

筆記の様子を記録した連続静止画像からのストローク情報の生成

浜口 拓輝^{*1} 加藤 直樹^{*1} 杉原 敏昭^{*2}

A development of the method for generating strokes data from continual images recorded a series of strokes

Hiroki Hamaguchi^{*1}, Naoki Kato^{*1} and Toshiaki Sugihara^{*2}

Abstract – We propose the method for generating strokes data like tablet PC from continual images recorded a series of strokes. It is how to sample the characteristic points from the final image from among a series of continual images and then the characteristic points compare continual images to sort. And we also calculated the character recognition rate of the Hiragana and Kanji thorough experiments which are taking picture rate are 15 and 30 frames per second. As a result, the agreement rate of the character of the Hiragana became 90% or more.

Keywords: Tabletop computer, Strokes data

1. はじめに

知識の共有や創造を目的として、少人数で対面して行われる会議はオフィスにおいてよく見られる光景である。近年、このようなリアルタイムかつ複数人で行われる会議を、コンピュータを使って支援し、円滑化・効率化・活性化するシステムの開発が行われている。

その一つとして、テーブル面にコンピュータ画面を表示する Tabletop Computer 環境の開発が進んでいる。図1はその一例で、画面をビデオカメラで記録し、書き込んだマジックペンの筆跡を静止画像として保存できるシステム Interactive Station である [1]。たとえば、Interactive Station では、Microsoft Word で文章を表示した状態で、画面上にマジックペンで書き込みを行なうと、筆跡をオブジェクトとして Word に取り込むことができる。また、単にテーブル面にマジックペンで書き込みを行うため、複数人で同時に筆記できるという、従来のペン入力機器の課題であった同時複数入力を可能にしている。

このシステムでは、マジックペンの筆跡を静止画像として保存することはできるが、筆跡をペン入力タブレット等からの入力と同じように、筆点列（ストローク）データとして取得することはできない。しかし、筆跡をストロークデータとして取得できれば、筆記再生や、オンライン方式の文字認識、ジェスチャ認識を行なうことができるようになる。

本稿では、上記の利点を、画面をビデオカメラで記録する方式のハードウェアでも受けられるようにすることを目標とする基礎研究として、書き込みの様子を連続的



図1 Interactive Station の外観
Fig.1 Looks of Interactive Station.

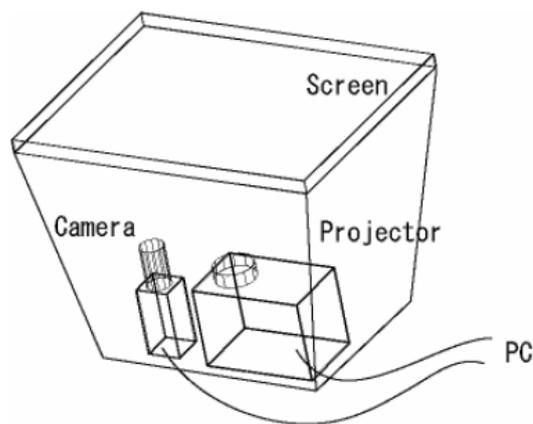


図2 Interactive Station のシステム構成
Fig.2 System composition of Interactive Station.

*1: 東京学芸大学
*1: Tokyo Gakugei University
*2: 株式会社リコー
*2: Ricoh CO., Ltd.

