

博士論文

日本語筆跡に現れる個人性の抽出と
オンライン筆者照合に関する研究

中村 善一

2008年12月22日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学)授与の要件として提出した博士論文である。

中村 善一

審査委員：

木戸出 正繼 教授	(主指導教員)
千原 國宏 教授	(副指導教員)
波部 斉 助教	(副指導教員)
木村 文隆 教授	(三重大学)

日本語筆跡に現れる個人性の抽出と オンライン筆者照合に関する研究*

中村 善一

内容梗概

本研究の目的は、オンライン筆者照合に用いる筆跡を署名に限定するのではなく、より一般的な漢字文字列を対象としてどのような特性値に個人性が現れるかをまず明らかにし、その結果に基づいたオンライン筆者照合方法を提案し実験によりその有効性を評価することである。本論文は以下の6章により構成される。

第1章では、オンライン筆者照合の現状と問題点を示し、本研究の目的を明らかにする。

第2章では、漢字に近いストローク構造を持つカタカナ文字列を対象にして予備実験を行う。筆跡に個人性が現れるのは各個人が習得している書写技能に個人差があるためであると考え、書写技能に基づく特性値を定義し、それら特性値に個人性が現れることを示す。さらに、筆者照合および筆者識別実験を行い、抽出した特性値を用いて筆者識別・照合が可能であることを示す。

第3章では、漢字文字列を対象に個人性を表す特性値について検討する。カタカナ文字列で用いた書写技能に基づく特性値は筆跡鑑定の検査項目と類似しているため、特性値を筆跡鑑定の知見に基づいて整理した。これら特性値を抽出し分析することで、個人性が現れやすい特性値を明らかにし、さらに、筆者照合および筆者識別実験により筆者認識の可能性を示す。

第4章では、抽出した特性値を用いた筆者照合方法を提案する。特性値の種類ごとに本人間と他人間の距離分布を求め、それに基づいて特性値の選択を行う。つぎに、真正筆跡の各特性値は参照筆跡の平均値に近いものが多く、遠いものは

* 奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文, NAIST-IS-DD0361027, 2008年12月22日.

少ないという考えに基づき，各特性値の参照筆跡の平均値からの偏差の度数分布を基にした識別器を提案し，その有効性を筆者照合実験を行うことにより明らかにする．

第5章では，筆者照合における重要な課題である偽造筆跡の排除に対する提案手法の有効性を検討する．本物筆跡のみを用いて設計した識別器に，十分に訓練された偽造筆跡を入力して実験を行い，提案手法が偽造筆跡に対しても有効であることを示す．さらに，第3・4章で用いた漢字文字列以外のデータを収集し，その文字列データに対して同様の筆者照合実験を行い，提案手法の有効性を示す．

第6章で，本研究で得られた成果をまとめるとともに，今後の課題と方向性を示す．

キーワード

オンライン筆者照合，特性値，筆跡鑑定，特性値選択，識別器，バイオメトリクス

Studies on Extraction of Individuality Revealed in Japanese Handwriting and Online Writer Verification*

Yoshikazu Nakamura

Abstract

The purpose of this research is to investigate what kind of feature parameters determine individuality of a string of kanji characters, and to propose an online writer verification method using these feature parameters. Furthermore, the verification method is evaluated by verification experiments. The dissertation consists of the following six chapters.

In Chapter 1, the state of the art of online writer verification is briefly mentioned and the problems are elucidated, which are the target issues in the research.

In Chapter 2, preliminary experiments using katakana characters similar in stroke shape to kanji characters are conducted. Since I consider that the difference in individual calligraphic skill causes handwriting individuality, I extract the feature parameters based on calligraphic skill and then show that these feature parameters express handwriting individuality. Furthermore, writer verification and identification experiments are performed and the experimental results show that these parameters are significant enough to identify and verify individual writers.

In Chapter 3, the feature parameters that express individuality of a string

* Doctoral Dissertation, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0361027, December 22, 2008.

of kanji characters are investigated. The feature parameters are redefined based on the knowledge of document examiners because the feature parameters extracted from the knowledge of calligraphic skill are similar to the inspected characteristics used by document examiners. The feature parameters are extracted and analyzed. Consequently, the feature parameters that express the individuality are revealed. Furthermore, the potential for writer verification is indicated by writer verification and identification experiments.

In Chapter 4, the method of writer verification using the feature parameters is proposed. First, the feature parameters effective in writer verification are selected based on the distribution of within-writer distances and between-writer distances in the feature space. Next, two classifiers based on frequency distribution of each feature parameters' deviation from the average of genuine handwriting samples are designed. The idea for two classifiers is that if the handwriting sample is genuine, it has a large number of features near to the average of reference handwriting samples and a small number of features far from the average. The proposed method is evaluated by verification experiments. The experimental results show that the proposed method is effective in writer verification.

In Chapter 5, the effectiveness of the proposed method to determine skilled forgeries is examined. The writer verification experiments using well-practiced forgeries are performed. The experimental results show that the proposed classifiers are effective against skilled forgeries. Furthermore, the experiments using different kanji characters is carried out and the experimental results show that the proposed method is effective in writer verification.

Chapter 6 briefly concludes the achievements in this research and mentions future avenues of exploration and elaboration.

Keywords:

Online writer verification, Feature parameters , Knowledge of document examiners , Feature selection , Classifier , Biometrics

目次

第 1 章	序論	1
1.1	研究の背景	1
1.2	筆者認識技術の現状と課題	3
1.3	研究の目的	8
1.4	本論文の構成	10
第 2 章	カタカナ文字列に現れる個人性の抽出	13
2.1	緒言	13
2.2	特性値の定義	14
2.2.1	書写技能	14
2.2.2	特性値の定義	16
2.3	特性値の評価	21
2.3.1	評価に用いるデータ	21
2.3.2	評価方法	22
2.3.3	評価結果	23
2.4	特性値の筆者識別力の検討	28
2.4.1	実験方法	28
2.4.2	実験結果	30
2.5	結言	32
第 3 章	漢字文字列に現れる個人性の抽出	33
3.1	緒言	33

3.2	特性値の定義	34
3.2.1	筆跡鑑定の知見	35
3.2.2	特性値の定義	36
3.3	特性値の評価	41
3.3.1	漢字文字列の選択基準	41
3.3.2	評価に用いるデータ	41
3.3.3	評価方法	43
3.3.4	評価結果	45
3.4	特性値の筆者識別力の評価	53
3.4.1	実験方法	53
3.4.2	実験結果	54
3.5	結言	59
第4章	オンライン筆者照合手法の提案	61
4.1	緒言	61
4.2	想定する筆者照合システム	62
4.3	特性値の選択方法	63
4.3.1	筆者内距離と筆者間距離の分布に基づく特性値選択法	63
4.3.2	選択結果	65
4.4	筆者照合手法	67
4.4.1	特性値の偏差の度数分布に基づく筆者照合手法	67
4.4.2	比較のために用いる筆者照合手法	70
4.5	筆者照合実験	70
4.5.1	実験方法	70
4.5.2	実験結果と検討	71
4.5.3	任意定数設定方法の妥当性の検討	72
4.5.4	ユークリッド距離法, 多数決法との比較	74
4.5.5	特性値の種類が照合結果に及ぼす影響の検討	75
4.6	結言	76

第 5 章	偽造筆跡と他文字列に対する提案手法の性能評価	79
5.1	緒言	79
5.2	偽造筆跡に対する照合性能評価	80
5.2.1	偽造筆跡データの収集と分析	80
5.2.2	筆者照合実験による照合性能評価	84
5.3	他文字列に対する照合性能評価	86
5.3.1	データ収集	86
5.3.2	特性値の選択	86
5.3.3	照合実験	89
5.4	結言	91
第 6 章	結論	93
	謝辞	97
	参考文献	99
	研究業績	105

目次

1.1	想定する筆者照合システムの処理の流れ	11
2.1	点画の分解	17
2.2	天地左右のあきと字間の説明	18
2.3	送筆部曲直の説明	20
2.4	評価に用いた筆跡の例	22
3.1	ペン高度 al と方位 az	34
3.2	文字切り出し	38
3.3	始筆部, 送筆部, 終筆部への分割	38
3.4	筆記タイミング	39
3.5	方向と曲直 Φ	40
3.6	収集したデータの例	42
3.7	筆者内距離と筆者間距離の分布の概念図	44
3.8	文字列の平均ペン方位の分布 (箱ひげ図)	47
3.9	「永」4画目終筆部の平均筆記速度 (箱ひげ図)	50
3.10	「奈」1画目送筆部の方向 (箱ひげ図)	51
3.11	全特性値と選択した特性値を用いた場合の ROC	55
3.12	選択した動的特性値と静的特性値を用いた場合の ROC	57
3.13	選択した字画の特性値と全特性値から選択した特性値を用いた 場合の ROC	58
3.14	各部の特性値を組み合わせていった場合の ROC	59

4.1	筆者内距離と筆者間距離の分布の概念図	64
4.2	特性値の偏差の度数分布 (全サンプル, 選択した特性値 270 個)	67
4.3	偏差度数比識別器の ROC 曲線	73
4.4	偏差度数識別器の ROC 曲線	73
4.5	ROC 曲線の比較	74
5.1	収集したデータの例「奈良永遠」の本物と偽造筆跡	80
5.2	収集したデータの例「大和郡山」	87

表目次

1.1	バイオメトリクスの比較	2
1.2	海外でのオンライン筆者照合	5
1.3	国内でのオンライン筆者照合	6
2.1	書写技能の分類	15
2.2	各特性値の F 値の平均, 最大, 最小	24
2.3	F 値が 200 以上の特性値	25
2.4	各特性値の筆者ごとの変異係数 (標準偏差 / 平均値) の平均	26
2.5	全筆者で相関係数 r が $ r \geq 0.671$ (母相関係数 $\rho = 0.8$ の信頼率 95% の信頼限界) の特性値の組み合わせ	28
2.6	筆者単独で相関係数 r が $ r \geq 0.829$ ($\rho = 0.9$ の信頼率 95% の信頼限界) の特性値の組み合わせ (他の筆者は $ r < 0.671$)	28
2.7	識別結果	30
2.8	文字ごとの識別率	31
2.9	照合結果	31
3.1	定義した特性値一覧	37
3.2	分散分析とクラスカル・ウォリス検定の結果	46
3.3	H 値上位 200 個に含まれる特性値の数	48
3.4	評価値 S	52
3.5	照合実験と識別実験の結果	54
3.6	部分ごとの照合実験結果	56

4.1	評価値 S	65
4.2	と の設定結果	69
4.3	照合実験結果	71
4.4	全特性値を用いた照合実験結果	72
4.5	特定の特性値の種類を除外したときの照合結果	75
5.1	偽造筆跡データ 566 個の内訳	81
5.2	偽造筆跡の評価値 S	83
5.3	照合実験結果	85
5.4	評価値 S	88
5.5	照合実験結果	89
5.6	「奈良永遠」と「大和郡山」の照合実験結果の比較 (選択した特性値を使用した場合)	90

第 1 章

序論

1.1 研究の背景

本人を確認することは，銀行口座の開設，クレジットカードの利用，特定場所への入退室など社会生活のいたるところで必要であり，重要な問題である．本人を確認する方法としては以下のものがある．

- 本人の所有による確認： ID カード，印鑑，鍵など．
- 本人の記憶による確認： 暗証番号，パスワードなど．
- 本人の身体的特徴や行動的特徴による確認： 顔，指紋，虹彩，音声，筆跡（署名）など．

本人の所有による確認は，盗難や偽造の危険性があり，本人の記憶による確認は，本人が忘れるあるいはパスワードや暗証番号が盗まれるという危険性がある．それらに比べ，本人の身体的特徴や行動的特徴による本人確認は，盗難や忘れるなどの心配がなく，より安全な本人確認方法であるといえる．この本人の身体的特徴や行動的特徴による個人認証はバイオメトリクス (Biometrics) と呼ばれ，近年特に注目されている．身近な例では，ATM(Automated Teller Machine) 利用時に，掌あるいは指の静脈パターンにより本人確認を行うなどがある．バイオメトリクスが注目される背景には，近年の情報化社会の進展に伴い電子商取引などネットワーク上でのサービスの需要が増大する中，ネットワーク

における本人確認が重要な課題となっていること，またアメリカで発生した同時多発テロ以降，安全性を確保するために個人認証技術が重要視されてきているという状況がある．

バイOMETRICS技術は，大きく身体的特徴と行動的特徴の二つに分類できる．前者は，指紋，掌形，顔，虹彩，静脈などであり，後者は，声紋（音声），署名（筆跡）などである．表 1.1 に主なバイOMETRICSの比較を示す [1]．ここで，普遍性とは誰もが持っている特徴であること，唯一性とは本人以外は同じ特徴を持たないこと，永続性とは時間の経過とともに変化しないことである．また，拒否率とは本人を間違えて拒否する誤りであり，受入率とは他人を間違えて受け入れられる誤りである．

表 1.1 バイOMETRICSの比較

情報	普遍性	唯一性	永続性	コスト	拒否率 (%)	受入率 (%)	データ量 (Byte)
指紋	高	高	高	低	1.0	0.01	250
掌形	高	中	中	高	0.1	0.1	10
顔	高	低	低	中	5	5	2,000
虹彩	高	高	高	高	10	10^{-6}	200
静脈	高	中	中	高	1.0	0.01	500
声紋（音声）	高	低	低	低	10	10	1,500
署名（筆跡）	高	低	低	中	5	5	1,000

表を見れば個人認証に用いる生体情報としては指紋が良いように思われるが，指紋は犯罪捜査に用いられていることや，入力装置が接触型であることから他人が接触したセンサに自分の指を置くことへの抵抗を感じるなど利用者の受容性という点で若干の問題がある．このように，それぞれの生体情報には一長一短があり，要求されるセキュリティレベルや利用方法に応じて多様なバイOMETRICSが必要になると考えられる．

本研究で取り上げる筆跡は，他の生体情報に比べて精度の点では劣るものの以下のような特長がある．

- 署名による本人確認に代表されるように，従来より社会的に受け入れられ

ている個人認証方法であり，指紋など他のバイオメトリクスに比べ利用者の受容性が高い．

- 行動特性による個人認証であるため本人の意思確認を伴う．
- 身体的特徴を用いる個人認証では登録データが盗まれればデータの変更は不可能であるが，筆跡による個人認証は登録筆跡を変更することが可能である．
- 電子文書への貼り付けが可能である．

このような特長を持つことから，筆跡による個人認証はバイオメトリクスの中でも重要な技術の一つである．

1.2 筆者認識技術の現状と課題

筆者認識技術は，その入力方法によりオフライン方式とオンライン方式に分類される．オフライン方式とは，既に書かれている筆跡をスキャナなどを用いて光学的に入力する方式であり，オンライン方式とは主にタブレットと電子ペンを用いてペンの座標位置，筆圧やペンの傾きなどの情報を時系列データとして入力する方法である．

両者を比較すると，他人が真似ることが困難であると考えられる筆記速度や筆圧，ペンの傾きなどの動的情報を利用できることから，オンライン方式の方がオフライン方式よりも認証精度が高いといえる．また，オフライン方式はカウンターサインや筆跡鑑定などの従来の業務や用途の自動化での利用が期待できるが，オンライン方式はネットワークでの利用も可能であることからカウンターサイン以外での幅広い場面での個人の自動認証へ応用可能であるという利点を持つ．さらに，最近ではタブレット以外にも Tablet PC，PDA(Personal Digital Assistant)，任天堂 DS のようなゲーム機，また NTT ドコモで携帯を利用した筆跡の送受信サービスが開始される [2] など，ペン入力が様々な場面で利用可能になってきている．

以上のように，得られる情報の多さやネットワークでの利用が可能であること，またオンラインペン入力の利用環境が整いつつある状況から，本研究ではオ

ンライン方式を研究対象とする。

オンライン方式による筆者照合に関しては、主に欧米でオンライン署名照合を中心に研究が行われており、その主要な要素技術は、特徴抽出と照合・判定である。用いられる特徴量は、時系列特徴と特徴パラメータに大別できる。時系列特徴としては、ペンの座標 $(x(t), y(t))$ 、筆圧 $p(t)$ 、ペンの速度 $v(t)$ と速度の xy 成分 $(v_x(t), v_y(t))$ 、ペンの加速度 $a(t)$ 、ペンの傾き（高度と方位） $(al(t), az(t))$ などの時間関数やセグメントごとの曲率や向きなどの局所の特徴の時系列がある。また、特徴パラメータは筆跡全体の筆記時間、ペンアップ回数、大きさ、筆記速度の平均値や最大値などの全体的特徴が主に用いられている。さらに、時間関数を直交変換し、その係数を特徴パラメータとして用いるなどがある。照合・判定については、時系列特徴には DP マッチング (Dynamic Programming matching)、隠れマルコフモデル、相互相関などが、また、特徴パラメータにはユークリッド距離、統計的識別器、ニューラルネットワークなどが用いられる。

1993 年までのオンライン署名照合技術については、Plamondon らにより詳しくサーベイされている [3,4]。1994 年以降に提案されたオンライン署名照合の手法には表 1.2 のようなものがあり、現在では筆跡の時系列特徴に対して DP マッチングや隠れマルコフモデルを用いて照合する手法が主流となっている。

本研究は日本語筆跡を対象にしているため、以下に日本語筆跡を対象にしたオンライン筆者照合の現状を概観し、問題点を挙げる。欧米と同様に、オンライン筆者識別の研究はオンライン署名照合に関するものがほとんどであり、筆跡の動的情報であるペンの座標、速度、筆圧、傾きなどの時系列データに対して DP マッチングを行うなど、筆跡を時間関数と捉えて照合するものが主流である。現在までに国内の研究者により提案されている手法には表 1.3 のようなものがある。なお、表中、署名以外の筆跡を対象としているのは、山崎らと Kameya らのみである。

このように日本語筆跡を対象にしたオンライン筆者照合の研究は署名照合を中心に行われているが、

- (1) 筆跡を署名に限定した研究がほとんどで、一般的な文字列を対象としていない、

表 1.2 海外でのオンライン筆者照合

特徴パラメータ		
研究者	特徴量	照合方法
Nelson, et al. [5] (1994)	グローバル特徴 (25 個)	ユークリッド距離, マハラノビス距離, 線形識別関数
Lee et al. [7] (1996)	特徴パラメータ (42 個, 正規化した 49 個)	多数決法
Wu, et al. [9] (1997)	$x(t), y(t)$ の LPC ケプストラム係数	NN(Neural Network)
Lee, et al. [15] (2004)	セグメント特徴とグローバル特徴	NN
Nanni, et al. [18] (2008)	$x(t), y(t), p(t), al(t), az(t)$ をローパスフィルタ後 DCT	Linear Programming Descriptor Classifier
時系列特徴		
研究者	特徴量	照合方法
Yang, et al. [6] (1995)	筆跡に沿って抽出された正規化された角度の時系列	HMM(Hidden Markov Model)
Nalwa [8] (1997)	ジッタ, 筆跡の長さ方向に沿った物理量 (重心, トルク, 慣性モーメント) の系列	相互相関
Dolfing, et al. [11] (1998)	セグメント分割後, 各セグメントを 32 次元の特徴ベクトル (13 個の静的特徴, 13 個の動的特徴, 6 個の構造特徴) で記述	HMM
Jain, et al. [12] (2002)	筆跡を等距離でリサンプリング. 各サンプル点での静的特徴 (7 個) と速度の系列	String matching(DP マッチング)
Feng, Wah [13] (2003)	$x(t), y(t)$, Nalwa の重心とトルク	EPW(Extreme Points Warping)
Bovino, et al. [14] (2003)	$x(t), y(t), v(t), a(t)$	セグメントに分割後, セグメント単位に DP マッチングにより真偽を判定. その結果を統合して最終判定
Kholmatov, et al. [16] (2005)	$x(t+1) - x(t), y(t+1) - y(t)$, 曲率	DTW により参照署名との最大距離, 最小距離とテンプレートとの距離を求め 3 次元ベクトル化, 主成分分析により 1 次元に圧縮後, 線形識別器により判定.
Fierrez, et al. [17] (2007)	$x(t), y(t), p(t)$, 速度, 角度, 曲率, 加速度の時系列	HMM
特徴パラメータ + 時系列特徴		
研究者	特徴量	照合方法
Kashi, et al. [10] (1997)	グローバル特徴 (23 個) と筆跡の接線の傾きと導関数の系列	ユークリッド距離 (グローバル特徴) と HMM(傾きと導関数) の統合

表 1.3 国内でのオンライン筆者照合

特徴パラメータ		
研究者	特徴量	照合方法
楊他 [25] (1996)	ストローク毎に幾何学的特徴 4 種類, GA により特徴発見 ストローク毎に差分角度を求め P 形フーリエ記述子で表現	照合ネットワーク, メンバシップ関数 ベクトル量子化によるカテゴリ化, LVQ による参照ベクトル学習, 重み付け
山崎, 小松 [26] (1996)		
時系列特徴		
研究者	特徴量	照合方法
Sato, Kogure [19] (1982)	$x(t), y(t), p(t)$, 字形, 運動, 筆圧の時系列	DP マッチング, 3 種類の偽距離, マハラノビス距離
Yoshimura, et al. [21] (1991)	$x(t), y(t), p(t)$, 字形, ペンの移動方向, 筆圧の時系列	DP マッチング
李他 [22] (1992)	加速度の時系列を木構造で表現	木マッチング
金他 [23] (1992)	筆圧, 筆速, 握り圧の時系列	ストローク単位で DP マッチング
Zhao, et al. [24] (1996)	Sato, Kogure を改良, 局所的な安定性に基づく重み付け	DP マッチング, 2 種類の偽距離, マハラノビス距離
山中他 [28] (2000)	ペンの傾き, 座標位置, 筆圧の時系列	DP マッチング, (注) 対象は英字署名
Komiya, et al. [29] (2001)	ペンの移動距離・方向, 筆圧, ペンの傾きの時系列	DP マッチング
Kondo, et al. [30] (2003)	筆跡重心からの距離と角度, 筆圧, ペンの傾きの時系列	DP マッチング後の 3 種類の距離を特徴ベクトルとしパーセプトロンに入力
中西他 [31] (2004)	$x(t), y(t)$ を DWT によりサブバンド分解 (局所差分信号)	適応信号処理
Kameya, et al. [32] (2006)	$x(t), y(t), p(t), al(t), az(t), v(t)$, ペンの移動方向の時系列	CDP(Continuous Dynamic Programming)
特徴パラメータ + 時系列特徴		
研究者	特徴量	照合方法
田口他 [20] (1988)	$x(t), y(t), p(t)$, 大局的なパラメータ (7 個), セグメント毎の筆力とペン傾度の波形 (時間波形と微分波形)	すべての特徴量が許容範囲内であれば本物と判定
菊池, 赤松 [27] (2000)	高感度筆圧ペンを試作, 特徴パラメータ (3 個), ストローク毎の筆圧波形	パラメータと波形間距離による 2 段階照合

- (2) 筆跡のどのような特性に個人性が現れるのかという検討が十分なされていない，
- (3) 動的特性が重要視され，動的特性と静的特性を総合的に用いて真偽を判定するという観点に乏しい，
- (4) 日本語筆跡が文字単位，ストローク単位に書かれることを意識して特徴量を抽出するという観点に乏しい，
- (5) 特徴量として筆跡の時間関数を用いる方法は，時間波形を登録するため盗まれれば筆跡の再現が可能となる，

という問題がある．

(1)，(2)に関連した一般的な日本語文字を対象に筆跡のどのような特性に個人性が現れるかを検討した研究として，吉村ら [33] は枠内自由手書きカタカナに対して，モーメントに関する量，縦横に関する量，ストロークの湾曲と長さに関する量，特徴点間の距離と傾きに関する量を特性値として計測し，実験的にそれら特性に個人性が現れることを報告している．また，尺長ら [34] は手書き文字の個人性情報抽出をテクスチャ解析の立場から行っている．また，英字アルファベット文字に関しては Srihari ら [35] が大規模なデータベースを用いて個人性の研究を行っている．しかし，これらの研究はオフライン方式によるものである．他にオフライン方式による筆者識別の研究として，文字の直線性と曲線性に注目したもの [36] や，加重方向指数ヒストグラム特徴を自己想起型ニューラルネットワークの入力とするもの [37] などがある．

一般的な漢字文字列を対象としたオンライン方式の筆者照合に関しては，次のような報告がなされている．山崎ら [26] はストロークの差分角度を特徴パラメータとし，これを P 形フーリエ記述子に変換後，個人の特徴を反映する複数のカテゴリーに分類し，個人の特徴の現れ方に応じてカテゴリーに重みを施すという個人性抽出手法を提案している．また，Kameya ら [32] は，漢字文字列の任意の部分の時間波形に対して CDP(Continuous Dynamic Programming) アルゴリズムを適用して筆者照合を行っている．

