LI-6 単語画像生成モデルに基づく低品質印刷単語認識

A Word Image Generation Model for Holistic Word Recognition

石寺 永記† サイモン・ルーカス‡ アンディー・ダウントン‡
Eiki Ishidera Simon Lucas Andrew Downton

1．はじめに

博物館等に保存されている古い文書を電子化することは重要である。エセックス大学とロンドン自然史博物館は、蝶や蛾の膨大な標本を管理するためのアーカイブ・カードを電子化し、世界中の生物学研究者にその情報をインターネットで公開するシステムを試作している[1]。このアーカイブ・カードは、20 世紀初頭から現在までタイプライターを用いて作成/更新され続けてきたために、インクのかすれや文字の接触/つぶれを多く含んだ低品質文書であり、単語画像から文字を一文字ずつ切り出すことが難しいという問題がある。そこで、本論文では単語画像生成モデルを用いた単語認識方式を提案する。

2．単語画像生成モデル

本章では単語画像生成モデルを定義する。ここでは、筆記者が予め与えられた単語の綴りを予め決められた範囲に記入し、フォントの種類等も予め与えられているとする。単語画像生成モデルは四つのパートから成る（図１）。

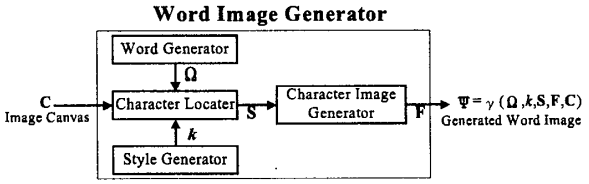

図１．単語画像生成モデル

Word Generator は、これから紙に記入される単語の綴り $\Omega=\{\omega_1,\omega_2,\ldots,\omega_n\}$を与え、単語の出現確率 $P(\Omega)$は一定と仮定する。ここで、Ωは n 個の文字から成り、ω_iは i 番目の文字のカテゴリーである。本論文では、Ωは予め辞書に登録されたものだけが記入される場合を考える。

Style Generator はフォントの種類や太さ等のスタイル k を与え、あるスタイルの出現確率 $P(k)$は一定と仮定する。

Character Locater は、Image Canvas **C**（予め決められた単語を記入するための範囲）と単語の綴りΩとスタイル k の情報から個々の文字の位置とサイズをセグメント情報 $\mathbf{S}=\{s_1,s_2,\ldots,s_n\}$として決定する。ここで、i 番目のセグメント $s_i=\{w_i,h_i,x_i,y_i\}$は矩形情報で、幅と高さ(w_i,h_i)と矩形中心の座標(x_i,y_i)から成る。また、i 番目のセグメントは(i-1)番目のセグメント s_{i-1} と ω_i、ω_{i-1} とも相関があると仮定すると、セグメント情報 **S** の出現確率は次のように書ける。

$$P(\mathbf{S}\mid\Omega,k,\mathbf{C})=P(s_1\mid\omega_1,k,\mathbf{C})\prod_{i=2}^{n}P(s_i\mid s_{i-1},\omega_{i-1},\omega_i,k,\mathbf{C})$$

Character Image Generator は、**S** によって与えられた領域に実際の文字パターン $\mathbf{F}=\{f_1,f_2,\ldots,f_n\}$を書き込む。このとき、i 番目の文字パターン f_iは、カテゴリー ω_i とスタイル k だけと相関を持つと仮定すると、**F** の出現確率は以下のように書ける。

$$P(\mathbf{F}\mid\Omega,k,\mathbf{S},\mathbf{C})=\prod_{i=1}^{n}P(f_i\mid\omega_i,k)$$

これらの結果から、単語画像Ψが生成される確率は以下のようになる。

$$\begin{aligned}P(\Psi\mid\mathbf{C})&=P(\Omega,k,\mathbf{S},\mathbf{F}\mid\mathbf{C})\\&=P(\Omega)P(k)P(s_1\mid\omega_1,k,\mathbf{C})P(f_1\mid\omega_1,k)\\&\prod_{i=2}^{n}P(f_i\mid\omega_i,k)P(s_i\mid s_{i-1},\omega_{i-1},\omega_i,k,\mathbf{C})\end{aligned}$$