に記録した画像から、ストロークデータを生成する手法を提案し、本手法の有効性について検討と考察を行なう。

2. ストロークデータ生成手法

2.1 前提条件

本手法では、図2のようなマジックペンによる書き込み過程を記録した連続的な画像群を入力対象とする。十分に短い間隔で書き込みの様子を記録できる場合、画像の差分を取ることで、ストロークデータを生成することができるが、本研究では、現実的な撮影レートである30fps (frames per second) および15fpsで記録した静止画像群からストロークデータを生成することを目指す。

2.2 提案手法

連続静止画像(図3)のうち、最終画像から特徴点を抽出し、抽出した特徴点を最初の連続静止画像から順番に比較することにより、抽出した特徴点に対応する図形画素が出現する順に番号付けを行なう。この処理によって、すべての特徴点を書いた順に並び替えることができ、ストロークデータが生成できる(図4)。

2.3 前処理

前処理として膨張処理と細線化処理を最終画像に施す。注目画素 P_0 の8近傍画素の集合を(式1)のように定義する。

$$N_8 = \{P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8\} \quad \dots (式1)$$

また、画素 P_k ($k \in N_8$) の諧調値を $B(P_k)$ ($k \in N_8$) と表記し、背景画素(マジックペンで書かれていない部分)の諧調値を0、図形画素(マジックペンで書かれている部分)の諧調値を1とする。

(1) 膨張処理

マジックペンでの書き込みを行なうと、マジックペンによるかすれや、筆記の様子をビデオカメラで記録するとき生じるかすれがあることが予想される。かすれがあると適切な細線化処理を行うことができないため、かすれを取り除くために、最終画像に膨張処理を行なう(図5中)(式2)の2つの条件がすべて満たされたとき $B(P_0) = 1$ と置き換える処理を画像中のすべての画素において独立に行う。(式2)は、背景画素に注目したとき、その画素を中心とする8近傍中に1つでも図形画素があれば、その背景画素を図形画素に置き換える処理を示す。

$$B(P_0) = 0 \quad \dots (式2)$$

$$B(P_k) = 1 \quad (k \in N_8)$$

(2) 細線化処理

膨張処理を行なった最終画像に対し、細線化処理を行う(図5右)。細線化処理にはHilditchによる細線化[2]

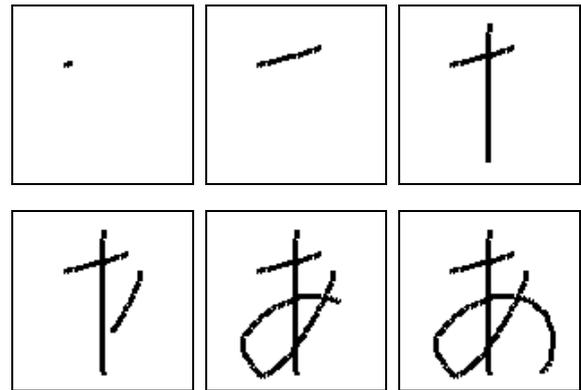


図3 連続静止画像の抜粋例

Fig.3 An excerpting example of continual images.

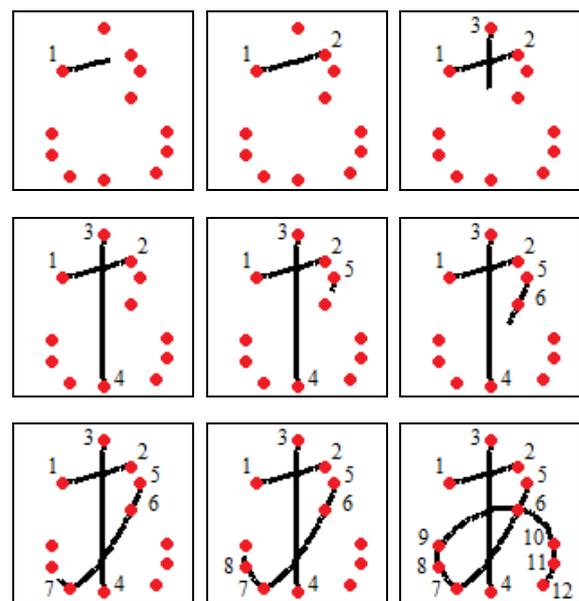


図4 ストロークデータ生成手法の概要

Fig.4 Outline of the method of generating strokes data.



図5 最終画像(左)と処理画像

Fig.5 The final image (left) and processing images.