しかしながら，オンライン入力された日本語筆跡に対して文字の形態やペンの動きを含めて，どのような特性に個人性が現れるかについては十分に検討されて

いないのが現状である。

また、オンライン入力された英字アルファベット文字の個人性については、ほとんど報告がない。最近になり、Shekhawat ら [38] は 8 文字のアルファベット文字から特徴パラメータを抽出し、その個人性について検討している。しかし、抽出された特徴パラメータは、ストローク数、筆順、ペンの速度、加速度と文字の高さの 5 種類のみであり、筆者照合のためには十分な特徴パラメータの数ではない。

(3) について、従来の研究では偽筆の排除に重点が置かれるあまり、筆跡の形は真似られるが速度や筆圧などの動的特徴は真似できないとの考えから、時系列特徴を用い音声認識で成功した DP マッチングや隠れマルコフモデルなどの手法を導入することが主流になっている。

しかしながら、本来筆跡の真偽を判定するには、まず筆跡の形の異同を検証することが大前提であり、その上で動的特徴の異同を問題にすべきであると筆者は考えている。残念ながら、このような観点から日本語筆跡のオンライン照合を捉えたアプローチは筆者の知る限り存在しない。

(4) については、筆跡全体あるいはストローク単位での特徴抽出が行われているが、文字単位に特徴抽出する試みは無い。しかしながら、一般的に日本語筆跡を書く場合、我々は一文字を単位に筆記している。したがって、文字単位に特徴抽出を行うというアプローチが必要であると考えられる。

(5) に対する解決策は、特徴パラメータに基づく照合方法を用いることであるが、先に概観したように特徴パラメータに基づく照合手法の提案は非常に少ない。

1.3 研究の目的

日本語筆跡を用いたオンライン筆者照合には、前節で述べたような問題が残されている。特に筆者が問題と感じるのは、時系列特徴を用いる方法が主流であるため、筆跡の形（静的特徴）や筆記動作（動的特徴）のどのような特性に個人性が現れるのかということをまず明らかにし、その結果を踏まえてオンライン入力された筆跡の真偽を静的特徴と動的特徴を総合的に用いて判断するというアプ

ローチが欠如していることである。

また、日本語筆跡は文字単位、ストローク単位に楷書体で書かれることが多く、単語単位に続け書きされる英字とは基本的に書字行動に相違がある。このため、欧米での署名照合に用いられる特徴パラメータではなく日本語筆跡の階層構造に注目し、筆跡全体、文字単位、字画単位に静的・動的特性値を抽出することで、筆跡の個人性が表されるのではないかと考えられる。さらに、日本では本人確認に印鑑が用いられることが多いことから、日本語の署名は欧米とは違い他人に真似られないように書くという意識に乏しい。

したがって、オンライン筆者照合を考える際に用いる筆跡としては特に署名に限定するのではなく、より一般的な漢字文字列を対象とすることで、登録筆跡の一般化を図ることが重要である。

そこで、本論文の目的は、

- 署名に限定するのではなく、より一般的な漢字文字列を対象として、どのような特性値に個人性が現れるかをまず明らかにする、
- その結果に基づいたオンライン筆者照合方法を提案し、実験によりその有効性を評価する、

ことである。

抽出する特性値に関しては、筆跡に個人性が現れるのは各個人が習得している書写技能に個人差があるためであると考えられること、また筆跡鑑定の知見が大いに参考になることから、これらに基づく特性値を定義する。また、照合方法については、抽出する特性値が文字列、文字、字画単位で定義され多数であることから、多数の特性値を総合的に用い真偽を判定する識別器が必要になる。そこで、真正筆跡の各特性値は参照筆跡(学習サンプル)の平均値に近いものが多く、遠いものは少ないという考えに基づき、各特性値の参照筆跡の平均値からの偏差の度数分布を基にした識別器を提案する。

漢字文字列を対象としているため署名への適用も可能であるが、一般的な漢字文字列の方が普段書き慣れた署名に比べ個人内変動が大きいと考えられるので、漢字文字列を用いる筆者照合の方が署名照合より困難な課題であるといえる。また漢字文字列を対象とすることで、署名の代わりにシステム共通のパスワードや

利用者が定めたパスワードを用いた個人照合システムや、システムに複数文字登録しておきシステムが提示する文字を順次入力して個人を照合するなど、筆跡を利用した個人照合システムの構築が可能となる。このようなシステムでは、偽造者は登録されている文字列とその筆跡の特徴を盗む必要があり、署名が他人に晒される機会があることを考慮すると、署名を用いる場合よりも偽造が難しいと考えられる。

本論文では、利便性を重視したアクセスコントロールやトラッキングを行う図 1.1 のようなオンライン筆者照合システムを想定している。このシステムは、登録システムと照合システムとから構成される。登録システムでは、まず利用者が ID (Identification) とシステムが設定した漢字数文字で構成された共通のパスワードを複数回 (10 回程度) タブレットから入力する。入力された筆跡から特性値を抽出し特性値ファイルに出力する。つぎに、登録ユーザーから抽出した特性値ファイルから照合に有効な特性値を選択し、照合のための辞書を作成する。照合システムでは、利用者が ID と登録時に入力したシステム共通のパスワードを入力することで、特性値を抽出し辞書と比較する。その比較結果に従って判定を行い本人を確認する。

このシステムでは、システム共通のパスワードを用いるため他人が書いた同一文字列を自動的に複数収集可能であり、筆者照合に有効な特性値を選択するために特別なデータベースを作成する必要がないという利点がある。さらに、システムによるパスワードの更新が定期的に行えるという利点もある。

本研究の成果はオンライン筆者照合だけに留まらず、筆跡に現れる個人性を客観的に計測したことになり、主観的な要素も含まれる筆跡鑑定の分野に対しても貴重な知見を提供できると考えられる。

1.4 本論文の構成

本論文の構成は以下のとおりである。

第 2 章では、本研究のアプローチが適切であるかを見極めるため、漢字に近いストローク構造を持つカタカナ文字を対象にして予備実験を行う。小規模なカタカナ文字列のデータベースに対して、まず筆跡に個人性が現れるのは各個人が習

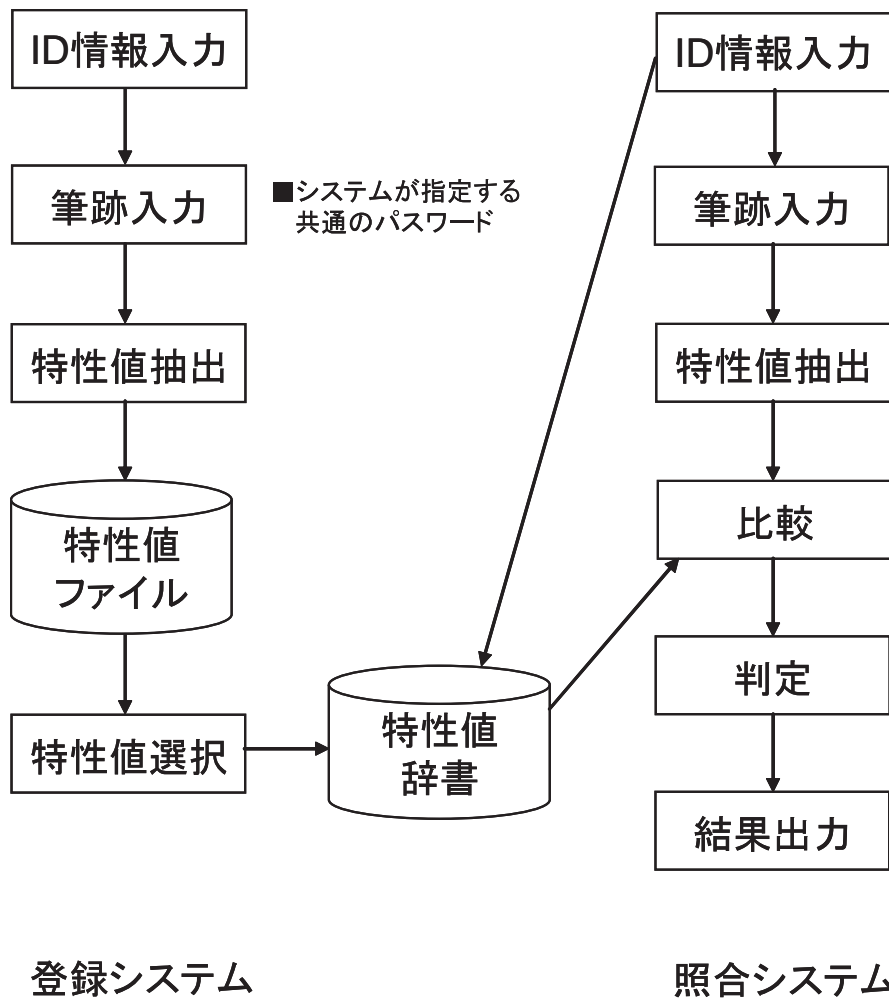


図 1.1 想定する筆者照合システムの処理の流れ

得している書写技能に個人差があるためであると考え、書写技能に基づく特性値を定義し、それら特性値に個人性が現れることを示す。さらに、ユークリッド距離による筆者照合および筆者識別実験を行い、抽出した特性値の筆者識別力を評価する。その結果、本アプローチで個人照合が可能であることを示す。

第3章では、漢字文字列を対象に個人性を表す特性値について検討する。カタカナ文字列で用いた書写技能に基づく特性値は筆跡鑑定の検査項目と類似しているため、特性値を筆跡鑑定の知見に基づいて整理する。つぎに、これら特性値を

抽出し分析することで、個人性が現れやすい特性値を明らかにする。さらに、筆者照合および筆者識別実験を行うことで、これら特性値による筆者照合の可能性を示す。

第4章では、抽出した特性値を用いた筆者照合方法を提案する。抽出した特性値が筆跡全体、文字単位、字画単位、始筆部・送筆部・終筆部単位で定義され多数であることから、特性値の種類ごとに本人間と他人間の距離分布を求め、それに基づいて特性値の選択を行う。つぎに、真正筆跡の各特性値は参照筆跡の平均値に近いものが多く、遠いものは少ないという考えに基づき、各特性値の参照筆跡の平均値からの偏差の度数分布を基にした識別器を提案し、その有効性を筆者照合実験を行うことにより確かめる。あわせて、特徴パラメータに基づく照合で用いられるユークリッド距離による方法および多数決法との比較を行い、両手法より照合精度が高いことを示す。

第5章では、筆者照合における重要な課題である偽造筆跡の排除に対する提案手法の有効性を検討する。本物筆跡のみを用いて設計した識別器に、十分に訓練された偽造筆跡を入力して実験を行い、提案手法が偽造筆跡に対しても有効であることを示す。さらに、第3・4章で用いた漢字文字列以外のデータを収集し、その漢字文字列データに対して同様の筆者照合実験を行い、提案手法の有効性を示す。

最後に第6章で、本研究で得られた成果をまとめるとともに、今後の課題と方向性を示す。

第2章

カタカナ文字列に現れる個人性の抽出

2.1 緒言

筆者認識における基礎的な研究課題として、筆跡のどのような特性に個人性が現れるかを検討し、コンピュータにより機械的、客観的に個人性を抽出することがあげられる。序論でも述べたように、この立場からの研究としてはオフライン方式による吉村ら [33] や尺長ら [34] の研究がある。しかし、オンライン方式の筆者認識については署名照合を中心に研究が行われているため、照合率の高さや偽筆が問題となり、筆跡のどのような特性に個人差があるのか、またそれら特性が個人内でどの程度変動するのかについては明らかにされていないのが現状である。

そこで本章では、漢字に近いストローク構造を持つカタカナ文字を対象にして、本研究のアプローチが適切であるかを見極めるための予備実験を行う。タブレットよりオンラインで入力された小規模なカタカナ文字列のデータベースに対して、まず筆跡に個人性が現れるのは各個人が習得している書写技能に個人差があるためであると考え、書写技能を考察し、書写技能に基づいて機械的に抽出できる特性値を定義する。つぎに、それら特性値が個人間で差があるかどうかを実験的に分析することにより筆跡の個人性がどのように現れるかを明らかにする。

さらに，ユークリッド距離による筆者照合および筆者識別実験を行い，抽出した特性値の筆者識別力を評価する．

以下，2.2 節では書写技能に基づく特性値の定義を行う．つぎに，2.3 節では一元配置分散分析，相関分析，変異係数（標準偏差 / 平均）によって特性値の評価実験を行い，これを分析する．最後に，2.4 節では定義した特性値を用いた筆者認識の可能性を調べるために，識別実験と照合実験を行った結果と検討について述べる．最後に，2.5 節でこの章のまとめを行う．

2.2 特性値の定義

2.2.1 書写技能

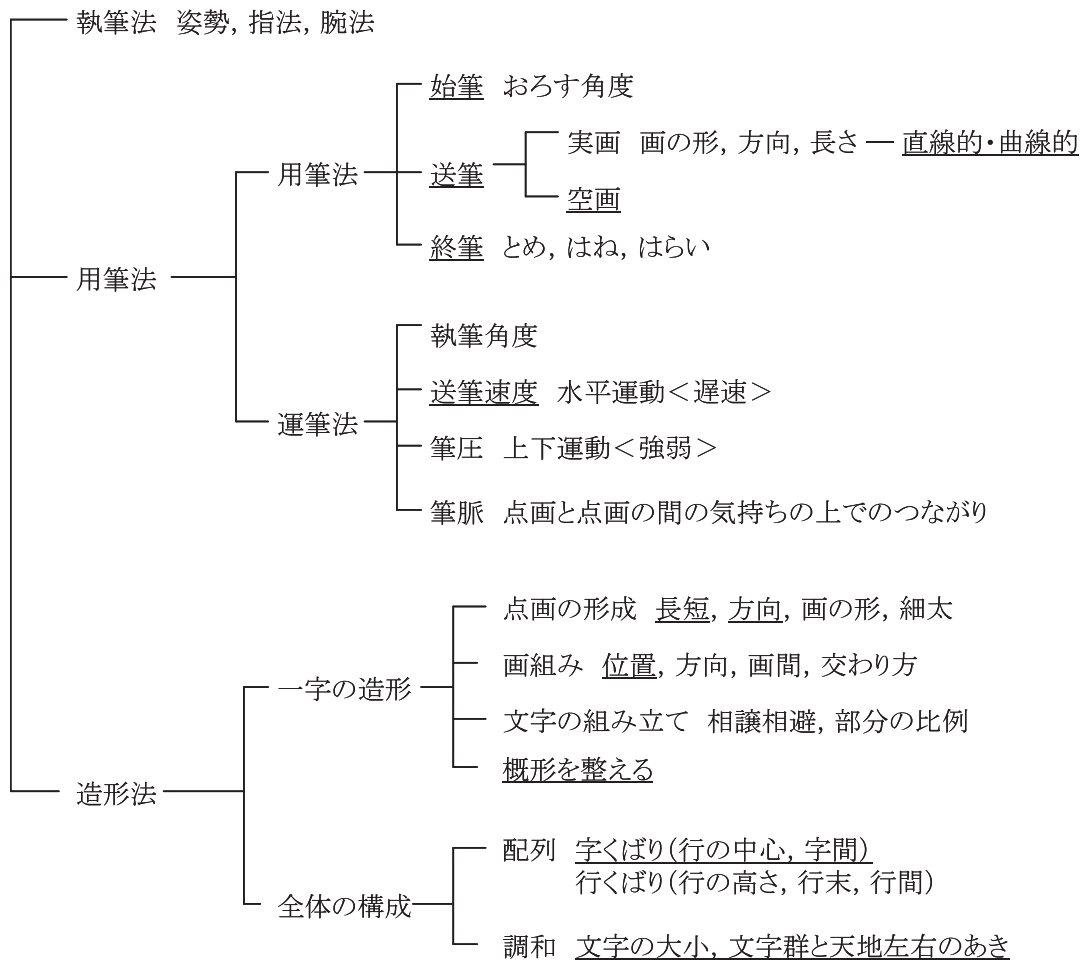
筆跡に個人性が現れる主な要因は，書字学習の過程で生じた書字習慣の差であると考えられる．すなわち，我々は小学生の頃より書写教育に親しみ，より美しく美しい文字を書くための練習が繰り返され，その中で書き癖が固定され，恒常的な個人性が筆跡の中に現れる．したがって，筆跡のどのような特性に個人性が現れるかを明らかにするには，書写技能を考察し，書写技能に基づく特性の抽出が必要である．

書写技能 [39] は，造形法，用筆法，執筆法の 3 つに大きく分類され，概観すると表 2.1 のようになる．

造形法は，点画の長さ・方向・形をどうするのか，点画をどのように組み合わせるのかという一字の造形と，どのような大きさの文字をどのように配列するのかという全体の構成からなる．用筆法は，どのように書き始め，どのような速度・筆圧で点画を書き，最後はどう処理するのか（とめ，はね，はらい）という筆使いに関する技能である．また，執筆法は，書くときの姿勢や筆記用具の持ち方に関するものである．

本研究では，タブレットよりオンラインで入力される筆跡を扱うので，ペンの座標位置およびアップダウン情報の時系列データより機械的に求められる特性を抽出対象とする．したがって，表 1 に示すすべての書写技能を対象とすることはできず，特に執筆法に関しては対象外となる．以下に対象とする特性について述

表 2.1 書写技能の分類



(注) _____は対象とする特性を表す。

べる。なお，具体的な特性値は次節で定義する。

全体の構成に関しては，書写技能とほぼ一対一に対応する特性として，文字の大きさ，文字群と天地左右のあき，文字の中心位置，字間を対象とする。一字の造形に関しては，点画相互間の相対的な長さ・間隔・交差角度などの特性を考慮すると，特性の数が非常に多くなる。そこで，我々が文字を形作る基本となるの

は、点画をどの位置から、どの方向に、どれくらいの長さで書くかであると考え（特に直線的な点画が多い漢字やカタカナを対象とした場合）、点画の始点位置・方向・長さを対象とする。また、文字の概形としては最も基本的な外接長方形を考え、その縦横比を特性とする。

用筆法に関しては、始筆、送筆、終筆の定義が明確ではないため、点画を機械的に始筆部、送筆部、終筆部に分解し各部の方向・速度を特性とすることで書写技能（おろす角度、とめ、はね、はらい）を反映する。また、送筆部については直線的であるか曲線的であるかを送筆部の湾曲の程度で抽出する。筆記面を離れているときのペンの動きである空画については、ペンがアップ状態での座標位置が計測できないため、便宜上、点画の終点から次の点画の始点を直線で結んだものを空画と定義し、その長さ・方向・速度を特性とする。

運筆法に関しては、点画の速度、加速度を特性とする。筆圧、執筆角度については、使用するタブレットとペンからでは計測できないので対象外とする。また、点画と点画の間の気持ちの上でのつながりである筆脈についても対象外である。

2.2.2 特性値の定義

2.2.1 で述べた特性を具体的に定義する前に、必要な記号と用語の説明を行う。対象とする筆跡は、タブレット上の一定枠内に横書きされた複数の文字を、一定の間隔でサンプリングしたものである。この筆跡を、

$$\{x(t), y(t), p(t); \quad t = 1, 2, 3, \dots, N\} \quad (2.1)$$

で表す。ここで、 $(x(t), y(t))$ は枠の左上隅を原点としたペンの座標位置を表す。 $p(t)$ はペンのアップダウン情報を表し、ペンダウンのとき 1、ペンアップのとき 0 となる。 t はサンプリングの番号を表す整数であり、 N は総サンプリング数である。また、各文字を画数をもとに切り出した後、その外接長方形の縦横の長さを 1 に正規化し外接長方形の中心を原点とする座標系に変換した座標系列を、 $x^*(t), y^*(t)$ で表すことにする。さらに、サンプリング間隔を時間の単位としたときのペンの速度を、

$$v(t) = \sqrt{(x(t+1) - x(t))^2 + (y(t+1) - y(t))^2} \quad (2.2)$$

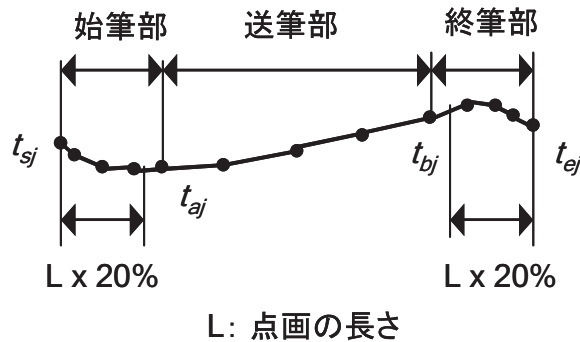


図 2.1 点画の分解

と定義する。

ペンのダウンからアップまでを点画，アップからダウンまでを空画とする。 j 番目の点画の始点を $(x(t_{sj}), y(t_{sj}))$ ，終点を $(x(t_{ej}), y(t_{ej}))$ で表す。また，点画は始筆部，送筆部，終筆部に次のように分解する。始点からの長さが，点画の長さの 20% を初めて超える点 $(x(t_{aj}), y(t_{aj}))$ までを始筆部とする。同様に，終点から始点へ向かったの長さが，点画の長さの 20% を初めて超える点 $(x(t_{bj}), y(t_{bj}))$ までを終筆部とする。したがって，残りの部分が送筆部となる。ただし，点画の長さは座標点列を直線で結んでいった距離である。図 2.1 に点画の分解の様子を示す。始筆部および終筆部を点画の 20% としたのは，収集したデータを観察することで始筆や終筆の状態を表すのに適切であると判断し，それを実験により確かめた結果である。

なお，以下の定義において，添え字 i は文字と字間の番号を，添え字 j は点画と空画の番号を表す。

造形法に関する特性値

- (1) 天地左右のあき： B_t, B_b, B_l, B_r

図 2.2 に示すように，筆記枠と文字列の外接長方形間の x 方向， y 方向の空白長とする。

- (2) 字間： L_i

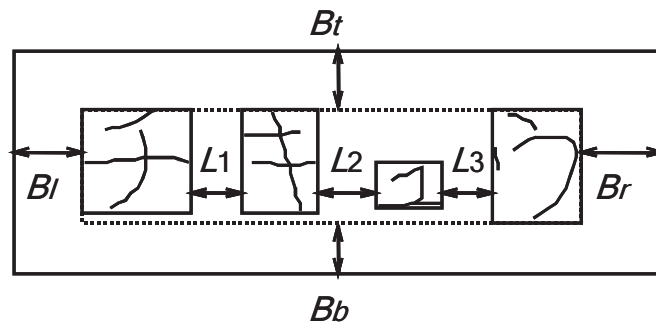


図 2.2 天地左右のあきと字間の説明

図 2.2 に示すように，隣り合う文字の外接長方形間の x 方向の空白長とする．

- (3) 文字の中心位置： (xc_i, yc_i)
外接長方形の中心座標とする．
- (4) 文字の大きさ： S_i

$$S_i = W_i \times H_i \quad (2.3)$$

ただし， W_i ， H_i は，それぞれ外接長方形の幅と高さである．

- (5) 文字の縦横比： R_i

$$R_i = \frac{W_i}{H_i} \quad (2.4)$$

- (6) 点画の始点位置： (xs_j, ys_j)
 $xs_j = x^*(t_{sj})$ と $ys_j = y^*(t_{sj})$ である．
- (7) 点画の方向： $T\theta_j$
 $(x^*(t_{sj}), y^*(t_{sj}))$ と $(x^*(t_{ej}), y^*(t_{ej}))$ を結ぶ直線方向とする．
- (8) 点画の長さ： Tl_j

$$Tl_j = \sum_{t=t_{sj}}^{t_{ej}-1} \sqrt{(x^*(t+1) - x^*(t))^2 + (y^*(t+1) - y^*(t))^2} \quad (2.5)$$

用筆法に関する特性値

(1) 点画の平均速度： Tv_j

$$Tv_j = \frac{\sum_{t=t_{sj}}^{t_{ej}-1} v(t)}{t_{ej} - t_{sj}} \quad (2.6)$$

(2) 点画の最大速度： Tvm_j

$$Tvm_j = \max_t v(t) \quad (2.7)$$

ただし, $t = t_{sj}, t_{sj} + 1, t_{sj} \cdots t_{ej} - 1$ である.

(3) 点画の最大加速度： Ta_{+j}, Ta_{-j}

$$Ta_{+j} = \max_t (v(t+1) - v(t)) \quad (2.8)$$

$$Ta_{-j} = \min_t (v(t+1) - v(t)) \quad (2.9)$$

ただし, $t = t_{sj}, t_{sj} + 1, t_{sj} \cdots t_{ej} - 2$ である.

(4) 空画の方向： $K\theta_j$

$(x(t_{ej}), y(t_{ej}))$ と $(x(t_{sj+1}), y(t_{sj+1}))$ を結ぶ直線の方向とする.

(5) 空画の長さ： Kl_j

$$Kl_j = \sqrt{(x(t_{sj+1}) - x(t_{ej}))^2 + (y(t_{sj+1}) - y(t_{ej}))^2} \quad (2.10)$$

(6) 空画の速度： Kv_j

$$Kv_j = \frac{Kl_j}{t_{sj+1} - t_{ej}} \quad (2.11)$$

(7) 始筆部初期方向： $SI\phi_j$

点画の始点と, 始点からの長さが初めて3(タブレットの分解能を単位として)以上になった点を結ぶ直線の方向とする.