図２に、**C**、Ω、**S** と **F** の例を示す。

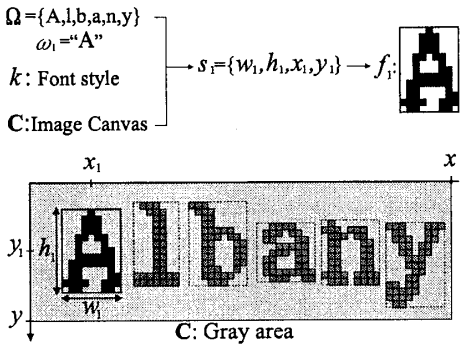

図２． パラメータの例

3．単語認識アルゴリズム

提案する単語認識アルゴリズムはトップダウンに単語画像Ψを生成し、入力画像Γと比較を行う。単語画像生成には、予め辞書に登録されている単語の綴りΩと、モデルが記憶しているフォントパターン k を用いる。

さて、入力画像Γを認識するときには、考えられる全てのパラメータによって単語画像Ψを生成しΓと比較することも考えられるが、計算コストが非常に大きくなってしまう。そこで、単語画像の生成に先立ち、入力画像からパラメータ（各個別文字の位置）を推定し、この値に基づいて単語画像を生成することにする。

パラメータの推定は三つのステップからなる。ステップ 1 では入力画像の黒画素の左端と上端から一文字目の位置を粗く推定する。ステップ２では個別文字のテンプレートを用いて一文字目の位置を詳しく推定する。ステップ３では、次の文字の位置を粗く推定した後に、個別文字のテンプレートを用いて詳しい位置を推定し、これを最後の文字まで繰り返す処理を行う。

† NEC マルチメディア研究所
‡ University of Essex

上記の処理の後に、推定された各個別文字の位置とテンプレートを用いて実際に単語画像Ψを生成し、入力画像Γとの市街区距離を求める。本論文の認識対象画像はタイプライタによる文書ゆえ、セグメントのサイズは一定とした。

3.1 ステップ1

ステップ 1 では、初めに入力画像Γの黒画素の左端（x_{min}）を求め、次に約一文字分の幅（w_s）の範囲で黒画素の上端（y_{min}）を求める（図３）。この処理により、一文字目の位置 s_1 を粗く求めることができる。

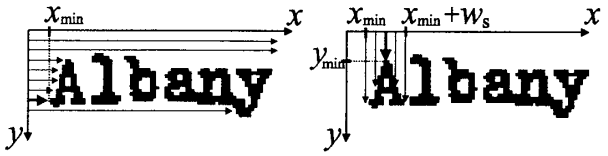

図３．　一文字目の位置の粗いサーチ

3.2 ステップ2

ステップ 1 で粗く推定されたセグメント中心の±7 画素分の領域（図 4 の Fixed area）で、個別文字のテンプレートと部分画像との市街区距離を求め、最小距離値の得られる座標を一文字目の位置 s_1 とする。これは、$P(s_1|\omega_1,k,\mathbf{C})$の値が Fixed area 内で一定、その外でゼロと近似することに対応する。このとき、セグメントのサイズは固定とした。

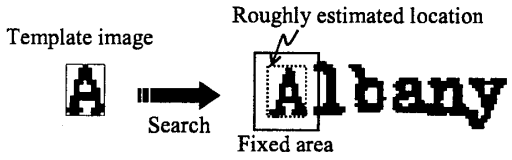

図４．一文字目の位置の詳しいサーチ

3.3 ステップ3

二文字目以降の位置 s_i は、まず初めに直前の文字の位置 s_{i-1} とカテゴリー ω_i と ω_{i-1} を用いて粗く求める。文字カテゴリーを、アセンダを持つクラス(AC)と持たないクラスの(NAC)2 クラスに分けると、s_i の中心座標(x_i,y_i)は以下の式で求めることができる。

$$x'_i = x'_{i-1} + \Delta x$$

$$y'_i = \begin{cases} y'_{i-1} + \Delta y & \text{If } \omega_{i-1} \text{ is AC and } \omega_i \text{ NAC} \\ y'_{i-1} - \Delta y & \text{If } \omega_{i-1} \text{ is NAC and } \omega_i \text{ AC} \\ y'_{i-1} & \text{else} \end{cases}$$