を行ってから、特定のパターンと一致したとき、残存している不要な図形画素を削るマスクパターン法の二段階で行なう。

Hilditchによる細線化を行うと、図形画素は、図6のM1~M4の左側パターンのように、上下左右いずれかのみ4連結で接続される。しかし、4連結では、まだ削ることができる不要な図形画素が残っているため、図6のM1~M4の左側パターンと一致した部分を、右側のパターン

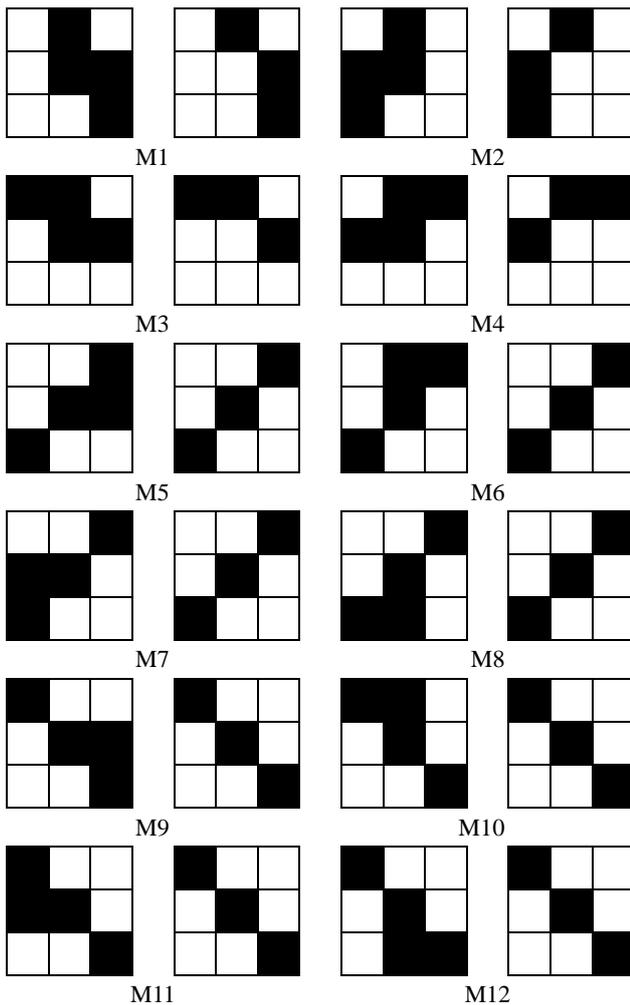


図6 マスクパターン

Fig.6 Mask pattern.

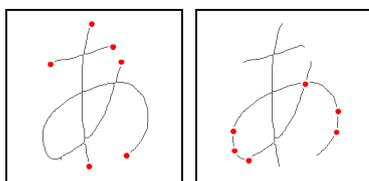


図7 特徴点の抽出

Fig.7 Sampling of characteristic points.

に置き換えるマスクパターン処理を行なう。その後、M5～M12パターンにおいても同様の処理を行う。

2.4 特徴点の抽出

特徴点の抽出は、端点の抽出とそれ以外の特徴点の抽出の二段階で行なう。

(1) 端点の抽出

まず、注目画素 P_0 に対して (式3) を満たすとき、注目画素 P_0 を端点である特徴点として抽出する (図7左)。(式3)は、注目している画素が図形画素であり、その8近傍画素の中に図形画素がただ1つ存在するとき、注目

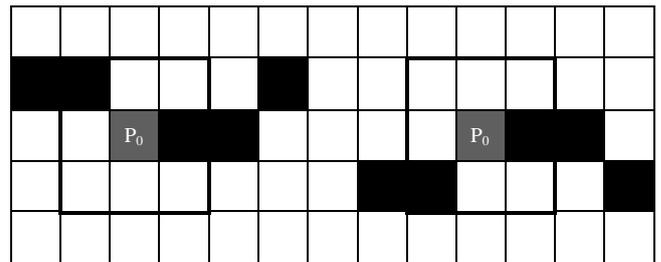


図8 特徴点抽出パターン

Fig.8 Sampling pattern of characteristic points.