(8) 始筆部方向： $SI\theta_j$

$(x(t_{sj}), y(t_{sj}))$ と $(x(t_{aj}), y(t_{aj}))$ を結ぶ直線の方法とする。

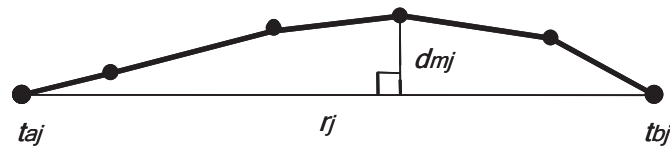
(9) 始筆部平均速度： SIv_j

$$SIv_j = \frac{\sum_{t=t_{sj}}^{t_{aj}-1} v(t)}{t_{aj} - t_{sj}} \quad (2.12)$$

(10) 送筆部曲直： SOc_j

$$SOc_j = \frac{d_{mj}}{r_j} \quad (2.13)$$

ただし、 r_j は図 2.3 に示すように送筆部の始点と終点を結ぶ直線の長さ、 d_{mj} は r_j に対して送筆部の各点から垂線を下ろしたときの最大の長さである。なお、 SOc_j は始点から終点への方角に対して、 d_{mj} が右にあるときは正、左にあるときは負の値をとる。



$$SOc_j = d_{mj} / r_j$$

図 2.3 送筆部曲直の説明

(11) 送筆部方向： $SO\theta_j$

$(x(t_{aj}), y(t_{aj}))$ と $(x(t_{bj}), y(t_{bj}))$ を結ぶ直線の方法とする。

(12) 送筆部平均速度： SOv_j

$$SOv_j = \frac{\sum_{t=t_{aj}}^{t_{bj}-1} v(t)}{t_{bj} - t_{aj}} \quad (2.14)$$

(13) 終筆部最終方向： $SY\phi_j$

点画の終端 2 点間の方向とする。ただし，2 点間の距離が 2 以下の場合
は， $SY\phi_j = 0$ とする。

(14) 終筆部最終速度： SYs_j

$$SYs_j = v(t_{ei} - 1) \quad (2.15)$$

(15) 終筆部方向： $SY\theta_j$

$(x(t_{bj}), y(t_{bj}))$ と $(x(t_{ej}), y(t_{ej}))$ を結ぶ直線の方向とする。

(16) 終筆部平均速度： SYv_j

$$SYv_j = \frac{\sum_{t=t_{bj}}^{t_{ej}-1} v(t)}{t_{ej} - t_{bj}} \quad (2.16)$$

2.3 特性値の評価

2.3.1 評価に用いるデータ

評価に用いるデータは，4 人の筆記者 (A, B, C, D) が，タブレット上に置いたデータ用紙の枠内 ($12\text{mm} \times 60\text{mm}$) に，「チキュウ」という文字を筆記したものであり，各人からそれぞれ 50 個収集した。データは，コードレスのスタイラスペン (ボールペン芯) を持つタブレットより， 0.1mm/point ， 100points/sec で入力した。図 2.4 にデータの一部を示す。

データとしてカタカナを用いたのは，カタカナの用筆法が漢字と同じであることと，カタカナの方が漢字より個性が現れにくいと考えられることによる。すなわち，カタカナで個性が現れれば，漢字については同程度かそれ以上の個性が現れることが期待できる。また，文字数を 4 文字にしたのはオンライン筆者照合で利用される署名が 4~5 文字であることから，それと同程度の文字数とした。「チキュウ」という文字を設定したのは，カタカナの中では画数の多い 3 画の文字を中心に多様な点画が含まれ，かつ意味のある単語という理由による。また，ボールペン芯を持つコードレスのスタイラスペンを使用したのは，普通の筆記状態に近い環境でデータを収集するためである。

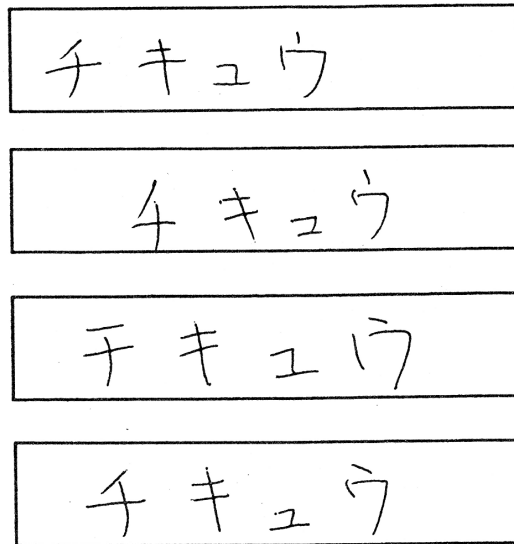


図 2.4 評価に用いた筆跡の例

2.3.2 評価方法

定義した各特性値を，

- (1) 個人間で差があるかどうか
- (2) 個人内の特性値のばらつきはどの程度か
- (3) 特性値間に強い相関があるかどうか

の観点から評価する．(1) については，特性値ごとに筆者 4 水準の一元配置分散分析 [40] を行い，その F 値の大きさにより評価する．(2) については，筆者ごとに各特性値の変異係数（標準偏差 / 平均値）を計算し，その値により評価する．(3) については，筆者ごとにすべての特性値の組み合わせについて相関係数を計算することにより評価する．

なお，特性値の数は，文字数が 4，点画の数が 11（添え字 j は「チ」第 1 画を 1，「ウ」第 3 画を 11 とする）であるので，全部で 251 個となる．ただし，4 人

の筆者の内 1 人が「チ」第 1 画を他の 3 人とは逆（左から右）に書くという特異な書き方をするので，これによる評価への影響を除去するため，「チ」第 1 画および第 1 空画に関する特性値は除く．したがって，評価する特性値の数は 230 個となる．

2.3.3 評価結果

分散分析の結果と検討

分散分析の結果，230 個の特性値の内 223 個で F 値が有意水準 5% の限界値 $F_{0.05}(3, 196) = 2.65$ より大きくなり，定義した特性値が個人間で有意な差があることが明らかになった．表 2.2，表 2.3 にその結果を示す．F 値は水準間の不偏分散と水準内の不偏分散の比として定義されるので，F 値が大きいことは，個人内の変動に比べて個人間の変動が大きいことを意味する．すなわち，F 値が大きい特性値ほど個人差を表しているといえる．このことをもとに検討を加える．造形法と用筆法に関する特性値を比較すると，平均的には造形法に関する特性値の方が用筆法より個人性が現れやすい．しかし，F 値の最大は平均ほど差がなく，特定の点画や空画において用筆法は造形法と同程度に個人性を表すといえる．表 2.2 より，点画や空画ごとに定義される特性値において，F 値の最大，最小の差が大きい．また，表 2.3 より F 値が 200 を越える特性値のほとんどは，「チ」と「ウ」に関するものである．このことは，文字や点画により個人性の現れ方に大きな差があることを示している．

つぎに，始筆部，送筆部，終筆部の特性について検討する．送筆部の方向・平均速度が点画のそれらに比べ F 値が大きく，また，終筆部の最終速度・平均速度の F 値も点画の平均速度と同程度かそれ以上である．これは，点画を分解することにより，個人性に関してより良い情報が得られることを示唆している．始筆部の F 値が，送筆部，終筆部に比べて小さい．始筆の方向については，書写教育において硬筆では素直な始筆が求められるため，個人差が現れにくいと考えられる．また，平均速度については，始筆部ではペンの移動速度が遅いため，移動速度の速い送筆部，終筆部に比べて個人差が現れにくいといえる．なお，送筆部の曲直の F 値が小さいが，これは対象とした文字に含まれる点画がほとんど直線

表 2.2 各特性値の F 値の平均，最大，最小

特 性 値	平均	最大	最小
天地左右のあき： B_t, B_b B_l, B_r	3.28 220.98	5.28 364.74	1.28 77.22
字間： L_i	13.25	25.43	3.81
文字の中心位置： xc_i yc_i	151.77 8.16	269.89 12.00	68.23 2.82
文字の大きさ： S_i	115.37	152.15	73.85
文字の縦横比： R_i	89.93	108.36	68.62
点画の始点位置： xs_j ys_j	80.25 103.18	193.68 450.45	7.59 5.14
点画の方向： $T\theta_j$	118.23	711.50	9.40
点画の長さ： Tl_j	109.92	222.83	17.12
点画の平均速度： Tv_j	38.48	130.85	2.97
点画の最大速度： Tvm_j	50.27	122.24	11.75
点画の最大加速度： Ta_{+j} Ta_{-j}	15.42 15.37	45.08 37.75	3.54 3.57
空画の方向： $K\theta_j$	74.97	179.58	24.29
空画の長さ： Kl_j	126.03	446.54	9.18
空画の速度： Kv_j	51.98	163.27	1.75
始筆部初期方向： $SI\phi_j$	19.98	39.23	1.70
始筆部方向： $SI\theta_j$	32.16	71.33	6.07
始筆部平均速度： SIv_j	24.18	75.40	1.97
送筆部曲直： SOc_j	17.82	40.54	1.57
送筆部方向： $SO\theta_j$	140.57	939.23	5.27
送筆部平均速度： SOv_j	76.71	240.21	5.34
終筆部最終方向： $SY\phi_j$	37.42	130.97	1.65
終筆部最終速度： SYs_j	53.20	136.40	10.01
終筆部方向： $SY\theta_j$	47.24	212.62	0.72
終筆部平均速度： SYv_j	47.63	96.38	10.57

表 2.3 F 値が 200 以上の特性値

F 値が 200 以上の特性値, () 内はその F 値					
$SO\theta_9^*$	$T\theta_9^*$	ys_3^*	Kl_9^*	Bl^*	Kl_{10}^*
(939.2)	(711.5)	(450.5)	(446.5)	(364.7)	(350.7)
xc_1	SOv_2^*	Tl_3	$SY\theta_9$	SOv_8	
(269.9)	(240.2)	(222.8)	(212.6)	(203.8)	

(注) *印は少なくとも 2 人の筆者間で分布が完全に分離する特性値を表す。

的な用筆で書かれるからであると考えられる。

上記以外の特徴的な事柄を述べる。まず、天地のあき、文字の中心位置の y 座標の F 値が左右のあきや x 座標に比べ極端に小さくなっている。これは、筆記された文字が一定枠内に横書きされ、縦方向への文字の配置の自由度が横方向に比べ小さいためと考えられる。また、空画の F 値が点画と同程度の大きさとなった。このことは、単に視覚的に見える筆跡だけではなく、ペンのアップ状態での動きも個人性を問題にする場合には重要であることを示している。

以上、F 値をもとに検討を行ってきたが、F 値が大きいことと 4 人の筆者間に差があることは等価ではない。すなわち、1 人の筆者のみに差があれば F 値は大きくなる。そこで、筆者間の特性値の分布を調べてみた。その結果、4 人の筆者を分離するような特性値は無かった。しかし、7 個の特性値(表 2.3)で少なくとも 2 人の筆者間の分布が完全に分離していた。

変異係数とその検討

表 2.4 に各特性の筆者ごとの変異係数の平均を示す。変異係数は標準偏差を平均で割ったものであるから、特性値の分布のばらつきを示す目安と考えることができる。すなわち、変異係数が小さいほどばらつきも小さいといえる。

文字の中心位置・縦横比、点画の長さ・平均速度、空画の方向・長さの変異係数が他の特性値に比べて全筆者で小さくなった。これは、一定枠が与えられるとこれに対する文字の配置や大きさが無意識の内にイメージされ、そのイメージどおりに筆記しようとするからではないかと思われる。逆に、始筆部初期方向・方

表 2.4 各特性値の筆者ごとの変異係数（標準偏差 / 平均値）の平均

特 性 値	筆者 A	筆者 B	筆者 C	筆者 D
天地左右のあき： B_t, B_b B_l, B_r	0.365 0.191	0.345 0.126	0.538 0.201	0.437 0.115
字間： L_i	0.242	0.238	0.302	0.222
文字の中心位置： xc_i yc_i	0.097 0.062	0.070 0.076	0.082 0.070	0.063 0.057
文字の大きさ： S_i	0.207	0.185	0.153	0.141
文字の縦横比： R_i	0.130	0.137	0.131	0.115
点画の始点位置： xs_j ys_j	0.373 0.243	0.587 0.336	1.224 0.609	0.509 0.258
点画の方向： $T\theta_j$	0.224	0.275	0.222	0.304
点画の長さ： Tl_j	0.075	0.119	0.078	0.065
点画の平均速度： Tv_j	0.171	0.178	0.143	0.119
点画の最大速度： Tvm_j	0.186	0.203	0.201	0.170
点画の最大加速度： Ta_{+j} Ta_{-j}	0.295 0.350	0.333 0.342	0.324 0.329	0.314 0.277
空画の方向： $K\theta_j$	0.084	0.125	0.090	0.067
空画の長さ： Kl_j	0.145	0.175	0.139	0.135
空画の速度： Kv_j	0.234	0.239	0.177	0.223
始筆部初期方向： $SI\phi_j$	6.060	0.832	1.208	0.512
始筆部方向： $SI\theta_j$	6.196	0.766	3.169	1.051
始筆部平均速度： SIv_j	0.319	0.317	0.294	0.223
送筆部曲直： SOc_j	3.324	1.215	4.809	1.616
送筆部方向： $SO\theta_j$	0.250	0.333	0.260	0.281
送筆部平均速度： SOv_j	0.162	0.209	0.156	0.138
終筆部最終方向： $SY\phi_j$	2.971	1.116	2.818	1.898
終筆部最終速度： SYs_j	0.573	0.544	0.949	1.252
終筆部方向： $SY\theta_j$	0.860	1.008	1.313	0.549
終筆部平均速度： SYv_j	0.300	0.341	0.317	0.258
全平均	1.032	0.430	0.835	0.457

向，送筆部曲直，終筆部最終方向・最終速度・方向については，変異係数が全筆者で大きくなった．始筆部，終筆部での方向の変異係数が大きくなる理由としては，それぞれ点画の長さの 20% を範囲としているので方向を考える 2 点間の距離が短くなり，わずかの座標のずれでも方向に変動が生ずることが考えられる．送筆部曲直については直線的な点画が多く，そのほとんどの平均値が 10^{-2} ~ 10^{-3} のオーダーとなり，わずかの特性値の変動によっても変異係数の値が大きくなると考えられる．

筆者間の変異係数の差をみると，全平均で筆者 A，C が B，D の 2 倍程度になっている．これは，始筆部初期方向・方向，送筆部曲直，終筆部最終方向の変異係数が大きいことが原因であり，始筆や終筆，送筆の形などを比較的安定に書く人と，そうでない人がいることを示している．

相関分析の結果と検討

表 2.5，2.6 に結果を示す．

相関が強い組み合わせは，左のあきと 1 文字目の中心位置 x 座標，右のあきと 4 文字目の中心位置 x 座標，点画の方向と送筆部の方向など，事前にある程度の相関が予想されるものがほとんどである．点画の方向と送筆部の方向は，表中の組み合わせ以外の点画においても全筆者で比較的強い相関（相関係数が最低で 0.533，平均 0.761）を示した．しかし，点画の平均速度と送筆部の平均速度，終筆部の最終速度と平均速度については，平均的には相関が強い（相関係数の平均 0.706 と 0.708）が筆者や点画によって相関の弱いものがあつた．

前後の文字の中心位置 x 座標間に強い相関がある．1 文字目と 2 文字目の相関も強く全筆者で相関係数が 0.633 以上であつた．これは文字の位置が直前の文字の位置に強く影響を受けることを示す．特に 3 文字目と 4 文字目の相関が強いのは，3 文字目が拗音であることで文字間の拘束が強いためではないかと考えられる．

表 2.6 より，筆者に依存して相関が強い特性値の組み合わせがあること，および筆者によってはそのような組み合わせが存在しないことが確認できた．したがって，このこと自体が個人性を表しているとも考えられる．

表 2.5 全筆者で相関係数 r が $|r| \geq 0.671$ (母相関係数 $\rho = 0.8$ の信頼率 95% の信頼限界) の特性値の組み合わせ

特性値の組み合わせ			
$B_l - xc_1^*$	$B_r - xc_3^*$	$B_r - xc_4^*$	
$L_1 - Kl_3$	$L_2 - Kl_6^*$	$L_3 - Kl_8$	$L_3 - K\theta_8$
$xc_2 - xc_3$	$xc_3 - xc_4^*$		
$T\theta_2 - SO\theta_2$	$T\theta_5 - SO\theta_5$	$T\theta_8 - SO\theta_8$	$T\theta_9 - SO\theta_9$
$T\theta_{10} - SO\theta_{10}$			
$Tv_7 - SOv_7$	$Tv_{11} - SOv_{11}^*$	$Tv_3 - SIv_3$	
$SYs_3 - SYv_3^*$	$SYs_8 - SYv_8$	$SYs_{11} - SYv_{11}$	
$Kl_{10} - Kv_{10}$			

(注) *印は $|r| \geq 0.829$ ($\rho = 0.9$ の信頼率 95% の信頼限界) の特性値の組み合わせを表す。

表 2.6 筆者単独で相関係数 r が $|r| \geq 0.829$ ($\rho = 0.9$ の信頼率 95% の信頼限界) の特性値の組み合わせ (他の筆者は $|r| < 0.671$)

筆者	特性値の組み合わせ
A	$Tv_6 - Kv_8$ $Tv_6 - SIv_3$
B	なし
C	$xs_{10} - SI\phi_{10}$ $Kl_7 - Kv_7$ $Kl_8 - K\theta_8$ $SY\phi_3 - SYs_3$
D	$SY\phi_{11} - SYs_{11}$

2.4 特性値の筆者識別力の検討

2.4.1 実験方法

2.3 節において、書写技能に基づいて定義した特性値に個人性が現れることを明らかにした。この節では、定義した特性値を用いて筆者の認識が可能であるか

どうかを検討するため，筆者識別実験ならびに照合実験を行う．

識別実験の方法

実験に用いるデータは評価に用いたデータと同じであり，奇数番目を参照筆跡，偶数番目を未知筆跡とした．識別は，分散の逆数で重み付けしたユークリッド距離を用いて行った．未知筆跡の筆者 i に対する距離を

$$D_i = \sqrt{\sum_{j=1}^n \frac{(f_j - m_{ij})^2}{\sigma_{ij}^2}} \quad (2.17)$$

n : 特性値の数

f_j : 未知筆跡の j 番目の特性値

m_{ij}, σ_{ij}^2 : 参照筆跡より求めた筆者 i の j 番目の特性値の平均と分散

で計算し，最小の距離を与える筆者を未知筆跡の筆者と判定する．

筆者識別はテキスト独立型とテキスト依存型に分類できる．そこで，本実験では「チキュウ」という 4 文字に共通あるいは依存しない特性値（天地左右のあき，字間，文字の中心位置・大きさ・縦横比）を用いた場合をテキスト独立型，その他の特性値を用いた場合をテキスト依存型として実験を行った．なお，テキスト依存型の識別では，文字数を 1 文字から 4 文字までのすべての組み合わせで実験した．

照合実験の方法

実験に用いるデータは識別実験と同じである．4 人の筆者それぞれについて，その筆者の未知筆跡を本物，他の 3 人の未知筆跡を偽物として実験を行う．したがって，本物 100 個（25 個 × 4 人），偽物 300 個（25 個 × 3 人 × 4 人）となる．

照合は，(2.17) 式で定義された距離 D_i を用いて行い，しきい値 T_i に対して $D_i \leq T_i$ ならば受理し，そうでなければ棄却する． T_i は本物の筆跡のみから決定するという立場から，筆者 i の参照筆跡内で距離 D を求め，その平均と標準偏差を D_{mi} ， D_{si} とし，

$$T_i = D_{mi} + w \cdot D_{si} \quad (2.18)$$

とする．ここで w は重み係数である．本物の距離 D は，正規分布に近い分布を

すると仮定すれば，ほぼ $D_{mi} + 4D_{si}$ の範囲内に収まると考えられる．したがって，本実験では $w = 4.0$ とした．

なお，実験は識別実験と同様にテキスト独立型と依存型について行った．

2.4.2 実験結果

識別実験の結果

表 2.7, 2.8 に識別実験の結果を示す．

テキスト独立型の識別率は 93% であり，これは単一文字での識別率とほぼ同程度である．しかし，文字の外接長方形の大きさと配置のみでの識別であることを考えると高い識別率であるといえる．テキスト依存型では，文字数の増加による識別率の向上が見られ，3 文字以上で 100% の識別率が得られた．また，全特性値を用いた場合も 100% の識別率が得られ，定義した特性値は筆者識別に有効であることが明らかになった．

文字ごとの識別結果を見ると，文字によって識別率に差があり，個人性の現れ方が文字によって異なる．特に，F 値 200 以上の特性値を多く持つ「ウ」では 100% の識別率が得られた．したがって，参照文字数の少ない筆者識別においては字種の選び方が問題となる．

表 2.7 識別結果

使用した特性値		識別率 (%)
テキスト独立		93.0
テキスト依存	1文字	96.8
	2文字	99.8
	3文字	100.0
	4文字	100.0
全特性値		100.0

表 2.8 文字ごとの識別率

「チ」	「キ」	「ユ」	「ウ」
98%	96%	93%	100%

表 2.9 照合結果

使用した特性値		FRR (%)	FAR (%)
テキスト独立		1.0	25.3
テキスト依存	1文字	1.5	24.6
	2文字	1.5	7.4
	3文字	1.3	1.5
	4文字	1.0	0.0
全特性値		0.0	0.0

照合実験の結果

表 2.9 に照合結果を示す。

テキスト独立型では識別実験と同様に、1文字による照合と同程度の照合率が得られた。しかし、偽物を誤って受理する他人受理率 FAR (False Acceptance Rate) が 25.3% と高く十分な照合結果とはいえない。テキスト依存型では文字数の増加による FAR の低下が顕著であり、4文字すべてを用いると FAR が 0% となった。これは、文字数の増加により偽物の距離が大きくなることを意味する。一方、本物を誤って棄却する本人拒否率 FRR (False Rejection Rate) は使用する特性値によらず 0% から 1.5% であった。この結果は、しきい値決定の際の仮定とよく一致している。

全特性値使用の場合、FRR、FAR とともに 0% となった。また、テキスト依存型で 4文字使用の場合、重み係数 w を変化させて実験を行った結果、 w が 4.2 から 4.9 の範囲で FRR、FAR とともに 0% であった。以上の結果から、4文字以上を用いれば、定義した特性値による筆者照合が有効であることが明らかになった。

2.5 結言

筆者認識における基礎的な研究として、タブレットよりオンラインで入力された筆跡に現れる個人性を、書写技能に基づく特性により抽出することを提案した。まず、書写技能を考察しコンピュータにより機械的に抽出可能な特性値を定義した。これら特性値を一元配置分散分析、変異係数、相関分析を用いて評価し、

- 文字や点画の違いにより現れ方に差があるものの、定義した特性値は個人性を表していること、
- 文字の配置や大きさに関する特性値の個人内でのばらつきは小さく、始筆部や終筆部の方向や送筆部の曲直などの局所的な特性値の個人内のばらつきが大きいこと、
- 始筆・送筆・終筆の形などを比較的安定に書く人と、そうでない人がいること、
- 筆者に依存せず相関が強い特性値の組み合わせと、筆者に依存して相関の強い組み合わせがあること

を示した。さらに、定義した特性値を用いてどの程度の筆者認識が可能であるかを検討するために、筆者識別実験と照合実験を行った。その結果、筆記者4人と小規模なデータベースではあるが、識別率100%、FRR、FARはともに0%であった。

以上の結果から、オンライン方式で入力された一定枠内に書かれた日本語筆跡に対して、筆跡の形態的特徴や動的特徴を表す特性値を用いて筆者認識を行える可能性が明らかになった。

第3章

漢字文字列に現れる個人性の抽出

3.1 緒言

第2章では、本研究のアプローチの有効性を確認するための予備実験として、筆跡に個人性が現れるのは各人が習得している書写技能に個人差があるためと考え、カタカナ文字に対して書写技能に基づく特性値を定義し、それら特性値が個人差を表すことを小規模なデータベースを用いて明らかにした。

本章では、より一般的な漢字文字列を対象にし、タブレットよりオンラインで入力された漢字文字列に現れる個人性の抽出と分析を行う。特性値については、前章では書写技能に基づいて定義したが、書写技能の要素は筆跡鑑定の検査項目や研究対象とされる書字運動とほとんど同じであり、ここでは今後の筆者照合への適応を考え、特性値数を極端に多くしないという観点から、筆跡鑑定の知見に基づいて特性値を整理した。また、筆跡鑑定の分野では筆記時間、筆記速度やペンの傾きなどの書字運動が研究対象となっており、これらについては特性値を追加した。

すなわち、筆圧やペンの傾きも含めて静的特徴と動的特徴を総合的に抽出するという立場から、筆跡鑑定の知見に基づく特性値により個人性を抽出することを提案する。そのため、前章ではペン座標とペンのアップダウン情報のみが入力可

能なタブレットを使用した。ここでは筆圧やペンの傾きも計測できるタブレットを使用する。

以下、3.2節で筆跡鑑定の検査項目である配字、字画形態、字画構成、筆圧の強弱、および筆跡鑑定に関連する研究で対象とされる筆記時間、筆記速度、ペンの傾きに着目し機械的に計測可能な特性値を定義する。つぎに、3.3節でこれら特性値が個人間で差があるかどうかを評価する。特性値ごとの分析は分散分析とクラスカル・ウォリス検定により、同じ種類の特性値を組み合わせた場合の評価は特性値空間での筆者内（本人間）距離と筆者間（他人間）距離の分布に基づいて行う。3.4節で特性値を用いた筆者認識の可能性を検討するため筆者識別・照合実験を行う。最後に、3.5節でこの章のまとめを行う。

3.2 特性値の定義

本研究で対象とする筆跡は、タブレット上の一定枠内に横書きされた複数の漢字である。なお、前節で述べたように使用するタブレットはカタカナを収集したタブレットとは異なり、一定の時間間隔でサンプリングしたペンの座標位置 $(x(t), y(t))$ 、筆圧 $p(t)$ とペンの高度 $al(t)$ 、方位 $az(t)$ の時系列データが得られる。図 3.1 にペンの高度と方位の定義を示す。

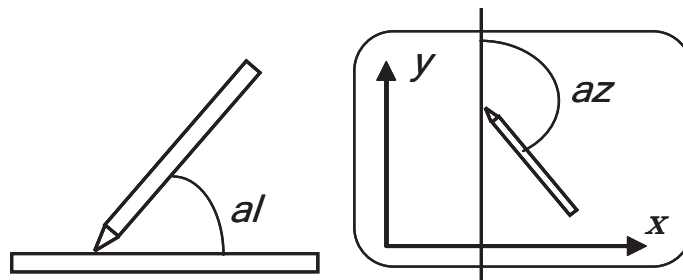


図 3.1 ペン高度 al と方位 az

3.2.1 筆跡鑑定の知見

筆跡鑑定の検査項目

筆跡鑑定の主要な検査項目は、配字、字画形態と字画構成である [41, 42] .