これは $P(s_i|s_{i-1},\ \omega_i,\omega_{i-1},k,\mathbf{C})$の最大値を与える s_i を粗い近似で求めることに対応する。ここでΔx=12, Δy=5 とした。次に、ステップ 2 と同様にテンプレートを用いて最良のセグメント中心を求める（図５）。

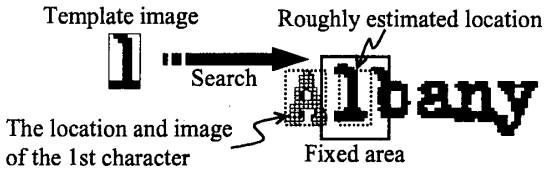

図５．　二文字目以降の位置の詳しいサーチ

3.4 単語画像生成

ステップ 1-3 で推定した位置にテンプレート画像を貼り付けて単語画像Ψを生成する。$P(f_i|\omega_i,k)$の値は一定とした。

単語辞書に登録されている単語数を N_Ω、モデルが記憶しているスタイルの数を N_k とすると、$N_\Omega \times N_k$だけ単語画像を生成する。入力画像Γとの市街区距離が最小の単語画像Ψを求めることで認識を行う。

4. シミュレーション

提案法を 4468 枚のアーカイブ・カードの画像を用いて評価した。単語辞書には約 2 万 8 千単語を登録した。フォントの種類は 6 種類を記憶した。提案法による認識率と従来法[2]による認識率を表 1 に示す。

表１．　認識率

	従来法	提案法
1st	89.4%	99.8%
10-best	94.6%	99.9%

提案方式による認識結果の例として、図 6 に、入力画像(Original Image)と生成された単語画像、正しい綴り(Transcription)と一位認識結果(Recognition Result)を示す。

Original Image	Generated Word Image (Recognition Result)	Transcription	Recognition Result
cypholoma	cypholoma	cypholoma	cypholoma
angulata	angulata	angulata	angulata
DOHERTYA	DOHERTYA	DOHERTYA	DOHERTYA
quinquelineata	quinquelineata	quinquelineata	quinquelineata
SICULODES	SICULODES	SICULODES	SICULODES
CHEVALIERELLA	CHEVALIERELLA	CHEVALIERELLA	CHEVALIERELLA
approximata	approximata	approximata	approximata
nummulalis	nummulalis	nummulalis	nummulalis
ACHROIA	SUFETULA	ACHROIA	SUFETULA
unicolor	tricolor	unicolor	tricolor

図６．　認識結果

図 6 から、提案方式はかすれやつぶれ、文字同士の接触がある場合にも単語を正しく認識することが可能である。しかし、"ACHROIA"の画像ではつぶれにより文字の線幅が大きく変わったために、"unicolor"の例では"n"の文字線幅と文字間隔が同時に変動したため誤認識となった。

5. おわりに

単語画像生成モデルを用いてトップダウンに単語画像を生成し、入力画像と単語単位で比較する認識方式を提案し、低品質印刷英単語の認識シミュレーションにおいて 99.8%の認識率を得た。処理速度の向上が今後の課題である。

謝辞

アーカイブ・カードのデータベースをご提供下さった、ロンドン自然史博物館の関係者に感謝します。

参考文献

[1] A. C. Downton et al.: "Constructing Web-Based Legacy Card Archives – Architectual Design Issues and Initial Data Acquisition", ICDAR'01, 2001.

[2] S. M. Lucas et al.: "Robust Word Recognition for Museum Archive Card Indexing", ICDAR'01, 2001.