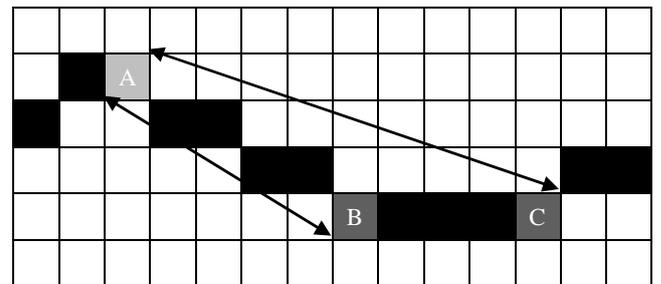


図9 1枚の静止画像に複数ある特徴点の並び替え

Fig.9 Sort of characteristic points that exist in an image.

している画素が端点であることを示す。

$$B(P_0) = 1 \quad \dots (式3)$$

$$B(P_k) = 1 \quad (k \neq N_8)$$

(2) 端点以外の特徴点の抽出

筆跡の中で、曲率が大きく変化している画素を特徴点として抽出する。さまざまな書き込みを想定し、実験を繰り返したところ、経験的に図8のような2つの特徴点抽出パターンが得られた。それぞれのパターンにおいて、 90° 、 180° 、 270° 回転させたときも含め、計8パターンとの比較を行う。ただし、図8において左右に3連結している画素については、一例であり、この部分は1つ以上連結していれば、いくつ連結していてもよいとする。そして、8パタンのいずれかと一致したとき、注目画素 P_0 を端点以外の特徴点として抽出する (図7右)。

2.5 特徴点の順序付け

抽出した特徴点と筆記の様子を記録した連続静止画像を順に比較し、抽出している特徴点のはじめて図形画素になる順を筆記順序とする (図4)。1枚の静止画像に、はじめて図形画素になる特徴点が複数ある場合については、直前に順序付けされた特徴点からの距離を求め、その距離が小さい順を筆記順序とする。たとえば、図9において、直前に順序付けされた特徴点を A、同時に図形画素に含まれた特徴点を B、C とすると、AB間とAC間の距離を比較し、AB間の距離の方が小さいため、A、B、Cと順序付けする。

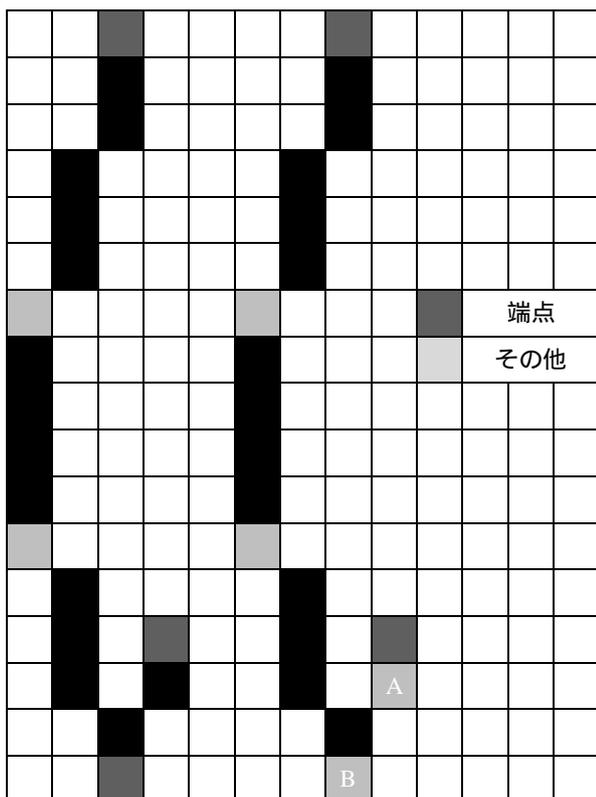


図 10 問題例と改善例

Fig.10 Examples of problem and improvement.