なお、書写技能では文字の各ストロークを点画と呼ぶが筆跡鑑定では字画と呼ぶため、今後は字画で統一する .

(1) 配字

文字の配列や枠と文字の位置関係など、文字の配置のことを配字という .
これに関しては、字間隔、文字の位置、文字の大きさ、文字の外形などが検査項目となる .

(2) 字画形態

字画の長短、湾曲、直線性、折れ、曲がり、止め、はね、はらいなどのような字画の部分的違いや、それらの部分の組み合わせ方によって、同じ字画が様々に変化する . このような、字画の形を字画形態という . 字画形態の検査では、入筆方向、反り、長さ、終筆の状態、折れ、曲がりなどが観察される .

(3) 字画構成

一文字中の字画相互間の組み立て方を字画構成という . 字画構成の検査項目としては、字画の交差位置、交差角度、字画の接合位置、字画相互間の長さ、字画間隔、字画の続け書きの有無などがある .

(4) 筆圧

既に書かれた筆跡から筆圧を定量的に測定できるものではないが、筆跡鑑定の実務においては必ず考慮されるものである . とくに、一字画内の始筆、送筆、終筆での相対的筆圧（筆圧の強弱）や、一文字内での字画間の相対的筆圧が観察される .

書字運動

筆跡鑑定分野では書字運動に関する研究が行われ、筆圧以外にも筆記速度や筆記時間、ペンの傾きについても検討されている [41, 42] .

3.2.2 特性値の定義

以下の観点から特性値を定義する .

先に述べた筆跡鑑定の検査項目は書写技能の要素と基本的に同じであるため、配字、字画形態、字画構成については、前章で提案した書写技能に基づく特性値を整理して用いる . 書写技能に基づく特性値から除外したものは、天地左右のあきと始筆部初期方向、終筆部最終方向と空画に関するものである . 天地左右のあきは文字の中心位置と強い相関があることが前章で明らかになったためであり、始筆部初期方向と終筆部最終方向については局所的な特性値は変動が大きいことと、始筆部と終筆部の方向に吸収できると考えたことによる . 空画については、筆跡鑑定の対象となっていないことと、特性値の数を極端に多くしないという考えからである . なお、曲直と方向については局所的な変動の影響を受けにくくするため、曲直はサンプリング点間の角度差分の平均値とし、方向はサンプリング点間の角度の平均値として再定義した . また、曲直と長さは抽出範囲を拡大し、曲直については送筆部のみでなく、文字列、文字、字画、始筆・送筆・終筆部のすべてで、長さについては字画以外に文字列と文字についても抽出した . その理由は、これら二つの特性値は、文字列や文字などの大局的部分においても有効であると考えられるからである .

動的特性値については、筆跡鑑定の検査項目である相対筆圧を反映する筆圧順位、筆圧順位を速度に拡張した速度順位、一文字内での各字画を書き始めるタイミングを表す筆記タイミングを定義する . また、タブレットから容易に得られる筆記時間、筆記速度、筆圧、ペンの傾きを用いる . 筆記速度以外は、書写技能に基づく特性値には含まれていない、新たに追加した特性値である . 特に筆圧順位、速度順位、筆記タイミングは従来の特徴パラメータに基づく筆者照合では用いられていない特性値であり、また、筆圧、ペンの傾き、筆記速度の平均値を文

字列（筆跡全体）、文字、字画、始筆・送筆・終筆部ごとに定義し用いることも、従来の特徴パラメータに基づく筆者照合では行われていない。なお、書写技能に基づく特性値の内、点画の最大加速度・最大速度および終筆部最終速度は特性値の数を多くしないという観点から除外した。

定義した特性値を、文字列（筆跡全体）、文字、字画、始筆・送筆・終筆部に關するものに分類し表 3.1 に示す。

表 3.1 定義した特性値一覧

表中, Y は定義した特性値を表す。

	特性値	文字列	文字	字画	始・送・終筆部		
					始筆	送筆	終筆
動的	筆記時間	Y	Y	-	-	-	-
	ペンダウン時間	Y	Y	Y	-	-	-
	ペンアップ時間	Y	Y	-	-	-	-
	筆記タイミング	-	-	Y	-	-	-
	筆圧順位	-	-	Y	Y	Y	Y
	平均筆圧	Y	Y	Y	Y	Y	Y
	平均ペン高度	Y	Y	Y	Y	Y	Y
	平均ペン方位	Y	Y	Y	Y	Y	Y
	速度順位	-	-	Y	Y	Y	Y
	平均筆記速度	Y	Y	Y	Y	Y	Y
静的	大きさ	Y	Y	-	-	-	-
	縦横比	Y	Y	-	-	-	-
	中心位置	Y	Y	-	-	-	-
	字間隔	Y	-	-	-	-	-
	相対始点位置	-	-	Y	-	-	-
	長さ	Y	Y	Y	-	-	-
	曲直	Y	Y	Y	Y	Y	Y
方向	-	-	Y	Y	Y	Y	

これら特性値を抽出するために、筆跡を筆圧情報により字画に分解後、字画の位置情報と筆順に基づいて文字単位に分割する。さらに、字画の始点からの長さが字画長さの 20% を始めて超えるサンプリング点までを始筆部とする。同様に、字画の終点から始点に向かう長さが 20% を始めて超える点までを終筆部とし、残りの部分を送筆部と定義する。図 3.2 に文字を切り出し外接長方形で抽出した様子を、図 3.3 に字画の分割の様子を示す。

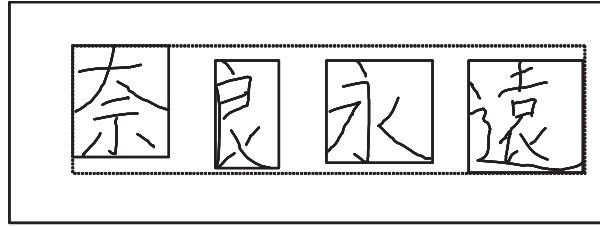


図 3.2 文字切り出し

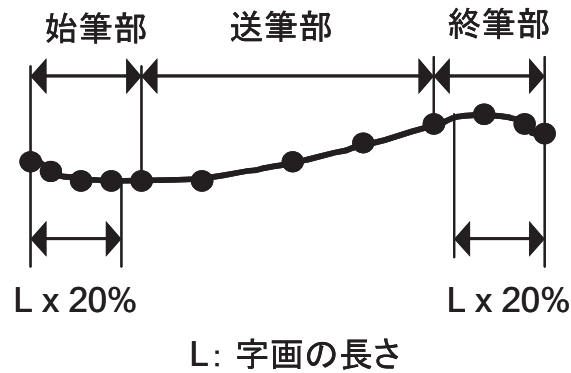


図 3.3 始筆部，送筆部，終筆部への分割

以下に，各特性値について説明する．

- (1) 筆記時間
書き始めから書き終わりまでの時間とする．
- (2) ペンダウン時間
筆圧 $p(t) > 0$ の時間とする．
- (3) ペンアップ時間
筆圧 $p(t) = 0$ の時間とする．
- (4) 筆記タイミング
一文字内の第 1 画目の書き始めを時間の起点としたとき，第 2 画目以降の各字面を書き始める時刻とする．ただし，一文字の筆記時間が 1 となるよう正規化を行っている．図 3.4 に筆記タイミングの定義を示す．

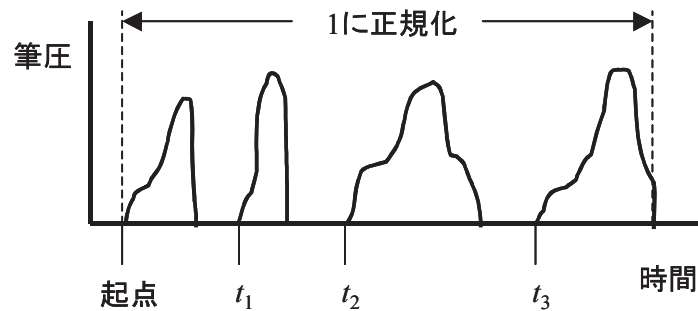


図 3.4 筆記タイミング

- (5) 筆圧順位

一文字内での各字画の平均筆圧を求め、一番大きい字画に 1、次の字画に 2 と順位付けを行ったものである。同様に、始筆・送筆・終筆部についても、各部の平均筆圧を求め順位付けを行う。
- (6) 平均筆圧

ペンダウン時の筆圧の平均値とする。
- (7) 平均ペン高度

ペンダウン時のペン高度の平均値とする。
- (8) 平均ペン方位

ペンダウン時のペン方位の平均値とする。
- (9) 速度順位

一文字内での各字画の平均筆記速度を求め、一番大きい字画に 1、次の字画に 2 と順位付けを行ったものである。同様に、始筆・送筆・終筆部についても、各部の平均筆記速度を求め順位付けを行う。
- (10) 平均筆記速度

ペンダウン時のサンプリング点間を順次直線で結んでいった距離の和を、ペンダウン時間で割った値とする。
- (11) 大きさ

外接長方形の面積とする。
- (12) 縦横比

外接長方形の縦と横の長さの比で表す。

(13) 中心位置

外接長方形の中心の xy 座標で表す。ただし, xy 座標の原点は筆記枠の左下隅とする。

(14) 字間隔

各文字を外接長方形で切り出したとき, 隣り合う外接長方形の x 方向の空白長で表す。

(15) 相対始点位置

文字の外接長方形の長辺が 1 となるように大きさの正規化を行い, 外接長方形の中心座標を原点としたとき, 各字画の始点の xy 座標で表す。

(16) 長さ

ペンダウン時のサンプリング点間を順次直線で結んでいった距離の和とする。

(17) 曲直

x 方向を基準としたとき, 筆跡の曲直をペンダウン時のサンプリング点間を結ぶ直線がなす角度 θ (反時計回りを正とする) の角度差分の平均値として図 3.5 のように定義する。ただし, 湾曲の方向により曲直は正負の値をとるので, 正の値のみを対象とする Φ_+ と, 負の値のみを対象とする Φ_- とに分けることとした。

(18) 方向

図 3.5 に示すサンプリング点間を結ぶ直線の角度 θ の平均値として定義する。

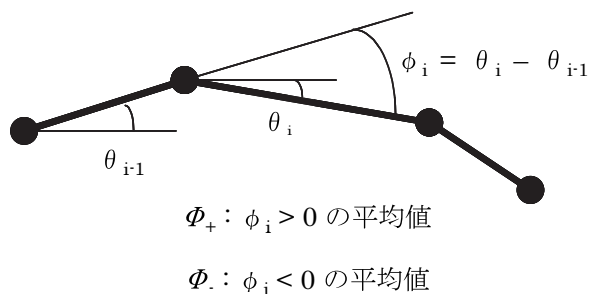


図 3.5 方向と曲直 Φ

3.3 特性値の評価

3.3.1 漢字文字列の選択基準

実験に用いる漢字文字列は、次の基準に基づいて選択する。

- (1) 文字数は 4 文字とする。
- (2) 文字列の方向は横書きとする。
- (3) 10 画以下の漢字を中心に選択する。
- (4) 横画，縦画，傾斜角（左払い，右払い）が文字列に含まれるように選択する。
- (5) 文字列は意味のあるものとする。

以下に、上記のように基準を設定した考え方を述べる。(1)については、利用者が ID とシステムが設定した共通のパスワードを入力する筆者照合システムを想定しているため、オンライン筆者照合で一般的に用いられている署名と同程度の 4 文字とした。(2)については、最近の文書や種々の記入欄が横書きであることから、横書きとした。(3)については、特性値が字画単位でも抽出され特性値の評価や選択を行うには全被験者で画数の統一が必要であることから、続け書きが起こりにくいと考えられる比較的画数の少ない文字を選択することとした。具体的には、常用漢字の約 52% が 10 画以下に含まれることから、10 画以下の漢字を中心とする。(4)については、漢字は横画，縦画，傾斜画（左払い，右払い）を中心に構成されるため、それらが文字列に含まれるように考慮した。(5)については、被験者の書きやすさから、文字列は意味のあるものとした。

3.3.2 評価に用いるデータ

「奈良永遠」という文字列を収集した。この文字列は上述の選択基準を満たすとともに、「奈良」は被験者にとって馴染みが深く、住所や所属の記載など比較的筆記機会が多いと考えられる文字である。「永」は永字八法といわれるように、漢字の基本的な運筆法を含んでいる。「遠」は 13 画であるが、「奈良永」に続く



図 3.6 収集したデータの例

意味のある文字であることから選択した。

被験者が、タブレット上に置いたデータ用紙の枠内 ($24\text{mm} \times 60\text{mm}$) に、「奈良永遠」という漢字を、コードレスの電子ペン (ボールペン芯) で筆記したものである。男性 35 人 (内 2 人は左利き)、女性 6 人、計 41 人の被験者から、1 度に 5 個、これを 1 週間以上の間隔をあけて各人 30 個以上のデータを収集した。評価には、この内から各被験者 30 個のデータを使用する。なお、使用したタブレットの仕様は、座標分解能 0.01mm 、傾き分解能 1.0° 、筆圧レベル 1024、サンプリング速度 200Hz である。収集したデータの例を図 3.6 に示す。

なお、字画に関連した特性値の評価を行う場合、一文字内での画数の統一が必要である。「奈」については 8 画で、「永」については 4 画で統一が可能であった。しかし、「良」と「遠」は字画間の繋がりが多く見られ、画数のばらつきが大きく統一が困難であった。そのため、「良」と「遠」の 2 文字に関しては、字画に関する特性値、始筆・送筆・終筆部に関する特性値を評価の対象から除外した。

3.3.3 評価方法

特性値の評価を行う場合、各特性値ごと (例えば「奈」1 画目の平均筆記速度など) に評価する方法と、同じ種類の特性値のまとまりごと (例えばすべての平均筆記速度など) に評価する方法が考えられる。以下、同じ種類の特性値のまとまりを特性値の種類と呼ぶことにする。特性値の種類ごとに評価することで、同じ種類の特性値をすべて用いたときの個人性の現れ方を評価でき、また文字や字画の形に依存せずに個人性を表す特性値を明らかにできる。

特性値ごとの評価

特性値ごとに筆者 41 水準の一元配置分散分析を行い、定義した特性値が個人間で差があるかどうかについて評価する。また、筆圧順位や速度順位など特性値の分布に正規性が仮定できないものが含まれるため、母集団の正規性を仮定せずに 3 つ以上の水準間の差の検定が行えるクラスカル・ウォリス検定 [43] も併せて行うこととした。

クラスカル・ウォリス検定とは、すべてのデータを順位付けしたとき、各水

準の順位平均に差があるかを検定するノンパラメトリックな手法である．そのための検定統計量 H は，各水準の順位平均 \bar{r}_i と全データ (N 個) の順位平均 $(N + 1)/2$ の差の二乗をとり，それに各水準のデータ数 n_i を掛けた値の和

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^a n_i \left(\bar{r}_i - \frac{N + 1}{2} \right)^2 \quad (3.1)$$

として定義される．ここで， a は水準の数である．したがって，この H の値が大きいくほど水準間の順位平均に差がある，すなわち筆者間で差がある特性値であるといえる．

特性値の種類ごとの評価

特性値の種類を評価するためには，複数の特性値の組み合わせに対応する必要がある．そこで，特性値ごとの分布ではなく，特性値空間での筆者内（本人間）距離と筆者間（他人間）距離の分布に基づいて特性値の種類を評価する．評価尺度はクラス間の分離性尺度である d -prime[44] を次に述べる観点から変形して用いた．筆者内距離の平均は筆者間距離の平均よりも小さいと仮定できることと，

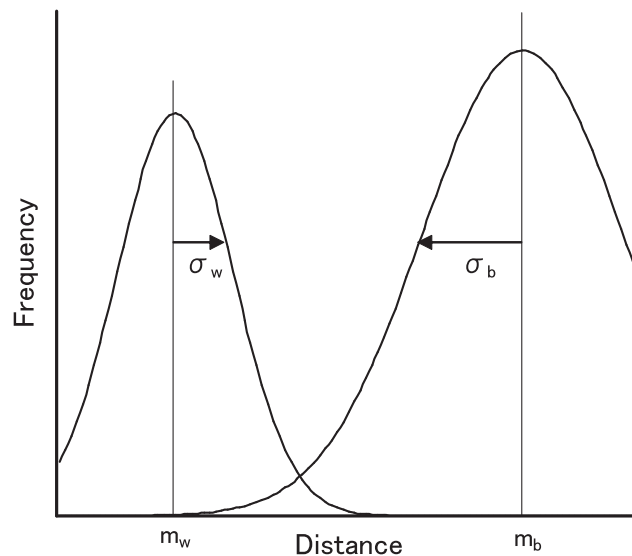


図 3.7 筆者内距離と筆者間距離の分布の概念図

距離の分布の対称性に保証がないので、実際に照合に影響を及ぼすのは筆者内距離の正の偏差と筆者間距離の負の偏差である。d-prime では、距離の分布の対称性を仮定して分散を用いているが、ここでは上述の考えにより、筆者内距離の正の偏差と筆者間距離の負の偏差の二乗平均を用いるという工夫を行った。以下に定義式を示す。

$$S = \frac{|m_w - m_b|}{\sqrt{\sigma_w^2 + \sigma_b^2}} \quad (3.2)$$

ここで m_w は特性値空間における筆者内距離の平均値、 m_b は筆者間距離の平均値、 σ_w^2 は m_w からの正の偏差の二乗平均、 σ_b^2 は m_b からの負の偏差の二乗平均である。なお、距離はユークリッド距離を用いた。図 3.7 に概念図を示す。

評価実験は、可能なすべての組合せ、筆者内 $41 \times C_{30}^2 = 17,835$ 組、筆者間 $30^2 \times C_{41}^2 = 738,000$ 組についてユークリッド距離を計算し、その平均 m_w, m_b と偏差の二乗平均 σ_w^2, σ_b^2 を求めることにより S 値を算出した。

3.3.4 評価結果

特性値ごとの評価結果

分散分析とクラスカル・ウォリス検定の結果、全特性値 563 個の内、「奈」1 画目、2 画目の送筆部の筆圧順位を除く 561 個が有意水準 1% で有意であると判定され、定義した特性値が程度の違いはあれ個人差を表すことが明らかになった。表 3.2 に分散分析の検定統計量 F とクラスカル・ウォリスの検定統計量 H の特性値ごとの平均値を示す。F 値が大きいことは、個人内の変動に比べて個人間の変動が大きいことを、また、H 値が大きいことは、各個人の順位平均の差が大きいことを意味する。なお、有意水準 1% の F 値は $F_{0.01}(40, 1189) = 1.609$ 、H 値は $\chi_{40}^2(0.01) = 63.691$ であり、各特性値の F 値、H 値の平均値がこれらに比べて高い値を持つことがわかる。

平均ペン方位の F 値が非常に大きいのは、左手書きの被験者がいるためである。図 3.8 に文字列の平均ペン方位の箱ひげ図を示す。筆者 g と u が左手書きの筆者であり、筆者 u が他の被験者から大きく離れていることが分かる。また、これら二人の左手書きの筆者を除いた分布を見ても、平均ペン方位の個人内の分散が小さくかつ個人間で差があることが確認できる。そこで、この二

表 3.2 分散分析とクラスカル・ウォリス検定の結果

上段:F値, 下段:H値

特性値	文字列	文字	字画	始・送・終筆部			
				始筆	送筆	終筆	
動的	筆記時間	120.40	68.39	-	-	-	-
		989.90	914.32	-	-	-	-
	ペンダウン時間	151.57	102.79	86.50	-	-	-
		1019.68	950.59	964.24	-	-	-
	ペンアップ時間	92.54	47.68	-	-	-	-
		945.76	881.93	-	-	-	-
	筆記タイミング	-	-	52.58	-	-	-
		-	-	823.61	-	-	-
	筆圧順位	-	-	36.90	20.00	12.92	20.51
		-	-	653.78	483.19	292.38	483.55
	平均筆圧	83.53	79.47	79.02	54.73	73.47	47.89
		910.06	886.90	877.22	772.39	866.03	750.77
	平均ペン高度	271.68	249.06	230.64	193.37	231.63	229.78
		1060.78	1044.62	1022.18	992.47	1020.90	1023.17
	平均ペン方位	3176.15	2822.58	2551.69	2215.11	2523.39	2494.72
		1185.43	1180.34	1176.24	1167.03	1175.31	1176.70
	速度順位	-	-	18.25	10.93	11.50	17.19
		-	-	441.73	321.69	335.09	398.11
平均筆記速度	234.13	150.89	62.67	12.23	72.79	50.44	
	1083.81	1027.12	778.40	335.62	841.65	708.42	
静的	大きさ	187.92	139.24	-	-	-	-
		1042.67	1018.11	-	-	-	-
	縦横比	89.36	61.40	-	-	-	-
		925.47	764.17	-	-	-	-
	中心位置 x	102.22	93.44	-	-	-	-
		843.77	847.83	-	-	-	-
	中心位置 y	32.84	33.85	-	-	-	-
		697.49	687.21	-	-	-	-
	字間隔	58.74	-	-	-	-	-
		798.06	-	-	-	-	-
	相対始点位置 x	-	-	61.84	-	-	-
		-	-	771.36	-	-	-
	相対始点位置 y	-	-	72.78	-	-	-
		-	-	773.57	-	-	-
	長さ	172.07	131.62	96.68	-	-	-
		1034.73	1000.50	880.77	-	-	-
	曲直 +	138.48	49.66	14.42	8.05	13.27	12.19
		994.54	726.95	385.84	259.16	426.39	403.08
曲直 -	113.65	45.27	14.79	8.55	13.86	15.06	
	993.39	728.10	430.84	272.09	432.15	459.83	
方向	-	-	30.48	16.04	45.71	27.81	
	-	-	568.17	435.52	715.83	531.93	

