしデータ数の比率を変化させたときの認識率

4.4 実験 3: 検出と自動ラベル付けの統合

実験 3 では上記の 2 つの提案手法を統合し、クエリ画像から文字領域の検出、認識、再学習の自動化に挑戦する。本システムでは事例ベース文字検出手法にて文字領域を検出した後、Reliability Check にてラベル付けするか決定する。Reliability Check では式 (2) を満たし、さらに 2 つの提案手法の認識結果が等しいかを調べる。検出した文字領域がこれら 2 つの条件を満たしたとき、その文字領域をラベル付ける。データベースへは文字検出時に生成したクラス内の特徴量を登録する。

実験に用いたデータについて説明する。train サブセットから各クラス 1,000 ずつ選択し、ラベル付きデータとしてデータベースに登録する。クエリ画像は extra サブセットから選択した 5,000 枚の画像を入力する。クエリ画像に含まれる文字数の合計は 13,215 である。識別器の性能評価のために test サブセットを用い、認識率を調べた。実験に用いたパラメータ t_d , d_c , n_c , t_l , t_u の値はそれぞれ 0.9, 30, 55, 0.3, 0.7 とする。

実験結果を表 2 に示す。ラベルなしデータから文字領域を検出、再学習をすることで約 1.3% 認識率が向上した。新たに文字領域を切り出し、ラベル付けされたデータ数は 4,037 であり、そのうち正しくラベル付けされた文字数はこのうち 73% である 2,947 であった。この結果より提案手法を組み合わせることによって識別器の性能を向上させ、人の介在なしにデータベースを拡大できることがわかった。しかしラベル付けされたデータには誤りも含んでいるため、更なる改善が必要である。実験 2 と異なり、文字でない領域も入力されるため、より正確な Reliability Check が必要である。Reliability Check が正確であっても検出結果が誤っていると学習可能なデータ数が減ってしまうため、文字検出、自動ラベル付けの両手法で精度を高める必要がある。

5. まとめ

本稿では大規模文字データベース構築のための処理を自動化のために事例ベース文字検出手法と自動ラベル付け手法を提案した。実験より事例ベース文字検出手法が低解像度画像に対しても正しく文字領域を検出、認識でき、自動ラベル付け手法ではラベルなしデータを精度高くラベル付けをすることができ、

認識精度を向上させることができた。また提案手法を組み合わせることで、文字領域の検出からデータベースへの登録までの一連の作業の自動化し、実験より、人の介在なしにデータベースを拡大させ、認識率を向上させることができた。

今後の課題として検出・認識手法の精度向上が挙げられる。自動ラベル付け手法で再学習データを決定する Reliability Check を改善し、誤ったラベルを持つデータの登録を減らすことも今後の課題である。

謝辞 本研究の一部は JST CREST の補助を受けた。ここに記して感謝する。

文献

- [1] M. Iwamura, T. Tsuji, and K. Kise, "Memory-based recognition of camera-captured characters," Proc. DAS, 2010.
- [2] M. Iwamura, T. Kobayashi, and K. Kise, "Recognition of multiple characters in a scene image using arrangement of local features," Proc. ICDAR, 2011.
- [3] A.D. Robert Nagy and K. Meyer-Wegener, "NEOCR: A configurable dataset for natural image text recognition," Proc. CBDAR, 2011.
- [4] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A.Y. Ng, "Reading digits in natural images with unsupervised feature learning," NIPS Workshop on Deep Learning and Unsupervised Feature Learning, 2011.
- [5] O. Chapelle, B. Schölkopf, and A. Zien eds., Semi-supervised learning, Cambridge, MIT Press, Sept. 2006.
- [6] X. Zhu and A.B. Goldberg, Introduction to semi-supervised learning, Morgan and Claypool Publishers, Sept. 2009.
- [7] J.-J. Lee, P.-H. Lee, S.-W. Lee, A.L. Yuille, and C. Koch, "Adaboost for text detection in natural scene," Proc. IC-DAR, 2011.
- [8] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," Proc. CVPR, 2010.
- [9] P. Sanketi, H. Shen, and J.M. Coughlan, "Localizing blurry and low-resolution text in natural images," Proc. IEEE Workshop on Applications of Computer Vision, 2011.
- [10] C. Yao, Z. Tu, and Y. Ma, "Detecting texts of arbitrary orientations in natural images," Proc. CVPR, 2012.
- [11] L. Neumann and J. Matas, "Real-time scene text localization and recognition," Proc. CVPR, 2012.
- [12] Y.-F. Pan, X. Hou, and C.-L. Liu, "A hybrid approach to detect and localize texts in natural scene images," IEEE Trans. on Image Processing, vol.20, no.3, pp.800-813, March 2011.
- [13] M. Tsukada, M. Iwamura, and K. Kise, "Expanding recognizable distorted characters using self-corrective recognition," Proc. DAS, 2012.
- [14] V. Frinken, M. Baumgartner, A. Fischer, and H. Bunke, "Semi-supervised learning for cursive handwriting recognition using keyword spotting," Proc. ICFHR, 2012.
- [15] Y. Ke and R. Sukthankar, "PCA-SIFT: a more distinctive representation for local image descriptors," Proc. CVPR, pp.506-513, 2004.
- [16] 佐藤智一, 岩村雅一, 黄瀬浩一, "空間インデクシングに基づく距離推定を用いた高速かつ省メモリな近似最近傍探索手法," 信学技報, pp.73-78, PRMU2012-142, Feb. 2013.
- [17] M. Klinkigt and K. Kise, "Using a reference point for local configuration of sift-like features for object recognition with serious background clutter," IPSJ Trans. on Computer Vision and Applications, vol.3, pp.110-121, Dec. 2011.
- [18] C. Wolf and J.-M. Jolion, "Object count/area graphs for the evaluation of object detection and segmentation algorithms," IJDAR, vol.8, no.4, 2006.