2.6 特徴点の補正

特徴点を抽出するとき、端点である特徴点か、端点以外の特徴点か判断をしている。ここで、ストロークデータにおける端点は、筆記する線分の始点または終点を指す。しかし、前述の特徴点抽出処理では、たとえば、“は”や“ほ”のハネの部分のように、曲率が大きく変化している部分が、端点である特徴点として抽出されてしまうことがある。

そこで、特徴点の順序付けを行ってから、端点 (x_1, y_1) と端点 (x_2, y_2) の間に特徴点が 1 つも存在しないとき、(式 4) と図形画素との交点を新たに特徴点として抽出する。(式 4) は、端点 (x_1, y_1) と端点 (x_2, y_2) を斜辺とする直角三角形を作り、斜辺以外の二辺のうち、長辺の垂直二等分線と交わる図形画素 (x, y) を特徴点とする処理を示す。

X 座標の差 > Y 座標の差のとき ... (式 4)

$$x = (x_1 + x_2) / 2 \quad (\min(y_1, y_2) \leq y \leq \max(y_1, y_2))$$

X 座標の差 < Y 座標の差のとき

$$y = (y_1 + y_2) / 2 \quad (\min(x_1, x_2) \leq x \leq \max(x_1, x_2))$$

すると、端点と端点の間に少なくとも 1 つは端点以外の特徴点が存在することになる。そこで、再度、特徴点の順序付けを行ない、端点が端点以外の特徴点に挟まれ

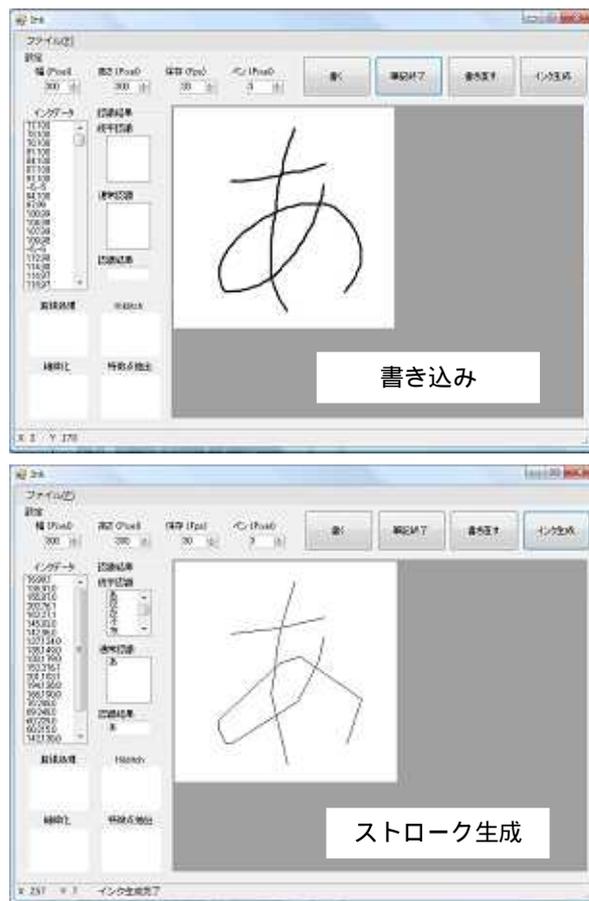


図 11 ストロークデータ生成ツール

Fig.11 Tool of generating strokes data.