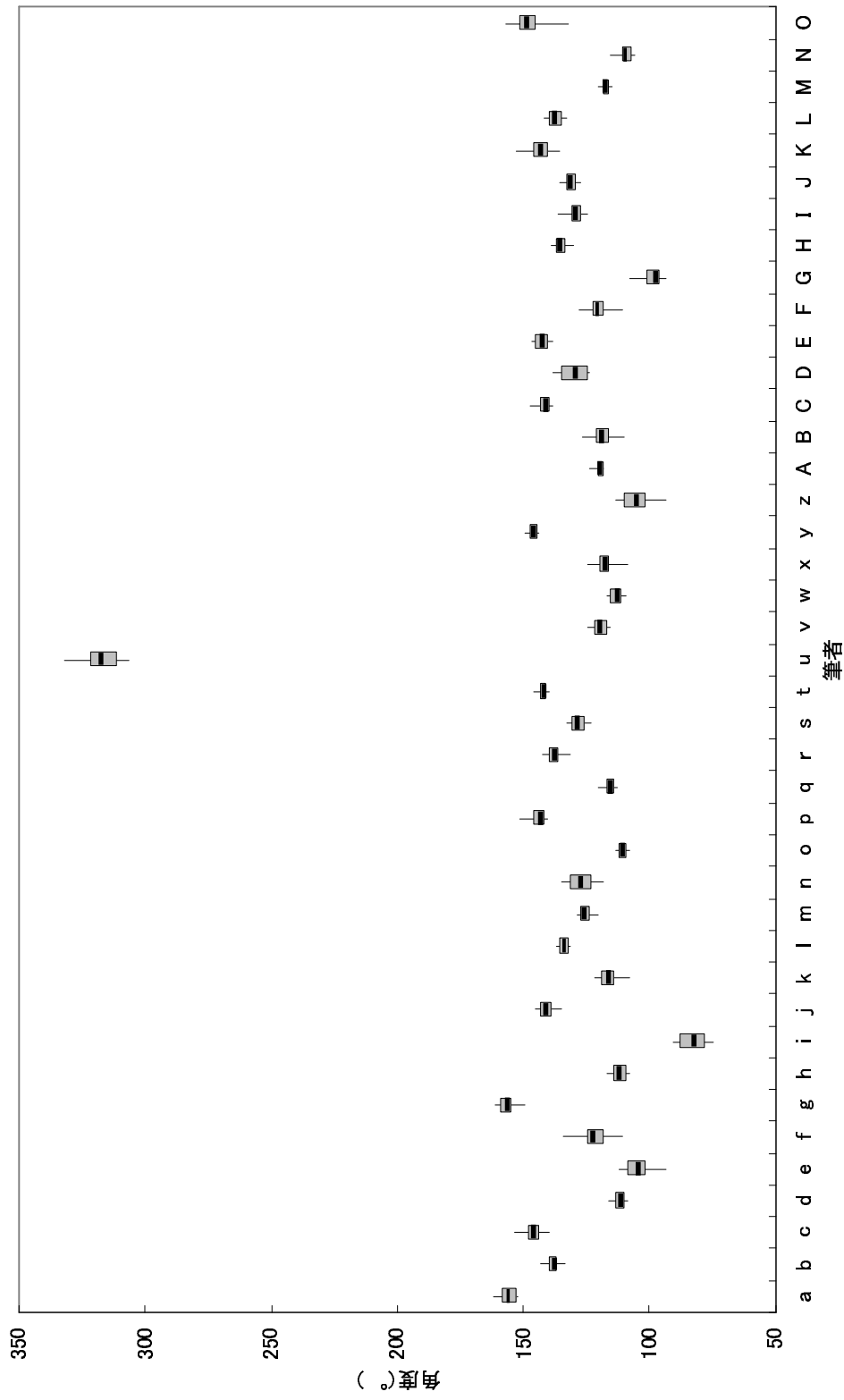


図 3.8 文字列の平均ペン方位の分布 (箱ひげ図)

人を除いた 39 人の右手書き被験者に対して分散分析とクラスカル・ウォリス検定を行った。その結果，F 値は $766.035 > F_{0.01}(38, 1131) = 1.627$ ，H 値は $1121.168 > \chi_{38}^2(0.01) = 61.162$ と高い値を示した。このことより，ペン方位の平均値は右手書きと左手書きで大きな差が出るだけでなく，右手書き被験者の間でも個人間で差が出る特性値であるといえる。

平均ペン高度と方位に関しては，文字列，文字，字画，始筆・送筆・終筆部の全体を通じて高い F 値と H 値を示している。一方，字画と始筆・送筆・終筆部の曲直，筆圧順位，速度順位，送筆部を除いた方向，始筆部平均筆記速度の F 値と H 値が低い値であった。動的特性値と静的特性値を比べると，動的特性値の方が高い F 値と H 値を示す傾向にあるといえる。

また，平均筆圧，平均筆記速度，曲直については，文字列や文字に比べ，字画や始筆・送筆・終筆部の F 値，H 値が低くなっている。これは，字画によって個人性の現れ方が違うこと，また，始筆，送筆，終筆など局所的な部分での計測であり，文字列や文字での計測よりも個人内での変動が大きくなることが原因と考えられる。

なお，F 値と H 値は同じ傾向を示し特に分析方法による大きな違いは見られなかった。しかし，平均ペン方位のように特異な値を持つ筆者が含まれる場合 F 値は非常に高くなる。このことと，特性値の分布が必ずしも正規分布ではないことを考え合わせると，クラスカル・ウォリス検定により特性値を評価する方が良いと考えられる。以下では H 値を中心に検討を行う。

表 3.3 に H 値の上位 200 に含まれる特性値の数を各部分ごとに示す。なお，平均ペン方位と平均ペン高度については部分にかかわらず，すべての特性値が上

表 3.3 H 値上位 200 個に含まれる特性値の数

文字列 (17)	文字 (56)	字画 (166)	始筆部 (108)	送筆部 (108)	終筆部 (108)
12	34	63	25	40	26
(10)	(26)	(39)	(1)	(16)	(2)

上段の () 内は特性値の総数，

下段の () 内の数字はペンの平均方位と高度を除いた数

位 200 までに含まれていた．そのことを踏まえて，平均ペン方位と高度を除いた個数についても括弧書きで示した．この結果から，各部分で抽出した特性値の総数に対する上位 200 までに含まれる特性値の数の割合は文字列，文字，字画，始筆・送筆・終筆部の順に高く，局所的な特性値より大局的な特性値の方がより個人性を表しやすいといえる．このことは，文字列と文字に関する特性値が下位 200 に含まれないことから裏付けられる．しかしながら，「永」4 画目終筆部の平均筆記速度（H 値=1022.92，順位 87 位）や「奈」1 画目送筆の方向（H 値=1012.33，順位 100 位）など局所的な特性値においても高い H 値を示すものもある．一方，筆圧順位，速度順位，字間隔については上位 200 に一つも含まれなかった．

図 3.9 に「永」4 画目終筆部の平均筆記速度の箱ひげ図を，図 3.10 に「奈」1 画目送筆部の方向の箱ひげ図を示す．これらの特性値は H 値が 1000 以上と高い値を持つが，図からもわかるように，特定の個人間を区別することは可能であるが，これら特性値だけで全筆者間を区別できるわけではない．したがって，定義した特性値を総合的に用いて真偽を判定する必要があるといえる．また，筆者によって個人内変動の大きさが異なり，安定して筆記する人とそうでない人がいることがわかる．

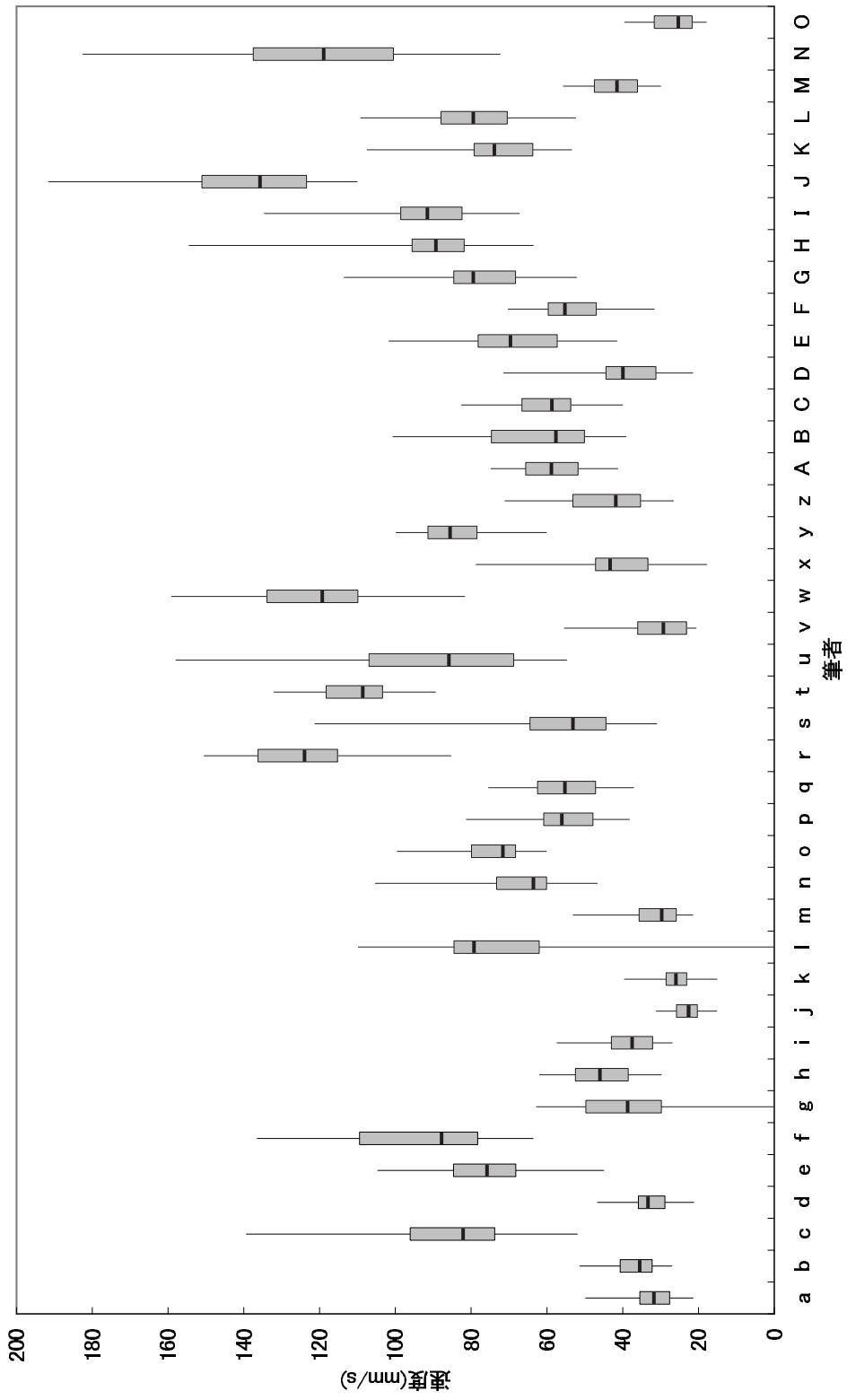


図 3.9 「永」4画目終筆部の平均筆記速度 (箱ひげ図)

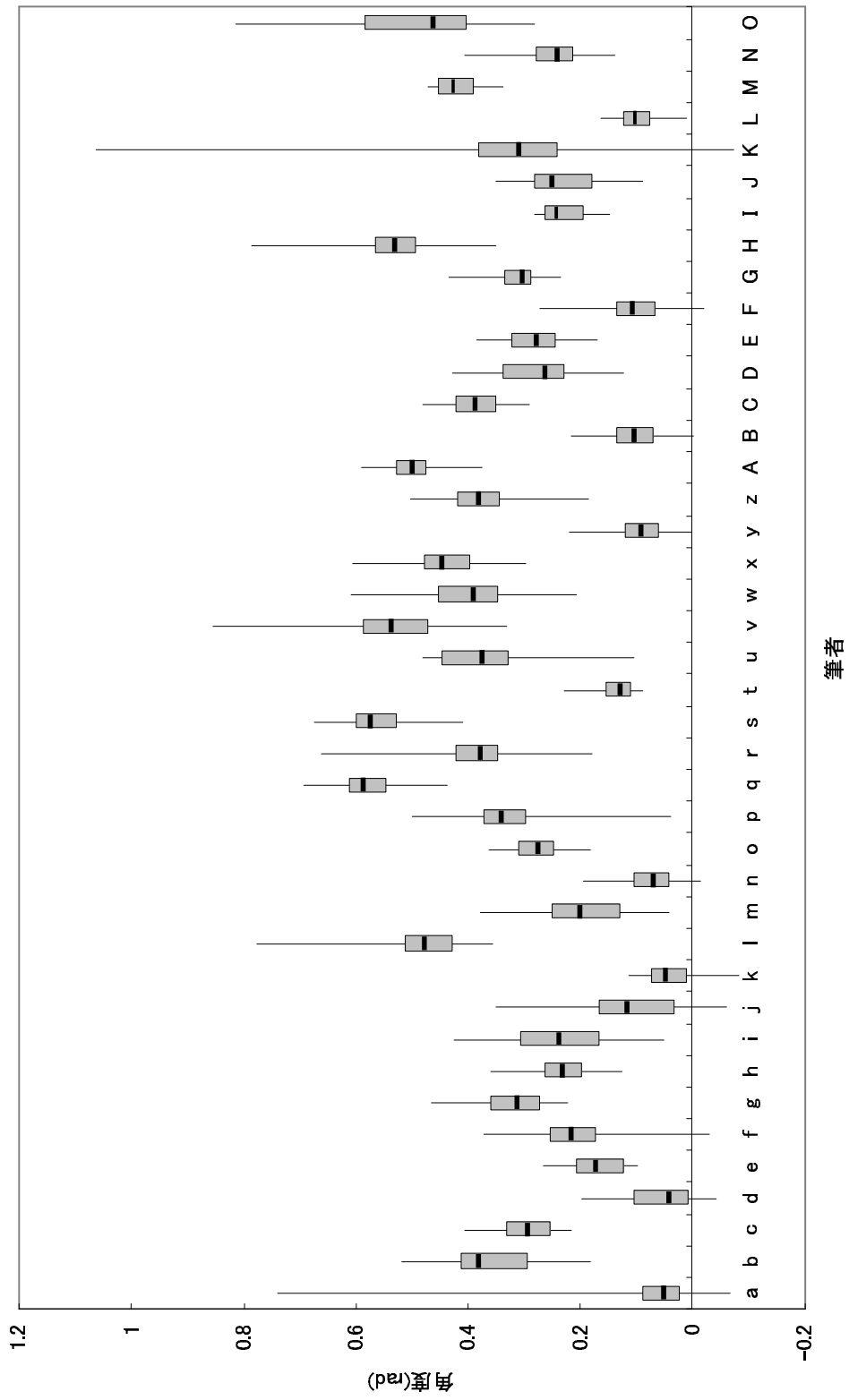


図 3.10 「奈」1 画目送筆部の方向 (箱ひげ図)

特性値の種類ごとの評価結果

表 3.4 に評価値 S の値を示す .

表 3.4 評価値 S

特性値	文字列	文字	字画	始・送・終筆部			全	
				始筆	送筆	終筆		
動的	筆記時間	0.692	0.630	-	-	-	-	
	ペンダウン時間	0.802	0.889	0.808	-	-	-	
	ペンアップ時間	0.585	0.467	-	-	-	-	
	筆記タイミング	-	-	0.592	-	-	-	
	筆圧順位	-	-	0.934	0.704	0.451	0.669	1.071
	平均筆圧	0.625	0.692	0.985	0.894	0.963	0.889	1.092
	平均ペン高度	0.949	0.985	0.999	1.007	0.992	0.993	1.005
	平均ペン方位	1.381	1.385	1.393	1.387	1.393	1.401	1.395
	速度順位	-	-	0.618	0.510	0.499	0.649	0.795
平均筆記速度	0.980	1.076	1.062	0.331	1.231	0.965	1.013	
静的	大きさ	0.910	0.991	-	-	-	-	
	縦横比	0.659	0.857	-	-	-	-	
	中心位置	0.717	0.896	-	-	-	-	
	字間隔	0.741	-	-	-	-	-	
	相対始点位置	-	-	1.264	-	-	-	
	長さ	0.882	1.011	1.302	-	-	-	
	曲直	0.842	0.790	0.363	0.244	0.293	0.373	0.447
方向	-	-	0.712	0.404	0.617	0.774	0.758	

表中、「-」は該当する特性値が存在しないことを、また「全」は同じ種類のすべての特性値を用いたときの S の値を示す。ただし、筆記時間など部分の違いによりスケールが異なる特性値の種類については、すべての特性値を用いた S 値は計算していない。

表 3.4 から以下のことがいえる。

- (1) 相対始点位置と字画の長さが高い S 値を示した。
- (2) 平均ペン方位がすべての部分で最も高い S 値を示した。しかし、部分間の差は無く、かつすべての平均ペン方位を用いても S 値にほとんど変化は無かった。平均ペン高度についても同じ傾向が見られた。
- (3) 筆圧順位は字画では平均筆圧と同等の S 値を示し、すべての筆圧順位を用

いることでより S 値が高くなった。

- (4) 平均筆記速度はすべての部分で速度順位よりも高い S 値を示した。
- (5) 曲直は字画，始筆・送筆・終筆部で最も S 値が低かった。
- (6) 始筆部の平均筆記速度と方向は他の部分に比べて S 値が低かった。なお，始筆部を除いた特性値をすべて用いたときの S 値は，平均筆記速度で 1.31 に，方向で 0.878 になった。

前述の特性値ごとの評価と比較すると，筆圧順位，速度順位，方向については特性値ごとの評価は低い特性値を組み合わせることにより他の特性値と同程度の筆者間の差を表すことが示された。また，字画，始筆・送筆・終筆部の曲直については，組み合わせることによる効果は現れず，特性値ごと，特性値の種類ともに低い評価となった。

特性値ごと，特性値の種類の評価結果より，筆跡鑑定の知見に基づき定義した特性値は程度の差はあるにしろ，個人性を表すことが明らかになった。とくに，S 値が大きく個人性をよく表すと考えられる特性値の種類は，平均ペン方位と高度，平均筆圧，筆圧順位，平均筆記速度，相対始点位置，長さであり，逆に S 値が小さく個人性が現れにくいと考えられる特性値の種類は字画と始筆・送筆・終筆部の曲直と，始筆部の方向と平均筆記速度である。

3.4 特性値の筆者識別力の評価

3.4.1 実験方法

実験に用いるデータは評価に用いたものと同じである。各被験者の 30 個のデータの内，10 個を参照筆跡，残りを未知筆跡とした。なお，参照筆跡の採り方を，最初の 10 個，次の 10 個，最後の 10 個と変え交差法による実験を行った。

識別，照合に用いる筆者 i に対する距離を，

$$D_i = \sqrt{\sum_{j=1}^n \frac{(f_j - m_{ij})^2}{\sigma_{ij}^2}} \quad (3.3)$$

で定義する。ここで， n は特性値の数， f_j は未知筆跡の j 番目の特性値， m_{ij} ，

σ_{ij}^2 は参照筆跡より求めた筆者 i の j 番目の特性値の平均と分散である。識別実験では、未知筆跡と各筆者間の距離を求め、最小の距離を与える筆者を未知筆跡の筆者と判断する。照合実験では、求めた距離 D_i がしきい値 T に対して、 $D_i < T$ ならば本物、そうでなければ偽物と判定する。なお、しきい値は全筆者共通の値とし、しきい値を 0.05 きざみで変化させたとき、本物を偽物と誤る FRR と偽物を本物と誤る FAR がほぼ等しくなる値とした。

3.4.2 実験結果

表 3.5 に実験結果を示す。実験には、全特性値 563 個と評価結果から選択した 270 個の特性値を用いた。なお、特性値の選択に関する考え方は以下のとおりである。

- (1) 評価結果より H 値, S 値ともに低い字画と始筆・送筆・終筆部の曲直, 始筆部の平均筆記速度と方向を除く。
- (2) 平均ペン方位と高度については, すべての部分の特性値を組み合わせても S 値にほとんど変化が無いことから, 文字列, 文字, 字画のみを用いる。
- (3) 筆圧に関しては, 筆圧順位はすべての特性値を組み合わせることで平均筆圧と同等の S 値が得られることと, 筆圧情報は用いる電子ペンの経時変化等により変動が大きいことが経験的に分かっていることから, 筆圧順位を選択する。
- (4) 筆記速度については, S 値の比較から平均筆記速度を選択する。

表 3.5 照合実験と識別実験の結果

使用した特性値	FRR	FAR	識別率
全特性値	6.22%	6.37%	90.37%
選択した特性値	3.21%	3.17%	96.54%

全特性値を用いた場合の筆者識別率は 90.37% , FRR 6.22% , FAR 6.37% であり, 定義した特性値が筆者識別力を持ち, これら特性値を用いた筆者識別・

照合の可能性が示された。さらに、特性値を選択することで識別率が 96.54%、FRR が 3.21%、FAR が 3.17% に改善され、特性値をうまく選択することでより筆者識別力が向上することも併せて示された。比較のため、全特性値を用いた場合と選択した特性値を用いた場合の ROC(Receiver Operating Curve) を図 3.11 に示す。図からも明らかなように、特性値を選択することで照合性能が向上していることがわかる。

なお、筆圧順位の代わりに平均筆圧を用いた場合、FRR 4.43%、FRR 4.56% であり筆圧順位を用いるほうが高い照合率が得られ、筆圧順位を用いる有効性が示された。ちなみに、筆圧順位を単独で用いた場合は、FRR 13.13%、FAR 14.78% であり、平均筆圧単独では、FRR 22.52%、FAR 23.39% であった。このことから、平均筆圧よりも筆圧順位の方が筆者識別力が高いと考えられる。

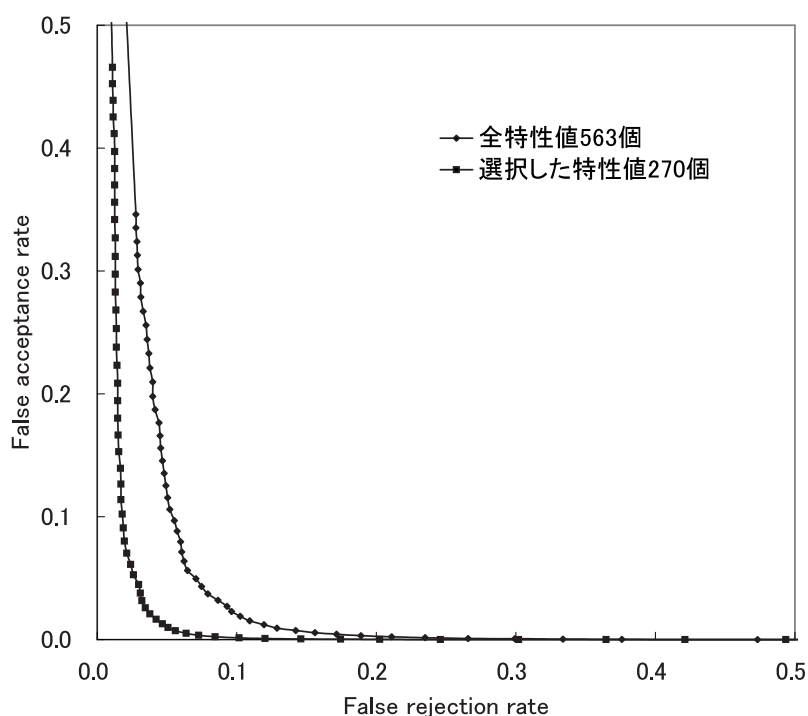


図 3.11 全特性値と選択した特性値を用いた場合の ROC

つぎに、動的特性値と静的特性値および各部分の違いによる筆者識別力について検討するために、照合実験を行った。その結果を表 3.6 に示す。表中「動的選択」など「選択」と記したのは選択した 270 個の特性値のみを用いた結果であり、例えば「動的選択 (160)」は選択した 270 個の特性値の内、動的特性値が 160 個含まれ、それら 160 個のみを用いたことを意味する。

表 3.6 部分ごとの照合実験結果

使用した特性値(個数)	FRR	FAR	EER	使用した特性値(個数)	FRR	FAR	EER
全特性値(563)	6.22%	6.37%	6.30%	選択(270)	3.21%	3.17%	3.19%
動的(345)	4.96%	5.04%	5.00%	動的選択(160)	3.58%	3.73%	3.65%
静的(218)	12.97%	14.10%	13.53%	静的選択(110)	5.12%	4.55%	4.84%
文字列(17)	5.45%	5.08%	5.27%	文字列選択(16)	5.33%	5.05%	5.19%
文字(56)	5.08%	4.68%	4.88%	文字選択(52)	4.59%	5.09%	4.84%
字画(166)	3.78%	3.53%	3.66%	字画選択(118)	3.62%	3.19%	3.41%
始送終(324)	7.85%	8.12%	7.98%	始送終選択(84)	6.79%	7.04%	6.91%
文字列文字(73)	4.67%	4.95%	4.81%	文字列文字選択(68)	4.84%	4.72%	4.78%
文字列文字字画(239)	3.74%	3.35%	3.54%	文字列文字字画選択(186)	3.46%	3.89%	3.67%

実験結果について検討を行う。部分ごとにすべての特性値を使用した場合と選択した特性値を使用した場合を比較すると、選択した特性値を用いた方が静的特性値で約 9%、動的特性値と始筆・送筆・終筆部の特性値で約 1% FRR と FAR が等しくなる等誤り率 EER(Equal Error Rate) が改善されており、他の部分ではほぼ同程度の EER を示した。このことから、特性値選択が各部においても有効であることが確認できた。なお、ここでの EER は FRR と FAR の平均を用いている。

この結果を踏まえて、以下では選択した特性値による各部の筆者識別力について検討する。まず、動的特性値と静的特性値との比較であるが、動的特性値の EER 3.65%、静的特性値の EER 4.84% と動的特性値の方が約 1.2% 低くなっている。さらに比較するため ROC を図 3.12 に示す。図からも明らかなように、動的特性値の方が良い照合性能を示した。動的特性値が既に書かれてある筆跡からは盗むことが困難であることを考えると、筆者照合にとっては有効な特性値であるといえる。しかし、静的特性値を動的特性値と組み合わせることで EER が

さらに 0.5% 程度改善されることから，筆者照合においては動的特性値と静的特性値を総合的に用いる必要性があることが確認できる．

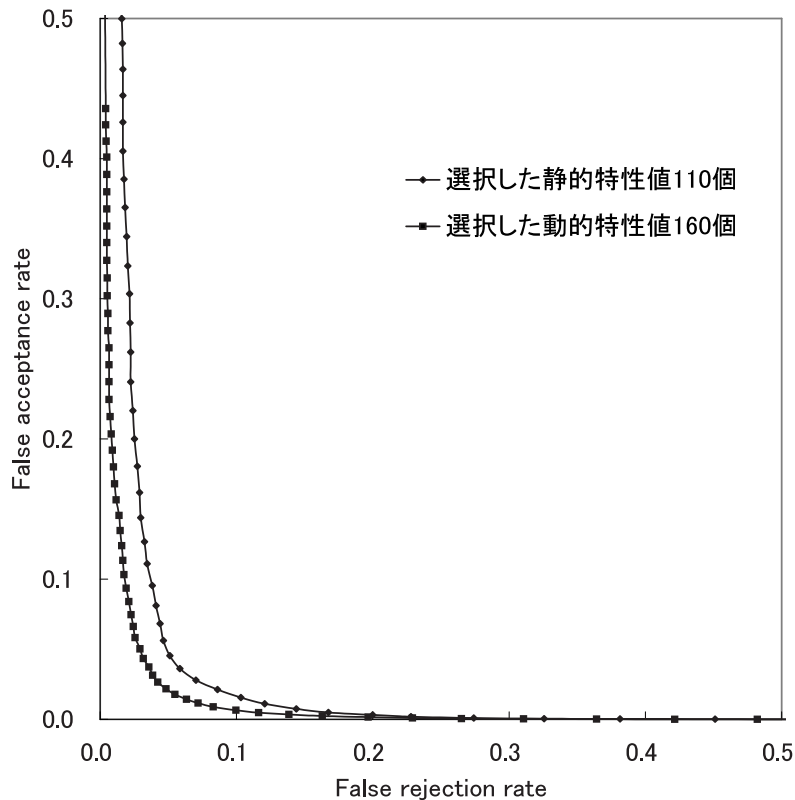


図 3.12 選択した動的特性値と静的特性値を用いた場合の ROC

つぎに，部分ごとについて比較すると，選択した字画の特性値を用いた場合が EER が一番低く約 3.4% であり全特性値から選択した 270 個を使用した場合より約 0.2% 程度高いという結果であった．これら二つを比較するために ROC を求め図 3.13 に示す．図より全特性値から選択した 270 個の特性を用いた方が FRR を小さくしたときの FAR が低くなり，良い照合性能を示すことがわかる．

選択した文字列の特性値を用いた場合の EER は約 5.2% ，文字の特性値は約 4.8% であった．ただ，文字列や文字単位に抽出した特性値，言い換えれば文字の内部構造に依存しない特性値だけを用いても 95% 程度の筆者照合率が得られ

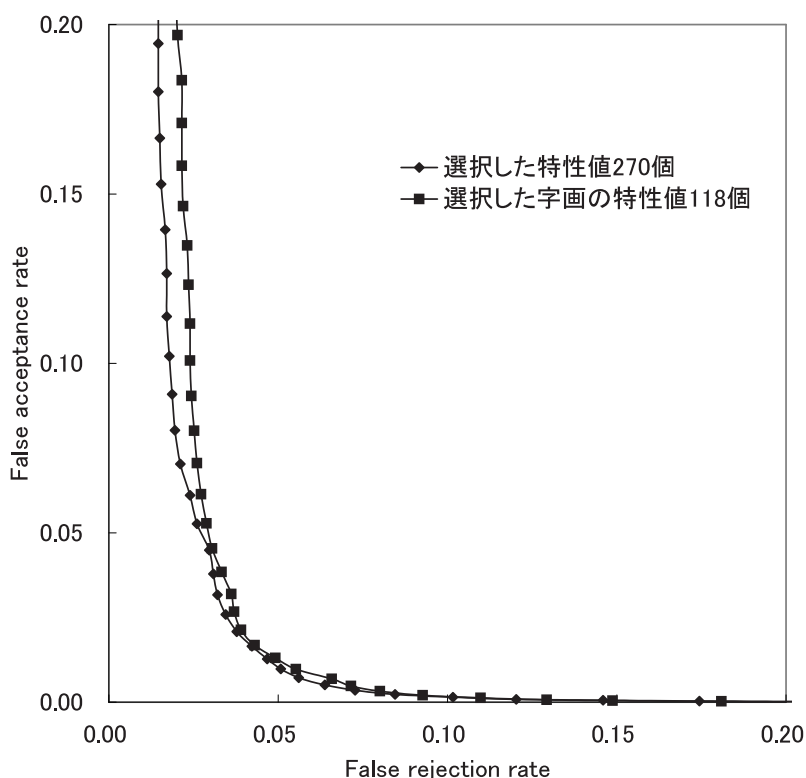


図 3.13 選択した字面の特性値と全特性値から選択した特性値を用いた場合の ROC

ることが明らかになった．特に，文字列の場合は 16 個の特性値のみを用いた結果であり興味深い．選択した始筆・送筆・終筆部の特性値による EER は約 6.9% と最も悪い結果であった．これは，特性値の評価のところでも述べたが，始筆・送筆・終筆部の特性値は局所的な特性値であり個人内変動が文字列や文字に比べ大きくなるためであると考えられる．

最後に，部分の組合せについて検討する．選択した文字列と文字の特性値を組み合わせた場合，それに選択した字面の特性値を加えた場合とさらに選択した始筆・送筆・終筆部の特性値を加えた場合の 3 通りについて ROC を求め図 3.14 に示す．図からも明らかなように，各部の特性値を組み合わせることで照合性能が改善されているのがわかる．すなわち，文字列，文字，字画，始筆・送筆・終筆部という筆跡の各部分において特性値を抽出し，それらを総合的に用いること

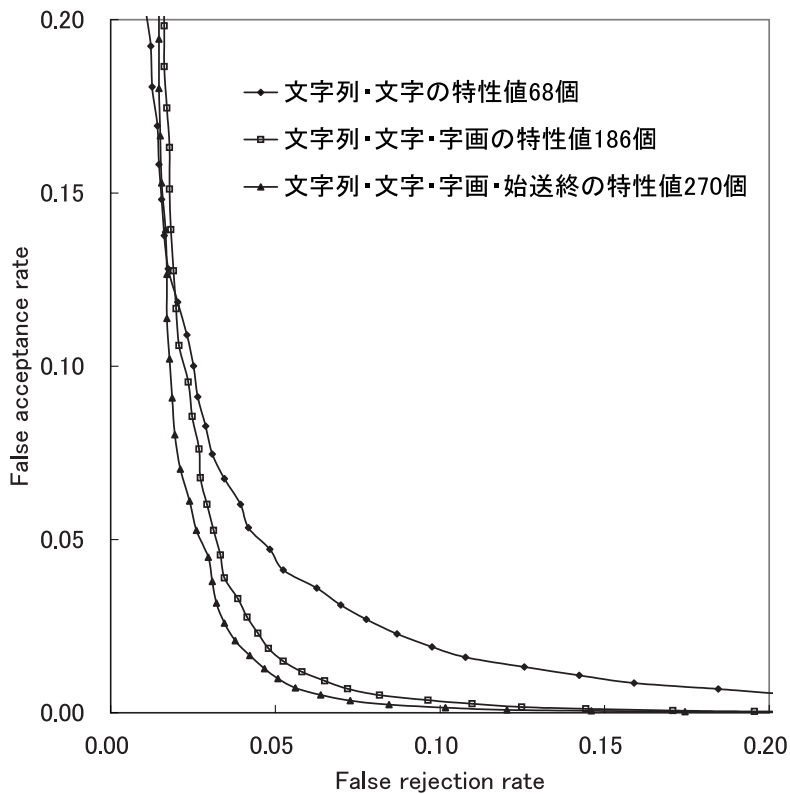


図 3.14 各部の特性値を組み合わせていった場合の ROC

が筆者照合に有効であることが明らかになり，本研究のアプローチの正しさが示されたといえる．

3.5 結言

タブレット上の一定枠内に手書きされた漢字文字列に現れる個人性を，筆跡鑑定の知見に基づく特性値により抽出することを提案し，分散分析とクラスカル・ウォリス検定により特性値ごとの評価実験を行った．また，同じ種類の特性値を組み合わせた場合の評価を特性値空間での筆者内（本人間）距離と筆者間（他人間）距離の分布に基づいて行った．

その結果，定義した特性値は程度の差はあるにしろ，個人性を表すことが明らか

かになった。特に個人性をよく表すと考えられる特性値の種類は、平均ペン方位と高度、平均筆圧、筆圧順位、平均筆記速度、相対始点位置、長さであり、逆に個人性が現れにくいと考えられる特性値の種類は字画と始筆・送筆・終筆部の曲直、始筆部の方向と平均筆記速度である。また、新たな特性値である筆圧順位は、従来の平均筆圧よりも個人識別力が高いことも明らかにした。

さらに、定義した特性値を用いて筆者識別・照合実験を行い、その有効性を確認できた。特に、同じ種類の特性値の評価結果を基に特性値を選択することで、全特性値を用いるよりも照合率が改善されること、文字の構造に依存しない文字列や文字の特性値のみを用いても約 95% 程度の照合率が得られること、筆跡の各部分における静的・動的特性値を総合的に用いることが筆者照合に有効であることを明らかにした。

第 4 章

オンライン筆者照合手法の提案

4.1 緒言

第 3 章において、タブレットよりオンライン入力された漢字文字列に対して筆跡鑑定の知見に基づき特性値を抽出し、それら特性値が個人性を表すとともに筆者識別能力を持つことを明らかにした。この章では、抽出した特性値を用いたオンライン筆者照合手法を提案する。

筆跡鑑定の知見に基づいて抽出した特性値は文字列、文字、字画および始筆・送筆・終筆部単位で定義され多数であるが、それら特性値は程度の差はあるが個人性を表しており、それらを総合的に評価し筆者照合を行う方法を考える必要がある。さらに、筆者照合システムの構築にあたっては限られた数の参照筆跡（学習サンプル）、例えば本人から 1～10 個程度、を用いて特性値の選択と識別器の設計を行う必要がある。

本章では、まず 4.2 節で想定する筆者照合システムについて述べる。つぎに、4.3 節において特性値選択法を示す。想定している筆者照合システムは、漢字数文字で構成された共通のパスワードを入力するものであり、利用者から共通の文字列が参照筆跡として事前に複数収集可能である。このことを利用して、特性値の選択に関しては、特性値の種類ごとに本人の筆跡間の距離と他人間の距離を計算し、その分布に基づいて特性値を選択する。4.4 節では、真正筆跡の各特性値は参照筆跡の平均値に近いものが多く、遠いものは少ないという考えに基づき、

各特性値の参照筆跡の平均値からの偏差の度数分布を基にした識別器を提案する．4.5 節では，提案した識別器の有効性を筆者照合実験を行うことにより確かめる．あわせて，特徴パラメータに基づくオンライン署名照合において Lee ら [7] が用いた多数決法および，一般的に用いられるユークリッド距離との比較を行い，両手法より照合精度が高いことを示す．最後に，4.6 節でこの章のまとめを行う．

4.2 想定する筆者照合システム

本論文では，第 1 章で述べたように利便性を重視したアクセスコントロールやトラッキングを行うオンライン筆者照合システムを想定している．利用者が ID を入力した後，システムが設定した漢字数文字で構成された共通のパスワードをタブレットから入力することで本人を確認する．なお，タブレットは市販のものをを用い，パスワードはタブレット上の一定枠内に横書きされる．

筆者照合の難しさを，単純に筆跡の個人内変動の大きさに依存すると考えると，筆者共通のパスワードは各筆者が定めたパスワードや署名に比べ個人内変動が大きいと考えられ，想定する筆者照合システム構築は署名照合より困難な課題であるといえる．そこで，目標とする性能は既に提案されているオンライン署名照合の性能を上回るものを目指すのではなく，同程度の性能を有するものとする．しかしながら，想定するシステムではパスワードを定期的に更新でき，さらに偽造者はパスワードとその筆跡の特徴を盗む必要があることから，署名照合より偽造に対して安全なシステムといえる．

性能評価の一つの指針として，本人を他人と誤る本人拒否率 FRR と，他人を本人と誤る他人受理率 FAR が等しくなる，等誤り率 EER がある．過去の日本語署名を対象としたオンライン署名照合の研究論文の EER が 1~2% であることから，想定するシステムの EER を 2% 程度とする．また，利用者の利便性を重視するという立場から，FRR を 0% に設定したときに，FAR が 10% 以下となることを目標とする．この値は，POS(Point-Of-Sale) アプリケーションでのガイドラインとして，Nelson ら [5] が用いた FRR 0.5% で FAR 20% 以下，Lee ら [7] が用いた FRR 0% で FAR 25% 以下を参考に，POS が対面で本人を確認

することを考慮し POS のガイドラインよりも条件を厳しくした。

4.3 特性値の選択方法

4.3.1 筆者内距離と筆者間距離の分布に基づく特性値選択法

特性値選択の考え方として、

- 全筆者共通の特性値の種類を選択する
- 筆者ごとに特性値を選択する

という二つがあげられる。また、特性値を圧縮するという観点から主成分分析等の適用も考えられる。しかし、抽出した特性値の数が多数であるにもかかわらず参照筆跡の数が少数であることから、主成分分析等の適用は不適であると判断した。

本研究では、全筆者共通の特性値の種類を選択するという考えに立つ。これにより、全筆者共通で筆者照合に有効であると考えられる特性値の種類が明らかになり、特定の筆者や特定の字画等に依存しにくい特性値が選択できる。

特性値の種類を選択するためには、複数の特性値の組み合わせに対応する必要がある。そこで、特性値ごとの分布ではなく、特性値空間での筆者内（本人間）距離と筆者間（他人間）距離の分布に基づいて特性値の種類を選択する。特性値選択のための評価尺度は第 3 章において特性値の評価に用いた S 値を使用する。

$$S = \frac{|m_w - m_b|}{\sqrt{\sigma_w^2 + \sigma_b^2}} \quad (4.1)$$

ここで m_w は特性値空間における筆者内距離の平均値、 m_b は筆者間距離の平均値、 σ_w^2 は m_w からの正の偏差の二乗平均、 σ_b^2 は m_b からの負の偏差の二乗平均である。

ただし、第 3 章で特性値の種類の評価を行ったときは、収集した全データ（41 人 × 30 個 = 1,230 個）に対してすべての組合せについて本人間距離と他人間距離を計算し S 値を求めた。しかし、ここでは筆者照合手法を検討する立場から実際のシステムに近い形での特徴選択を考える必要がある。

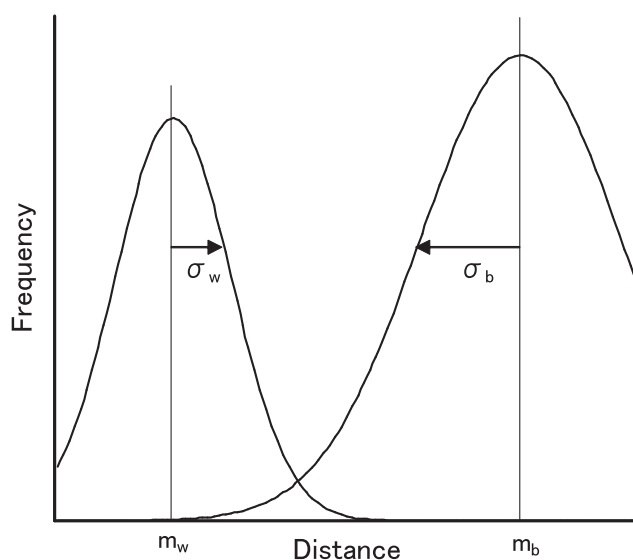


図 4.1 筆者内距離と筆者間距離の分布の概念図

そこで、筆者ごとに、参照筆跡のすべての組み合わせについて各特性値の種類に対する筆者内距離と筆者間距離を求め評価値 S を計算し、これを全筆者で平均する。この S の平均値に基づいて、以下の選択基準により、特性値を選択する。

- (1) 文字列、文字、字画の特性値に関しては、 S 値が小さい特性値を排除する。ただし、 S 値のしきい値は 0.707 とする。この値は、 σ_w と σ_b が等しいと仮定したとき、 m_w と m_b の差が σ_w と等しくなる値である。一般に $\sigma_w < \sigma_b$ と考えられるので、 S が 0.707 であるということは m_w と m_b の差が σ_w 以上であることを意味する（図 4.1 参照）。
- (2) 始筆・送筆・終筆部の特性値に関しては、(1) で選択した特性値と組み合わせることにより、さらに S 値が向上するものを選択する。これは、特性値の数が全字画数 $\times 3$ 個と非常に多くなること、また字画に関する特性値と同じ種類の特性値であることによる。
- (3) 筆圧と筆記速度に関する特性値は平均値と順位の 2 種類あるが、これらについては、どちらか一方を選択することとする。速度については、 S 値が大きい方を選択する。筆圧情報は用いる電子ペンの経時変化などにより

変動が大きいことが経験的に分かっているため，S 値に大きな差がない限り，筆圧情報の変動の影響を受けにくい筆圧順位を優先して選択する．

4.3.2 選択結果

使用する漢字文字列データは，第 3 章で用いた「奈良永遠」である．各筆者 30 個のデータの内最初の 10 個を参照筆跡とし，本人間のすべての筆跡のペア $10 \times 9 \div 2 = 45$ 組と他人間のペア $10 \times 40 \times 10 = 4,000$ 組について，特性値の種類ごとにユークリッド距離を計算し評価値 S を求め，全筆者 41 人の平均をとった．表 4.1 に S の値を示す．なお，表中「全」とあるのは，始筆・送筆・終筆部の特性値と他の部分の同じ特性値を組み合わせた場合の S の値である．

なお，全筆者共通の特性値を選択するため，一文字内での画数の統一が必要である．第 3 章の特性値の評価で述べたように，「奈」については 8 画で，「永」に

表 4.1 評価値 S

特性値	文字列	文字	字画	始・送・終筆部			全	
				始筆	送筆	終筆		
動的	筆記時間	0.934	0.986	-	-	-	-	
	ペンダウン時間	1.028	1.169	1.529	-	-	-	
	ペンアップ時間	0.891	0.960	-	-	-	-	
	筆記タイミング	-	-	1.134	-	-	-	
	筆圧順位	-	-	1.165	0.914	0.574	0.843	1.347
	平均筆圧	0.943	1.049	1.427	1.156	1.395	1.354	1.591
	平均ペン高度	1.206	1.269	1.290	1.306	1.280	1.292	1.304
	平均ペン方位	1.683	1.685	1.685	1.678	1.683	1.692	1.690
	速度順位	-	-	0.773	0.516	0.576	0.763	1.029
	平均筆記速度	1.096	1.213	1.304	0.388	1.479	1.273	1.640
静的	大きさ	1.031	1.216	-	-	-	-	
	縦横比	0.772	0.978	-	-	-	-	
	中心位置	0.941	1.081	-	-	-	-	
	字間隔	0.928	-	-	-	-	-	
	相対始点位置	-	-	2.022	-	-	-	
	長さ	1.062	1.197	1.670	-	-	-	
	曲直	0.933	0.892	0.612	0.348	0.510	0.470	0.560
方向	-	-	1.094	0.564	0.886	1.045	1.234	

については4画で統一を行った。また、「良」と「遠」は字画間の繋がりが多く見られ、画数のばらつきが大きく統一が困難であったため、この2文字については字画に関する特性値、始筆・送筆・終筆部に関する特性値を除外した。

前節で示した選択基準より特性値を選択する。まず、文字列、文字、字画の特性値については、S値が0.707未満である特性値の種類は字画の曲直のみであり、字画の曲直以外の特性値を選択する。始筆・送筆・終筆部の特性値のS値は0.707未満のものも含まれているので、すべてを組み合わせた場合と0.707以上の特性値を組み合わせた場合のS値を求めた。ただし、曲直についてはすべてが0.707未満であったので、すべてを組み合わせた場合のみS値を求めた。その結果、平均筆記速度と方向については始筆部を除いた方がS値が大きくなった。表4.1の「全」には、そのときの値を示している。始筆・送筆・終筆部の特性値については、表4.1より、組み合わせることでS値が向上するのは、筆圧順位、平均筆圧、速度順位、平均筆記速度、方向の5種類である。平均ペン高度と方位については、組み合わせたS値が字画のS値とほぼ同じであり組み合わせによる効果が期待できないと判断した。筆圧順位と平均筆圧との比較では、平均筆圧の方がS値が大きい。しかし、選択基準のところでも述べたように、筆圧情報は用いる電子ペンの経時変化等により変動が大きいことが経験的に分かっている。したがって、筆者照合に用いる特性値としては、筆圧順位のS値が1.347と比較的大きいことより、筆圧順位を選択する。また、速度順位と平均筆記速度については、平均筆記速度のS値が明らかに速度順位より大きいことから、平均筆記速度を選択した。

最終的には、平均筆圧、速度順位、字画および始筆・送筆・終筆部の曲直、始筆・送筆・終筆部の平均ペン高度と方位、始筆部の平均筆記速度と方向を除いた特性値を選択した。その結果、特性値の総数563個から270個が選択され、52%の削減が実現できた。

結果的に、選択した特性値の種類は、第3章で全データを用いて特性値を評価した結果に基づいて選択した特性値の種類と同じ結果になった。このことは、本章で用いた特性値選択手法の妥当性を示すものである。

4.4 筆者照合手法

4.4.1 特性値の偏差の度数分布に基づく筆者照合手法

全筆者共通の特性値を用いることから、個人内変動の大きな特性値を持つ筆者が存在する可能性がある。したがって、これら特性値の影響を低減する筆者照合方法が必要になる。そこで、特性値の参照筆跡の平均値からの偏差の度数分布に基づく照合手法を提案する。基本的な考え方は以下の通りである。

- (1) 真正筆跡であるならば、参照筆跡の平均値に近い特性値の数が多く、平均値から離れた特性値の数は少ない。
- (2) 偽造筆跡であれば、真正筆跡に比べ、参照筆跡の平均値に近い特性値の数は少なく、平均値から離れた特性値の数は多いと考えられる。

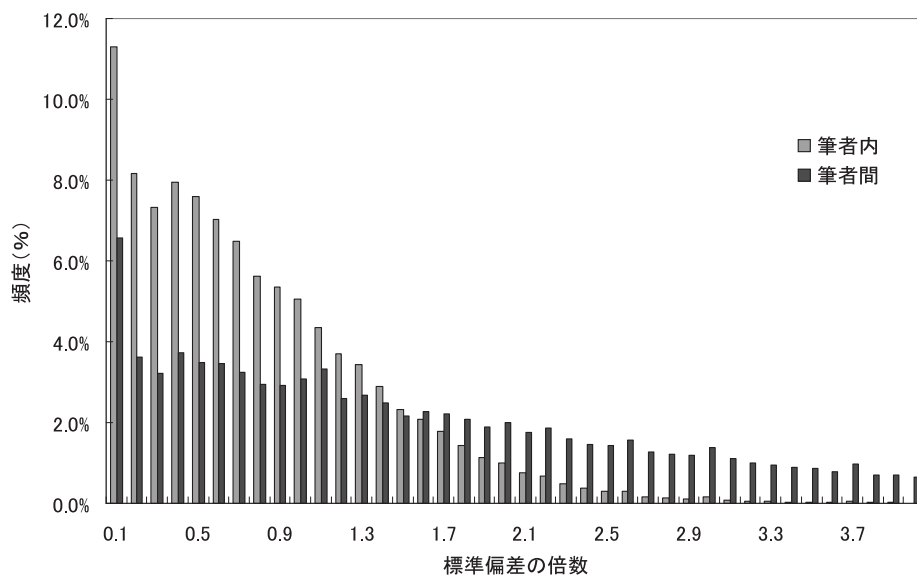


図 4.2 特性値の偏差の度数分布 (全サンプル, 選択した特性値 270 個)