たとき、端点以外の特徴点と判断のし直しを行なう。たとえば、図 10 では(式 4)によって、新しい特徴点 A を追加し、その後、特徴点 B を端点以外の特徴点と判断し直している。

3. 予備評価

3.1 目的

提案したストローク生成手法の有効性について検証するため、ひらがなと漢字について、入力した文字と文字認識結果の一致率を調べ、本手法の有効性の検証を行なう。

3.2 実装

提案したストローク生成手法の有効性を確かめるために、一文字の筆記を行なうと、指定の撮影レートで連続静止画像を作成し、その静止画像群を元にストロークを生成し、単文字認識にかけるツールを試作した(図 11)。なお、単文字認識エンジンには Microsoft Windows XP Tablet PC Edition 2005 Recognizer を用いた。

3.3 実験環境

いずれの実験においても、ワコム製 17 インチの液晶ペンタブレット DTU-710 を用いる。また、一文字の筆記を

表1 漢字リスト

Table1 Kanji list.

学年	実験に用いた漢字				
1年生	一	五	人	町	右
	口	水	天	雨	校
2年生	引	丸	古	作	場
	太	東	歩	羽	岩
3年生	悪	開	銀	仕	拾
	真	炭	湯	氷	由
4年生	愛	街	給	固	士
	照	束	徒	夫	約
5年生	異	揮	己	至	純
	泉	値	扨	忘	論

表2 文字認識率

Table2 Character recognition rate.

	第一位候補	第三位候補以内
実験 1A	0.909	0.926
実験 1B	0.909	0.930
実験 2	0.284	0.439

行うキャンパスのサイズは、縦横 300 ピクセル(約 7.14cm)とし、通常の入力と同じように筆記を行なった。

3.4 被験者

被験者は、筆者 1 名であり、普段からペン入力タブレットを使用している。

3.5 タスク内容

筆者が筆記したひらがな 46 字種各 5 回に対する認識実験を 30fps (実験 1A) および 15fps (実験 1B) の撮影レートで行う。また、ランダムに選んだ、小学校 1 年生から 5 年生で習う漢字、各学年 10 字種各 5 回に対する認識実験を 30fps (実験 2) の撮影レートで行う。なお、実験 2 の結果より、撮影レートを 15fps に設定した実験を行なう必要がないと判断したため、行わなかった。実験に用いた漢字のリストを表 1 に示す。

3.6 実験結果

実験 1A, 1B, および実験 2 において、入力した文字と文字認識結果の一致率を表 2 に示す。

実験 1A, 1B, および実験 2 において、入力した文字に対し、誤認識率の高い文字と誤字候補を表 3 に示す。

4. 考察

4.1 文字認識率

入力をひらがなとしたとき、撮影レートを 30fps (実験 1A), 15fps (実験 1B) としたいずれの場合でも、第一位候補が入力文字と一致した率が 90% 以上を示した。しか

表3 誤認識率の高い文字

Table3 Character that wrong recognition rate is high.

	入力文字	誤字候補				
実験 1A	け	1		l		
	た	7	十	ナ	f	
	と	て	Z	乙		
	は	1		l	!	
	ほ	1		l	I	
	や	ち	ぺ	パ	人	
	よ	0	S	D		
実験 1B	け	1		l	I	
	た	7	十	ナ	l	
	と	て	Z	乙		
	は	1		l	I	
	ほ	1		l	I	
	や	ち	力	カ	九	
	よ	0	O			
実験 2	右	六	方	介	る	
	口	口	は			
	校	投	松	様	持	
	丸	も	B	の		
	岩	京	交	寒		
	悪	突	窓	恐	空	
	仕	に	は			
	拾	格	路	梓		
	炭	年	業	産	そ	
	湯	濟	場			
	氷	水	れ			
	街	例	卸			
	固	国	回	四	因	
	照	京	き	米	気	
	束	来	東			
	徒	従	値	種	任	
	異	算	業	量		
己	を	て	そ	を		
至	査	立	そ			
純	鉄	関	紙	立		
扨	排	協	州	洋		

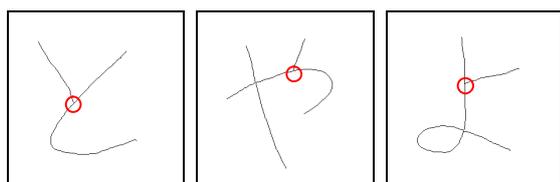


図 12 端点が取得できない例

Fig.12 Example that edge point can't be secured.