実際にすべてのデータを用いて選択した特性値の偏差の度数分布を求めると図 4.2 のようになり，上記 (1)，(2) の妥当性が確認できた．

この考えに基づく二つの識別器を以下に示す．なお，それぞれの識別器を偏差度数比識別器，偏差度数識別器と呼ぶことにする．

偏差度数比識別器

ここで提案する識別器は，上述の考え方を直接反映するものである．すなわち，偏差の絶対値が標準偏差の α 倍以内の特性値の数 N_α と β 倍以上の数 N_β との比 R を真偽判定の尺度とする．参照筆跡の i 番目の特性値の平均値を m_i ，標準偏差を σ_i とし， α と β を任意の定数， f_i を未知筆跡の i 番目の特性値としたとき，評価尺度 R を以下のように定義する．

$$N_\alpha = \left| \left\{ i : \frac{|f_i - m_i|}{\sigma_i} \leq \alpha \right\} \right| \quad (4.2)$$

$$N_\beta = \left| \left\{ i : \frac{|f_i - m_i|}{\sigma_i} \geq \beta \right\} \right| \quad (4.3)$$

$$R = \frac{N_\beta}{N_\alpha} \quad (4.4)$$

この R の値が小さいほど，未知筆跡は参照筆跡に近い，すなわち本物であると判断できる．そこで，任意のしきい値を T_R としたとき， $R < T_R$ であれば本物，そうでなければ偽物と判断する．

偏差度数識別器

この識別器は，偏差度数比識別器をさらに簡単にしたものであり，偏差度数比識別器における N_α のみを用いて真偽の判定を行う．すなわち，偏差の絶対値が標準偏差の α 倍以内である未知筆跡の特性値の数 N_α が多いほどより本物らしいと判断する．そこで，任意のしきい値を T_α としたとき， $N_\alpha > T_\alpha$ であれば本物，そうでなければ偽物と判断する．

任意定数 α と β の設定方法

識別器には 2 種類の任意定数 α と β があり，これらをどのように設定するかという問題がある．そこで，これら定数を以下のように設定する．全筆者の参照筆跡を用いた Leave-one-out 法による筆者照合実験を行うことにより決定する．具体的には，各筆者 n 個の参照筆跡の内 $n - 1$ 個を学習サンプル，残りの 1 個をテストサンプルとする照合実験を，テストサンプルを順次変えながら n 回繰り返す．これを α と β を変えながら行い，EER が最小になる α と β および，FRR が 0% のとき FAR が最小となる α と β を求める．この 2 種類の設定基準について検討する．

任意定数 α と β の設定結果

上述した設定方法により，任意定数 α と β の設定を行った．なお，設定に用いた参照筆跡は特性値選択に用いたものと同じである．偏差度数比識別器については， α と β を 0.1 刻みに変化させて実験を行った．また，偏差度数識別器についても， α を 0.1 刻みに変化させて実験を行った．EER が最小となる α と β の値と，FRR が 0% のとき FAR が最小となる α と β の値ならびに，そのときの EER と FAR を表 4.2 に示す．なお，EER は FRR と FAR の値が最も近くなるときの FRR と FAR の値を平均したものとした．

表 4.2 に示す結果に従って，偏差度数比識別器では $\alpha = 0.4$ ， $\beta = 2.7$ (EER 最小) と $\alpha = 0.3$ ， $\beta = 2.5$ (FRR が 0% のとき FAR 最小) に任意定数を設定

表 4.2 α と β の設定結果

	設定基準	α	β	EER	FAR (at FRR=0%)
偏差度数比識別器	EER 最小	0.4	2.7	0.14%	0.28%
	FAR (at FRR=0%) 最小	0.3	2.5	0.24%	0.27%
偏差度数識別器	EER 最小	2.4	-	0.24%	0.49%
	FAR (at FRR=0%) 最小	2.5	-	0.33%	0.46%

した。また，偏差度数識別器では， $\alpha = 2.4$ (EER 最小) と $\alpha = 2.5$ (FRR が 0% のとき FAR 最小) に設定した。

4.4.2 比較のために用いる筆者照合手法

提案した識別器の性能を評価するため，特徴パラメータによる筆者照合で一般的に用いられるユークリッド距離による方法，Lee ら [7] が用いた多数決法と提案手法の比較を行う。そこで，この両手法を以下に簡単に説明する。

ユークリッド距離による照合

未知筆跡の参照筆跡に対する距離を，

$$D = \sqrt{\sum_{i=1}^n \frac{(f_i - m_i)^2}{\sigma_i^2}} \quad (4.5)$$

で定義する。求めた距離 D が任意のしきい値 T_D に対して， $D < T_D$ ならば本物，そうでなければ偽物と判定する。

多数決法

多数決法は，偏差の絶対値が標準偏差の γ 倍以内の特性値の数 N_γ

$$N_\gamma = \left| \left\{ i : \frac{|f_i - m_i|}{\sigma_i} \leq \gamma \right\} \right| \quad (4.6)$$

が用いた特性値の数の過半数，すなわち $N_\gamma \geq n/2$ であれば本物，そうでなければ偽物と判定する。

4.5 筆者照合実験

4.5.1 実験方法

照合実験は，第 3 章で用いた 41 人の被験者から各人 30 個収集した 1,230 個のデータ「奈良永遠」に対して行った。各筆者 30 個のデータの内，特性値の選択と任意定数の設定に用いたのと同じ最初の 10 個を参照筆跡とした。未知筆跡

については、各筆者の残り 20 個を未知筆跡として照合実験を行った。したがって、未知筆跡の内 820 個 (20 個 × 41 人) が本物, 32,800 個 (20 個 × 40 人 × 41 人) が偽物となる。

なお、照合のためのしきい値については全筆者共通のしきい値を用いて実験を行った。また、EER については FRR と FAR の値が最も近くなるときの FRR と FAR の値を平均したものとした。

4.5.2 実験結果と検討

表 4.3 に照合実験の結果を示す。

すべての結果で、目標とした EER 2% 程度よりも良い 1% 以下となった。実験に用いるデータベースの違いもあり比較は難しいが、従来の日本語署名を対象としたオンライン署名照合の EER が 1~2% であることと、書き慣れた署名以外の筆跡を用いていることを考慮すると、これは良好な結果であるといえる。また、FRR を 0% としたときの FAR も目標とした 10% 以下となった。とくに、FAR 最小基準で識別器の任意定数の設定を行った偏差度数比識別器において、FAR が 3.63% と良い結果が得られた。

二つの識別器を比較すると、EER は 0.9% 前後で差がないが、FRR を 0% としたときの FAR は偏差度数比識別器の方が偏差度数識別器に比べて 3~5% 低くなった。また、識別器の任意定数 α と β の設定基準の違いによる照合率について比較すると、EER はほとんど差がないといえる。しかし、FRR を 0% にしたときの FAR を最小にする基準で α と β を設定した方が、FAR は低い値を示

表 4.3 照合実験結果

	α, β の設定基準	α	β	EER	FAR (at FRR=0%)
偏差度数比識別器	EER最小	0.4	2.7	0.92%	5.04%
	FAR(at FRR=0%)最小	0.3	2.5	0.86%	3.63%
偏差度数識別器	EER最小	2.4	-	0.82%	8.69%
	FAR(at FRR=0%)最小	2.5	-	0.89%	8.36%

表 4.4 全特性値を用いた照合実験結果

	α, β の設定基準	α	β	EER	FAR (at FRR=0%)
偏差度数比識別器	EER最小	0.8	2.4	1.80%	29.57%
	FAR(at FRR=0%)最小	0.3	1.9	2.13%	16.99%
偏差度数識別器	EER最小	1.3	-	2.19%	30.35%
	FAR(at FRR=0%)最小	1.7	-	1.89%	45.50%

した。特に偏差度数比識別器では、約 1% FAR が低下した。このことから、2つの設定基準間に顕著な差はないが、偏差度数比識別器については FRR を 0% にしたときの FAR を最小にする基準に基づいて α と β を設定した方が良いといえる。

特性値選択の効果を確認するために、全特性値 (563 個) を用いて、同じ方法で識別器の任意定数 α と β を設定し、照合実験を行った。

その結果を表 4.4 に示す。表 4.3 と比較すると、選択した特性値を用いた方が EER, FRR を 0% にしたときの FAR とともに大幅に改善されており、特性値選択が有効であることが示された。

4.5.3 任意定数設定方法の妥当性の検討

識別器の任意定数 α と β の設定方法の妥当性について検討した。二つの識別器について、 α と β を 0.1 刻みで変化させ、EER が最小となる α と β を求めた。その結果、偏差度数比識別器では $\alpha = 1.1, \beta = 2.5$ で EER が 0.56%、偏差度数識別器では $\alpha = 2.2$ で EER が 0.82% であった。これらの任意定数を用いた識別器と、参照筆跡を基に FRR を 0% にしたとき FAR を最小にする基準で設定した任意定数を用いた識別器との照合性能の比較を行うため、ROC を求め図 4.3 と図 4.4 に示す。

偏差度数比識別器では、提案した任意定数の設定方法による識別器の方が EER 付近での性能は若干劣るが、両者で大きな差はないといえる。また、偏差度数識別器では両者ともほぼ同程度の性能を持つ。以上のことより、識別器の任意定数

の設定方法は妥当な方法であると考えられる。

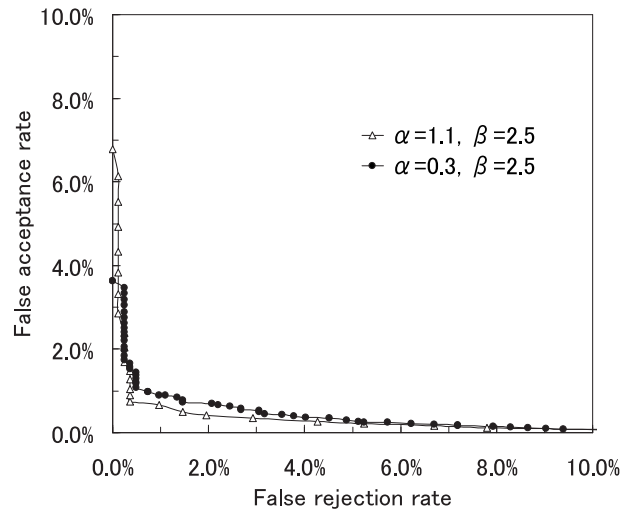


図 4.3 偏差度数比識別器の ROC 曲線

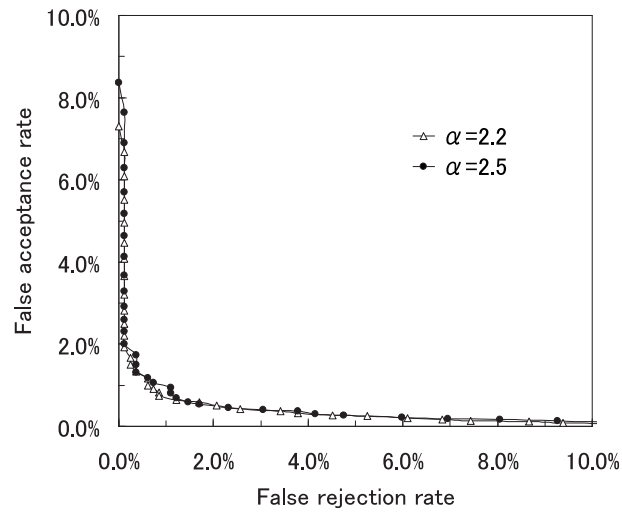


図 4.4 偏差度数識別器の ROC 曲線

4.5.4 ユークリッド距離法，多数決法との比較

ユークリッド距離法，多数決法との比較を行うため，ROC を求めた．その結果を，図 4.5 に示す．ただし，使用した特性値は選択した 270 個である．図より明らかのように，提案した偏差度数比識別器と偏差度数識別器の方が，ユークリッド距離，多数決法に比べ，照合性能が高い．なお，ユークリッド距離による照合では，EER は 2.89% ，多数決法では，EER は 1.63% であった．

提案した識別器の照合性能がユークリッド距離や多数決法に比べ高かった理由は，以下のように考えられる．全筆者共通で有効な特性値を選択することで，筆者によっては個人内変動が大きい特性値が含まれ，それが照合に悪影響を及ぼすと考えられる．ユークリッド距離の場合は，その影響を直接受けるが，提案した識別器は偏差の度数分布に基づくため，個人内変動が大きく平均から大きく外れる特性値があった場合にも，その影響を小さく抑えられることが，照合性能が高かった理由であるといえる．

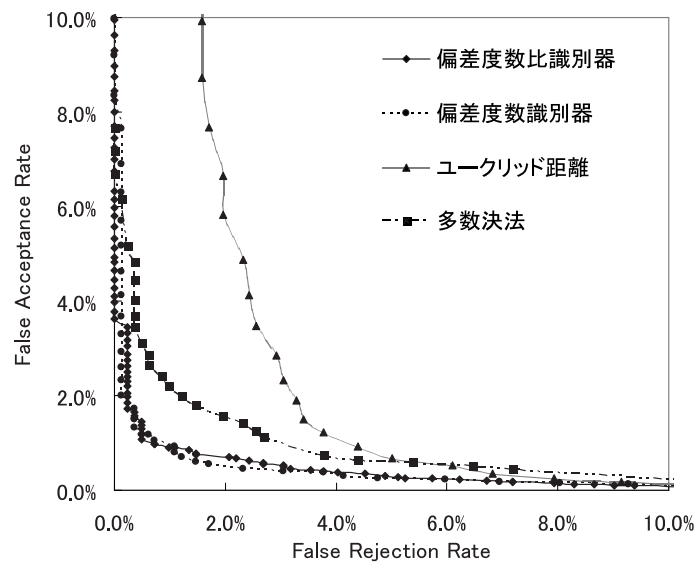


図 4.5 ROC 曲線の比較

多数決法は標準偏差の倍数 γ をしきい値とし、特性値の数が過半数であれば本物、そうでなければ偽物と判定するが、偏差度数識別器は参照筆跡を用いて標準偏差の倍数 α を設定し、照合では特性値の数をしきい値としている。したがって、多数決法は標準偏差の倍数と特性値の数という二つの任意定数の内、特性値の数を事前に固定した偏差度数識別器の特別な場合と考えられる。よって、参照筆跡から求めた偏差の度数分布より標準偏差の倍数を設定し、特性値の数をしきい値とする偏差度数識別器の方が、偏差の度数分布により柔軟に対応できるため、多数決法よりも高い照合性能を示したといえる。また、偏差度数比識別器は標準偏差の倍数に関して二つの任意定数 α と β を持ち、偏差の度数分布をより反映できるものと考えられる。

4.5.5 特性値の種類が照合結果に及ぼす影響の検討

選択した 270 個の特性値の内、特定の特性値の種類が、どの程度照合性能に影響を及ぼすかについて検討する。検討方法は、選択した 270 個の特性値から対象とする特性値の種類を除外して照合実験を行い、EER と FRR を 0% としたときの FAR を求めることで行う。すなわち、選択した 270 個を用いた場合の

表 4.5 特定の特性値の種類を除外したときの照合結果

除外した特性値の種類 (特性値の数)	EER	270個を用いた 場合の EER = 0.86% に対す る増加率	FAR (at FRR=0%)	270個を用いた 場合の FAR = 3.63% に対す る増加率
ペンダウン時間(17個)	0.79%	-8.14%	6.80%	+87.33%
筆記タイミング(10個)	1.15%	+33.72%	8.15%	+124.52%
筆圧順位(48個)	0.96%	+11.63%	6.55%	+80.44%
ペンの傾き(34個)	1.43%	+66.28%	10.95%	+201.65%
平均筆記速度(41個)	1.19%	+38.37%	7.73%	+112.95%
相対始点位置(24個)	1.52%	+76.74%	5.98%	+64.74%
長さ(17個)	1.11%	+29.07%	8.00%	+120.39%
方向(36個)	0.88%	+2.33%	3.78%	+4.13%

EER=0.86% , FAR=3.63% に対して, 求めた EER, FAR の増加率が大きいほど照合結果に大きな影響を与える, すなわち照合に有効な特性値の種類であるといえる.

対象とする特性値の種類は, 特性値の数が多く S 値が大きいもの, 具体的には特性値の数が 10 個以上ですべての部分あるいはすべての特性値を用いたときの S 値が 1 を超える, ペンダウン時間, 筆記タイミング, 筆圧順位, ペンの傾き (平均ペン高度と方位), 平均筆記速度, 相対始点位置, 長さ, 方向を対象とした. また, 照合には照合性能が高かった偏差度数比識別器を用いる.

実験結果を表 4.5 に示す. ペンダウン時間を除外した場合の EER 以外で増加率が正となり, 対象とする特性値の種類は照合に有効であるといえる.

特性値の種類を比較すると, ペンの傾きは他の特性値の種類に比べて EER, FAR とともに増加率が大きく, 照合に有効である. 相対始点位置も EER の増加率が一番大きく, 他の特性値の種類に比べて照合に有効であるといえる. 筆記タイミング, 平均筆記速度, 長さの EER と FAR の増加率は同程度であり, ペンの傾き, 相対始点位置に次いで増加率が大きい. 筆圧順位はこれらより若干増加率が劣る. 逆に, 方向は EER, FAR とともに増加率が小さく, 他の特性値の種類に比べて照合結果に与える影響は少ない. ペンダウン時間は EER の増加率がわずかではあるが減少するが, FAR の増加率は大きく照合に有効であると考えられる.

以上のように, 特性値の種類により照合結果に与える影響に違いはあるが, 定義した特性値は有効に機能しているといえる. また, 従来の特徴パラメータに基づく筆者照合では用いられていない特性値である筆圧順位と筆記タイミングが有効であることが確認できた.

4.6 結言

本章では, 漢字文字列から抽出した筆跡鑑定の知見に基づく特性値を用いたオンライン筆者照合方法について検討した. まず, 参照筆跡を用いて特性値の種類ごとに筆者内距離と筆者間距離を求め, その分布に基づいて全筆者共通の特性値の種類を選択した. つぎに, 特性値の平均値からの偏差の度数分布に基づく識別器を 2 種類提案し, 参照筆跡を用いた設計方法を示した.

提案手法の性能を評価するために筆者照合実験を行った。その結果、EER 0.82~0.92%、FRR を 0% に設定したときの FAR が 3.63~8.69% であった。これは目標とした EER 2% 程度、FRR 0% のときの FAR 10% 以下を満足する良好な結果である。さらにユークリッド距離による方法および多数決法との比較を行い、提案手法の方が良い結果を示した。

特性値の種類が照合結果に与える影響についても検討した。その結果、特性値の種類により照合結果に与える影響に違いはあるが、定義した特性値が有効に機能していること、特に平均ペン高度と方位、相対始点位置が照合に有効であることが明らかになった。また、従来の特徴パラメータに基づく筆者照合では用いられていない特性値である筆圧順位と筆記タイミングが有効であることを確認した。

想定している筆者照合システムは漢字数文字で構成された共通のパスワードを入力するものであり、利用者から共通の文字列が参照筆跡として事前に複数収集可能である。これを利用することで、提案した筆者照合方法は、照合に有効な特性値の種類が選択可能であり、参照筆跡から求めた特性値の偏差の度数分布に基づく識別器が設計でき、有効に機能することが確認できた。提案した偏差度数比識別器および偏差度数識別器は、アルゴリズムが簡単であり特性値の雑音に強い有効な識別器であるといえる。

第 5 章

偽造筆跡と他文字列に対する提案手法の性能評価

5.1 緒言

第 4 章において，真正筆跡の各特性値は参照筆跡の平均値に近いものが多く，遠いものは少ないという考えに基づき，各特性値の参照筆跡の平均値からの偏差の度数分布を基にした識別器を提案し，その有効性を照合実験により確認した．しかし，この照合実験で用いた筆跡データは故意に筆跡を真似た偽造筆跡を含んでいなかった．また，「奈良永遠」という筆跡のみを対象としていた．

そこで，本章では被験者が十分に訓練を積んだ「奈良永遠」の偽造筆跡を収集し，提案した識別器の偽造筆跡に対する識別性能を評価する．さらに，「奈良永遠」以外の文字列として，3.3.1 で述べた漢字文字列の選択基準に従い，かつ「奈良永遠」で含まれない文字の画数，3 画，10 画を含む「大和郡山」を対象として，識別性能を評価する．

まず，5.2 節で偽造データの収集と特性値の種類ごとに真筆内距離と真筆偽筆間距離の分布を求めることにより偽造データを分析した結果について述べる．さらに，偽造データを用いた筆者照合実験を行い提案識別器の偽造データに対する識別性能を評価する．つぎに，5.3 節で「大和郡山」を用いて照合実験を行い，「奈良永遠」と比較することで，提案手法の性能について検討する．最後に，5.4

節でこの章のまとめを行う。

5.2 偽造筆跡に対する照合性能評価

5.2.1 偽造筆跡データの収集と分析

偽造データの収集方法

偽造筆跡は、既に収集した「奈良永遠」の被験者 41 人の内の任意の 24 人の筆跡を対象に以下の条件で収集した。偽造者は、新たな被験者 13 人と対象とする筆跡の筆記者 24 人の内 8 人の計 21 人である。これら偽造者を各筆跡に対して 4~6 人割り当て、事前に筆跡のコピー（筆順情報も付与）を渡し十分な練習を行った上で対象筆跡を 5 回筆記して貰った（一部のデータで 5 回以上筆記したものを含む）。この収集条件は、既に書かれてある筆跡を手に入れることは可能であるが、筆記速度や筆圧などの動的特徴を観察することは容易でないことを考慮したものである。したがって、被験者は本物筆跡の形のみを情報として持ち、

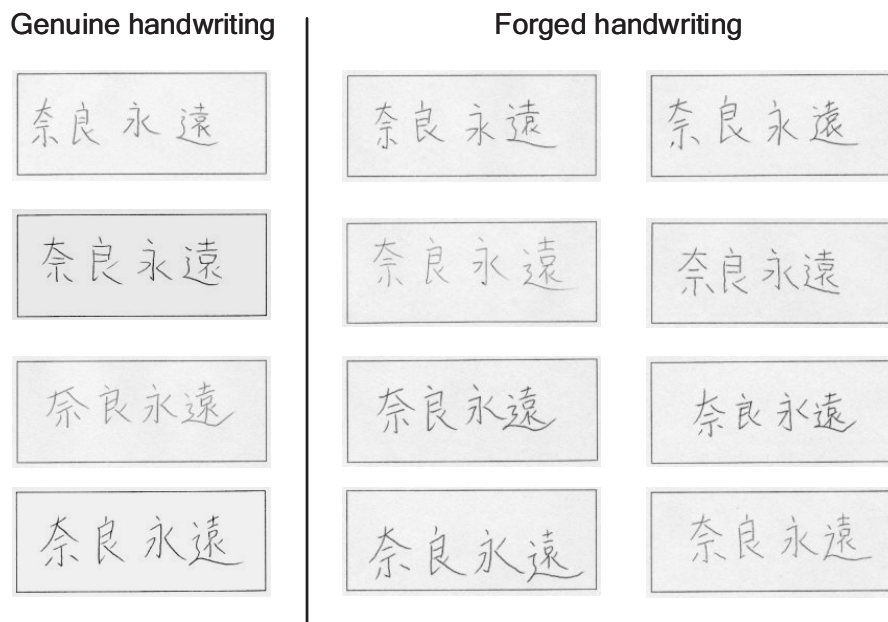


図 5.1 収集したデータの例「奈良永遠」の本物と偽造筆跡

表 5.1 偽造筆跡データ 566 個の内訳

対象筆記者	偽造者	偽造筆跡数	備考	
A	a, b, c, d, e 5人	18	a は10個筆記	
B	a, b, d, f, g 5人	21		
C	a, b, c, f, h 5人	23		
D	a, b, i, k 4人	20		
E	a, f, k, l 4人	14		
F	b, k, m, n, o, p 6人	29		
G	a, b, c, l, m 5人	25		
H	a, b, d, e, k 5人	29		
I	a, g, k, q 4人	20		
J	b, f, i, k, o 5人	25		
K	b, e, f, k, r 5人	23		
L	h, k, m, o, p 5人	25		
M	g, l, m, s, t 5人	25		
N	a, b, c, q 4人	20		
O	a, b, h, k, n, o 6人	30		
P	b, g, h, j, l, s 6人	29		
Q	b, g, k, l, p 5人	25		
R	b, c, d, k, m 5人	25		
S	f, h, p, s, t 5人	22		
T	a, j, m, s, t 5人	20		
U	b, d, k, o, u 5人	25		k は6個筆記
V	d, f, o, r, u 5人	25		
W	b, e, j, l, s 5人	25		
X	a, f, h, s, t 5人	23		

a, b, f, j, k, l, p, s の 8名は, 対象筆記者F, L, D, M, S, T, O, E と同一人物である.