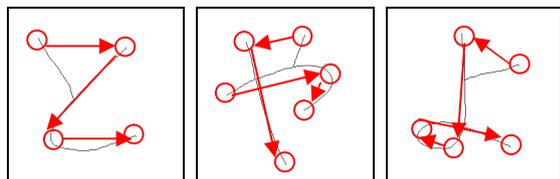


図 13 端点が取得できないストローク例

Fig.13 Example of stroke that edge point can't be secured.

し、入力を漢字（実験 2）にすると、実験 1 に比べ、文字認識率は大きく低下する（表 2）。これより、ひらがなにおいてのみ、本手法の有効性を示すことができたと考えられる。

4.2 撮影レートの与える影響

入力をひらがなとしたとき、実験 1A, 1B において、ほぼ同じ文字認識率を示した（表 2）。また、誤認識率の高い文字が同じ文字になる傾向があり（表 3）、ひらがなを入力文字としたとき、撮影レートの違いが、本手法に与える影響は少ないと考えられる。

4.3 誤認識率の高い文字の特徴

実験 1A, 1B では誤認識率の高い文字が同じ文字になり、誤認識の原因を考察したところ、傾向として二種類に分類することができる。

一つは、「け、た、は、ほ」のように文字全体を認識するのではなく、文字の扁の部分だけを認識しているグループである。これは、文字認識エンジンの問題であり、一文字に対応する認識枠をあらかじめ用意することにより、認識精度を高められそうである。

もう一つは、「と、や、よ」のように、ストロークの始点または終点として取得したい特徴点が、筆記中の重なりによって抽出できないというグループである（図 12）。適切に生成された時のストロークデータと比べると、筆記中の重なりによって抽出できない特徴点を除く特徴点に順序付けが施されたストロークデータが生成されており、これらが認識率を低下させた原因であると考えられる（図 13）。

実験 1A, 1B のように、ひらがなを入力文字としたとき、ストロークの始点または終点として取得したい特徴点が、筆記中の重なりによって抽出できなくなる文字は限られている。しかし、漢字では、多くの文字に、ストロークの始点または終点として取得したい特徴点が、筆記中の重なりによって抽出できないという問題が生じ、認識率を大きく低下させた原因であると考えられる。

また、漢字はひらがなより文字が複雑であり、似たような候補が多いことが認識率を低下させた原因であると考えられる。

5. おわりに

本稿では、書き込みの様子を連続的に記録した静止画像から、ストロークデータを生成する手法を提案し、その有効性を調べるために行なった実験と、その結果に対する考察を述べた。

実験から、ひらがなの入力に対して、30~15fps という限定的な撮影レートにおいて、文字認識結果の一致率が 90% 以上であり、本手法が有効である可能性を示すことができた。

また、誤認識の原因として、ストロークの始点または終点として取得したい特徴点が、筆記中の重なりによって抽出できなくなっていることがわかった。

今後の課題として、今回、認識率を低下させた原因を解消するためにアルゴリズムを改良することが挙げられる。特に、1 枚の静止画像に複数ある特徴点を並び替えるときに、ユークリッド距離による並び替え方法を採用しているが、今後、より精度よく並び替えられる方法を追求したい。また、今回、抽出した特徴点だけでは間引きが大きすぎ、筆記再生にたえられるストロークデータとはなっていない。今後、特徴点どうしの中の点をストロークデータに追加することも試みたい。

謝辞

本研究の一部は、科学研究費補助金・若手研究（B）19700112 の補助による。

参考文献

- [1] 新西, 伊賀, 桜井: Interactive Station: デジタル情報に手書きできるテーブルトップコンピュータ, ヒューマンインタフェースシンポジウム 2007 論文集, pp.245-248 (2007).
- [2] 安居院, 長尾: 画像の処理と認識, pp.68-72, 昭晃堂 (1992).