動的情報については納得がいくまで練習してもらうことで極端に筆記速度が遅くなったり，筆記が滞ることを極力排除した．

実験に用いたのは，文字切り出しやストロークの対応付けに失敗した筆跡を除いた 566 個（各筆跡 14～30 個）である．なお，表 5.1 に収集した偽造データの内訳を，また図 5.1 に収集した偽造データの一部を示す．

真筆内距離と真筆偽筆間距離の分布による分析

収集した偽造データの特性値がどの程度真筆と似ているのかを，第 3 章で特性値の評価に用いた S 値により評価する．

具体的には，対象とする筆記者 24 人の筆跡データに対して，本物 30 個のすべての組み合わせ $C_{30}^2 = 435$ 組と，本物 30 個と偽造筆跡 14～30 個のすべての組み合わせ 420～900 組のユークリッド距離を求め，(3.1) 式により特性値の種類ごとに S 値を計算した．評価には，対象とする筆記者 24 人の S 値の平均を用いる．また比較のため，筆記者 24 人の内，本人以外の筆跡を偽筆とした場合の S 値の平均を求めた．その結果を，表 5.2 に示す．

表より，次のことがいえる．静的特性値については，故意に真似た偽造筆跡の S 値がすべての特性値の種類で故意に真似ていない偽筆（筆記者自身の筆跡）の S 値よりも小さくなった．これは，練習を積むことで文字の大きさや形が本物に近づいたことを意味する．逆に，動的特性値については，故意に真似た偽造筆跡の S 値がほとんどの特性値の種類で大きくなった．これは，文字の形を真似ることに意識がいきスムーズな筆記動作が行えない結果生じたと考えられる．このことは，筆記時間，ペンダウン時間，ペンアップ時間の S 値の増加が大きいことから裏付けられる．

また，静的特性値の中で文字列と文字の曲直の S 値が両者間でほとんど差がない．これは，筆跡を文字列や文字などの字画に比べて長い筆記長さでみたとき，その形態的特徴（線の直線性や曲がり具合など）まで真似ることは難しいことを示唆している．

表 5.2 偽造筆跡の評価値 S

上段は筆記者24人の筆跡データを用いた S 値であり,
下段は24人の筆跡データと偽造筆跡データを用いた S 値である.

特性値	文字列	文字	ストローク	始・送・終筆部			
				始筆	送筆	終筆	
動的	筆記時間	0.836	0.857	-	-	-	-
		1.725	1.723	-	-	-	-
	ペンダウン時間	0.956	1.059	1.298	-	-	-
		1.677	1.749	1.827	-	-	-
	ペンアップ時間	0.811	0.807	-	-	-	-
		1.732	1.681	-	-	-	-
	筆記タイミング	-	-	0.918	-	-	-
		-	-	1.437	-	-	-
	筆圧順位	-	-	1.053	0.673	0.401	0.628
		-	-	0.963	0.872	0.874	1.204
	平均筆圧	0.739	0.852	1.163	1.026	1.178	1.048
		1.097	1.167	1.461	1.240	1.587	0.900
	平均ペン高度	0.872	0.952	0.998	1.034	0.984	0.978
		0.930	0.984	1.065	1.076	1.041	1.045
	平均ペン方位	1.261	1.278	1.312	1.310	1.310	1.328
		1.668	1.701	1.704	1.682	1.707	1.716
	速度順位	-	-	0.653	0.494	0.477	0.743
		-	-	0.906	0.303	0.402	0.563
	平均筆記速度	1.148	1.287	1.252	0.445	1.442	1.229
		1.430	1.510	1.417	0.557	1.578	1.260
静的	大きさ	0.870	1.030	-	-	-	-
		0.331	0.544	-	-	-	-
	縦横比	0.537	0.904	-	-	-	-
		0.226	0.519	-	-	-	-
	中心位置	0.828	1.011	-	-	-	-
		0.627	0.725	-	-	-	-
	字間隔	0.794	-	-	-	-	-
		0.592	-	-	-	-	-
	相対始点位置	-	-	1.784	-	-	-
		-	-	1.088	-	-	-
	長さ	0.786	0.962	1.388	-	-	-
		0.544	0.642	0.830	-	-	-
	曲直	0.866	0.876	0.565	0.326	0.435	0.449
		0.860	0.871	0.431	0.241	0.398	0.365
	方向	-	-	1.031	0.550	0.806	0.907
		-	-	0.568	0.353	0.522	0.535

5.2.2 筆者照合実験による照合性能評価

実験方法

照合実験は、各筆者 30 個のデータの内、最初の 10 個、計 240 個を参照筆跡とし、各筆者の残り 20 個と偽造筆跡、計 1,046 個（本物 480 個と偽物 566 個）を未知筆跡として用いた。ただし、故意に真似た偽造筆跡に対する照合性能を評価するため、各筆者の残り 20 個は本物の未知筆跡とし、他の筆者の偽筆としては用いていない。

また、使用する特性値は、第 4 章で選択した 270 個である。さらに、識別器の任意定数 α と β については、第 4 章で設定した値をそのまま用いた。このことで、偽造筆跡に関する事前の知識を用いずに、提案した筆者照合システムの偽造筆跡に対する性能評価が可能になる。

なお、照合のためのしきい値については全筆者共通のしきい値を用いて実験を行った。また、EER については FRR と FAR の値が最も近くなるときの FRR と FAR の値を平均したものとした。

実験結果と検討

表 5.3 に照合実験の結果を示す。なお、比較のため第 4 章で行った照合実験結果も併せて示す。表中データセット 1 とあるのが、故意に真似た偽造筆跡を含まない第 4 章で用いたデータである。また、区別するために本章で収集したデータセットをデータセット 2 と呼ぶことにする。

すべての結果で、両データセットの EER は差が無いが、故意に真似ていない偽造筆跡を対象としたデータセット 1 に比べて故意に真似た偽造筆跡を対象としたデータセット 2 の方が、FRR を 0% としたときの FAR が低くなった。特に偏差度数識別器では、FAR が 1% 台とデータセット 1 の 8% 台に比べて大幅に照合性能が高くなっている。このことは、提案した識別器が本物を真似た偽造筆跡に対して、本物を故意に真似ない偽造筆跡に対するのと同様以上の照合性能を持つことを意味する。

ただし、今回用いた偽造筆跡は静的特徴のみを偽造者に与えており、動的特徴

表 5.3 照合実験結果

	α	β	データセット1		データセット2	
			EER	FAR (at FRR=0%)	EER	FAR (at FRR=0%)
偏差度数比識別器	0.4	2.7	0.92%	5.04%	0.86%	4.24%
	0.3	2.5	0.86%	3.63%	0.86%	3.18%
偏差度数識別器	2.4	-	0.82%	8.69%	0.86%	1.94%
	2.5	-	0.89%	8.36%	0.86%	1.41%

データセット1: 真正筆跡 20個×41人

データセット2: 真正筆跡 20個×24人+偽造筆跡(566個)

を同時に与えた場合については検討していない。しかし、静的特徴と動的特徴を同時に真似ることは困難であると考えられることと、偽造者がパスワードとその筆跡の静的、動的特徴を同時に盗むことが容易でないことを考慮すると、提案した識別器は故意に真似た偽造筆跡に対しても有効に機能すると考えられる。

二つの識別器の性能の違いについて検討する。データセット1については偏差度数比識別器の方が FAR は低く、データセット2では偏差度数識別器の方が FAR が低くなった。このような結果になった理由として、以下のことが考えられる。5.2.1 の分析結果から故意に真似た偽造筆跡では、静的特性値はより本物に近づき、動的特性値はより本物から離れる。したがって、偏差度数比識別器では標準偏差の α 倍以内の特性値数 N_α と標準偏差の β 倍以上の特性値数 N_β との比を取るため、データセット1と2では FAR にあまり差が現れなかったと考えられる。しかし、偏差度数識別器では標準偏差の α 倍以内の特性値数 N_α のみが判定の基準になることから、データセット2では標準偏差の α 倍を超える特性値の数が増加するため、FAR が大幅に低下したと考えられる。結果から見れば、故意に真似た偽造筆跡に対しては偏差度数識別器が有効であり、故意に真似ない偽造筆跡に対しては偏差度数比識別器が有効である。総合的に見ると偏差度数比識別器の方が良い照合性能を示すといえる。

5.3 他文字列に対する照合性能評価

5.3.1 データ収集

「奈良永遠」では8画,7画,5画,13画の文字を用いたので,ここでは,3.3.1で示した漢字文字列の選択基準に従い,かつ「奈良永遠」では用いなかった3画と10画の漢字を含む意味のある文字列「大和郡山」(地名)を選択した。

男性15人(内2人は左利き),女性5人,計20人の被験者から,1日1回5個ずつ,これを1ヶ月から2ヶ月の期間をかけて各被験者から30個以上のデータを収集した。この内から各被験者30個,合計600個のデータを使用する。なお,この20人の被験者は「奈良永遠」を書いた被験者とは別人である。収集したデータの例を図5.2に示す。

このように文字列と被験者が「奈良永遠」と違うデータベースを構築することで,照合性能に影響すると考えられる筆記者と筆記文字列が同時に変わった場合の提案手法の照合性能を評価する。

5.3.2 特性値の選択

「奈良永遠」と同様に,各被験者の最初の10個を参照筆跡としてS値を計算し特性値選択を行った。ただし,「郡」については全筆者の画数の統一が困難であったため,字画に関する特性値と始筆・送筆・終筆部に関する特性値を除外した。

求めたS値を表5.4に示す。なお,比較のため「奈良永遠」のS値も併せて記載した。S値は「奈良永遠」と同様の傾向を示し,4.3.1で述べた特性値選択基準により全特性値644個から323個が選択され,49.8%の特性値が削減された。「奈良永遠」で選択されたが「大和郡山」で選択されなかった特性値の種類は文字列と文字のペンアップ時間,文字列の縦横比と中心位置のみであった。また,「奈良永遠」で選択されなかったが「大和郡山」で選択された特性値の種類は字画の曲直のみであった。

「奈良永遠」で選択されたが「大和郡山」で選択されなかった特性値の種類は,筆記した文字に依存する特性値ではなく,筆記のスムーズさや文字列の配置に関

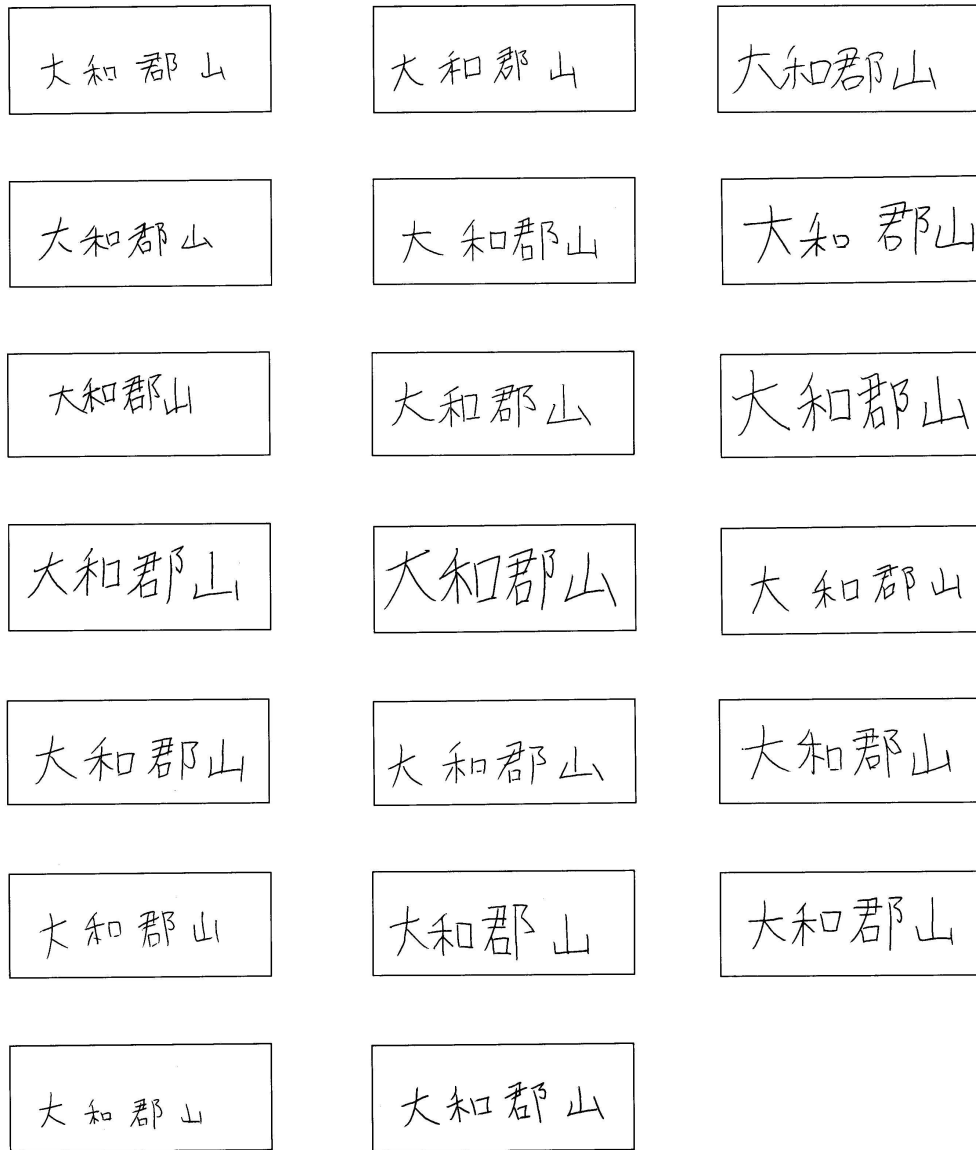


図 5.2 収集したデータの例「大和郡山」

表 5.4 評価値 S

表中、上段は「大和郡山」のS値を、下段は「奈良永遠」のS値を示す。

特性値	文字列	文字	字画	始・送・終筆部			全	
				始筆部	送筆部	終筆部		
動的	筆記時間	0.914	0.970	-	-	-	-	-
		0.934	0.986	-	-	-	-	-
	ペンドアウン時間	1.162	1.220	1.252	-	-	-	-
		1.028	1.169	1.529	-	-	-	-
	ペンアップ時間	0.667	0.668	-	-	-	-	-
		0.891	0.960	-	-	-	-	-
	筆記タイミング	-	-	0.951	-	-	-	-
		-	-	1.134	-	-	-	-
	筆圧順位	-	-	0.863	0.907	0.490	0.939	1.074
		-	-	1.165	0.914	0.574	0.843	1.347
	平均筆圧	1.049	1.127	1.288	1.231	1.183	1.181	1.476
		0.943	1.049	1.427	1.156	1.395	1.354	1.591
	平均ペン高度	1.189	1.272	1.276	1.296	1.273	1.259	1.282
		1.206	1.269	1.290	1.306	1.280	1.292	1.304
	平均ペン方位	1.857	1.861	1.843	1.838	1.847	1.836	1.845
		1.683	1.685	1.685	1.678	1.683	1.692	1.690
	速度順位	-	-	0.831	0.627	0.821	0.840	1.067
		-	-	0.773	0.516	0.576	0.763	1.029
	平均筆記速度	1.057	1.146	1.343	0.807	1.477	1.351	1.587
		1.096	1.213	1.304	0.388	1.479	1.273	1.640
静的	大きさ	0.962	1.059	-	-	-	-	-
		1.031	1.216	-	-	-	-	-
	縦横比	0.649	0.791	-	-	-	-	-
		0.772	0.978	-	-	-	-	-
	中心位置	0.601	0.801	-	-	-	-	-
		0.941	1.081	-	-	-	-	-
	字間隔	1.118	-	-	-	-	-	-
		0.928	-	-	-	-	-	-
	相対始点位置	-	-	2.477	-	-	-	-
		-	-	2.022	-	-	-	-
	長さ	1.073	1.192	1.592	-	-	-	-
		1.062	1.197	1.670	-	-	-	-
	曲直	0.824	0.833	0.720	0.336	0.528	0.594	0.586
		0.933	0.892	0.612	0.348	0.510	0.470	0.560
	方向	-	-	1.286	0.503	1.343	1.154	1.350
		-	-	1.094	0.564	0.886	1.045	1.234

する特性値である．また，「大和郡山」で字画の曲直が選択されたのは，「大」と「山」の各字画は「奈」や「永」に比べて字画の長さが長いため，曲直に個人性が表れやすいのではないかと推測できる．

全部で 61 個の特性値の種類の内，二つの文字列で選択の違いが生じたのは 5 種類であり，本研究で用いた特性値選択の方法は，文字列と被験者が変わってもほぼ同等に機能することが示された．

5.3.3 照合実験

「奈良永遠」と同様に，各被験者の最初の 10 個を参照筆跡として識別器の任意定数を設定し，筆者照合実験を行った．照合実験は，各被験者の最初の 10 個を参照筆跡，残り 20 個を未知筆跡とした．したがって，未知筆跡の内 400 個（20 個 × 20 人）が本物，7,600 個（20 個 × 19 人 × 20 人）が偽物となる．なお，照合のためのしきい値については全筆者共通のしきい値を用いて実験を行った．また，EER については FRR と FAR の値が最も近くなるときの FRR と FAR の値を平均したものとした．

照合結果を表 5.5 に示す．また，「奈良永遠」の照合結果と比較するために，選択した特性値を使用した場合の両文字列に対する照合結果を表 5.6 に示す．

表 5.5 より，両識別器とも全特性値を使用した場合と，選択した特性値を使用した場合とで EER については両者とも 2% 台で大きな差はないが，FRR を 0% にしたときの FAR は偏差度数比識別器で約 3~12% ，偏差度数識別器では約 21~29% 改善されてる．このことより，特性値選択の効果が確認できる．

表 5.5 照合実験結果

	設定基準	選択した特性値(323個)		全特性値(644個)	
		EER	FAR (at FRR=0%)	EER	FAR (at FRR=0%)
偏差度数比識別器	EER 最小	2.48%	6.55%	1.50%	18.93%
	FAR (at FRR=0%) 最小	2.47%	7.59%	2.13%	10.92%
偏差度数識別器	EER 最小	2.68%	10.20%	2.45%	31.72%
	FAR (at FRR=0%) 最小	2.71%	8.42%	2.26%	37.12%

表 5.6 「奈良永遠」と「大和郡山」の照合実験結果の比較（選択した特性値を使用した場合）

	設定基準	「奈良永遠」		「大和郡山」	
		EER	FAR (at FRR=0%)	EER	FAR (at FRR=0%)
偏差度数比識別器	EER 最小	0.92%	5.04%	2.48%	6.55%
	FAR (at FRR=0%) 最小	0.86%	3.63%	2.47%	7.59%
偏差度数識別器	EER 最小	0.82%	8.69%	2.68%	10.20%
	FAR (at FRR=0%) 最小	0.89%	8.36%	2.71%	8.42%

「奈良永遠」の結果と比較すると、選択した特性値を用いた場合 EER が偏差度数比識別器で 1.6%、偏差度数識別器で 1.8% 程度高くなっている。また、FRR を 0% としたときの FAR は、識別器と任意定数の設定基準により差がある。偏差度数比識別器では EER を最小にする基準で任意定数を設定した場合 1.5% 程度、FAR を最小にする基準では 4% 程度高くなっている。偏差度数識別器では EER を最小にする基準で任意定数を設定した場合 1.5% 程度高く、FAR を最小にする基準ではほぼ同程度となっている。

「奈良永遠」より照合性能が劣った理由としては、画数が少なく文字線構造が簡単な「大」と「山」を含んでいること、また被験者が違うことなどが考えられる。しかし、画数が少なく文字線構造の簡単な文字を含んでも、EER が 2.5% 程度であり、FRR を 0% にしたときの FAR が目標とした 10% 以下であった。したがって、提案した筆者照合方法は文字列や被験者の違いによる影響を受けるが、その影響は小さいといえる。

二つの文字列に対する実験結果から、以下のことがいえる。横書きされた 10 画以下を中心とした 4 文字の漢字文字列を選択することで、その内の 2~3 文字で画数の統一が可能となる。このように 4 文字中 2~3 文字で画数の統一が可能であれば、提案手法は目標とした EER 2% 程度、FRR を 0% にしたときの FAR 10% 以下をほぼ満足すると考えられる。

また、総合的にみると、識別器については EER の差はほとんど無いが FRR を 0% にしたときの FAR は偏差度数比識別器のほうが小さくなり、良い照合性能を示すといえる。識別器の任意定数の設定基準については、FRR を 0% にし

たとき FAR を最小にする基準と、EER を最小にする基準はほぼ同程度であるが、FAR を最小にする基準の方が若干良い照合性能を示すといえる。

5.4 結言

本章では、提案した筆者照合方法の訓練された偽造筆跡と他文字列に対する識別性能を評価した。

その結果、提案した筆者照合方法は、訓練された偽造筆跡に対して、偽造者自身の筆跡による偽造筆跡に対してよりも良い照合性能を持つことが明らかになった。これは、訓練された偽造筆跡は静的特徴は良く真似ているが動的特徴に対しては既知でないため真似られず、参照筆跡に近い特性値の数は増えるが同時に参照筆跡から遠い特性値の数も増えた結果である。一般的に筆跡の形を真似ようとするれば動的特徴が真似られず、逆に動的特徴を真似ようとする静的特徴が真似られないと考えられる。したがって、提案手法のように静的特徴と動的特徴を総合的に評価する手法は偽筆排除の有効な方法であるといえる。

また、「奈良永遠」以外の文字列「大和郡山」に対する照合実験より、提案した筆者照合方法は文字列や被験者の違いによる影響を受けるが、その影響は小さく、目標とした EER 2% 程度、FRR を 0% にしたときの FAR 10% 以下をほぼ満足するものであることが確認できた。

以上より、想定する、システムが設定した共通の漢字文字列を入力するオンライン筆者照合システムは、3.3.1 で示した漢字文字列の選択基準を満たす 4 文字の文字列を用いることで、従来のオンライン署名照合と同程度の照合性能を有し、故意に真似た偽造筆跡に対しても有効であることが示された。

第 6 章

結論

本論文は，オンライン筆者照合において登録筆跡を署名に限定するのではなく，より一般的な漢字文字列を対象に，どのような特性値に個人性が現れるのかをまず明らかにし，その結果に基づいたオンライン筆者照合方法を提案し，実験によりその有効性を評価することを目的としたものである．

特性値については，書写技能やそれを包含する筆跡鑑定の知見に基づいて静的特性値と動的特性値を定義し，それら特性値が個人性を表すことと，照合に有効な特性値を明らかにした．さらに，システムが設定した共通の漢字文字列（パスワード）を入力するオンライン筆者照合システムを想定し，利用者から 10 個程度の参照筆跡を収集することで，全利用者共通で照合に有効な特性値を選択する方法と，特性値の偏差の度数分布に基づく識別器の設計方法を提案し，それらが有効に機能することを示した．

以下に，本研究で得られた主要な成果を要約する．

1. 本研究のアプローチの正しさを確認するため，漢字と字画の形態が似ているカタカナ文字列を対象に予備実験を行った．4 人の被験者が一定枠内に横書きしたカタカナ文字列「チキュウ」から，書写技能に基づく特性値を機械的に抽出し，それら特性値に個人性が現れるかについて分析を行い，以下のことを明らかにした．
 - (a) 一元配置分散分析の結果，抽出した 230 個の特性値の内 223 個が有

意水準 5% で有意となり，抽出した特性値は文字や点画の違いにより現れ方に差があるものの個人性を表していること．

- (b) 変異係数を求めることにより，文字の配置や大きさに関する特性値の個人内でのばらつきは小さく，始筆部・終筆部の方向や送筆部の曲直などの局所的な特性値の個人内のばらつきが大きいこと．
 - (c) 始筆・送筆・終筆の形などを比較的安定に書く人と，そうでない人がいること．
 - (d) 特性値間の相関係数を求め，筆者に依存せず相関が強い特性値の組み合わせと，筆者に依存して相関の強い組み合わせがあること．
2. 書写技能に基づく特性値を用いた筆者照合・識別の可能性を，1 と同じカタカナ文字列「チキュウ」に対して筆者照合・識別実験を行い検討した．被験者 4 人に対する結果ではあるが識別率 100% ，照合率 100% となり，書写技能に基づく特性値を用いた筆者認識の可能性を示唆するものであった．
3. 文字列を一般的な漢字とし，特性値も書写技能に基づく特性値を包含する筆跡鑑定の知見に基づく特性値として整理，追加を行い，被験者数を 41 人と拡張することで，より一般的な検討を行った．漢字文字列の選択基準を示し，それを満たす一定枠内に横書きされた漢字 4 文字「奈良永遠」から，筆跡鑑定の知見に基づく特性値を抽出し，
- (a) これら特性値に対して，一元配置分散分析，クラスカル・ウォリス検定を行った結果，抽出した 563 個の特性値の内 561 個が有意水準 1% で有意となり，抽出した特性値は程度の差はあるが個人性を表すことを明らかにした．
 - (b) 特性値空間から距離空間へ写像し筆者内距離と筆者間距離の分布を求めることにより，同じ種類の特性値を組み合わせたときの特性値の識別能力について評価を行い，全筆者共通で照合に有効な特性値を明らかにした．また，この評価尺度を用いた，特性値選択の基準を示した．
 - (c) 個人性をよく表すと考えられる特性値の種類は，平均ペン方位と高度，平均筆圧，筆圧順位，平均筆記速度，相対始点位置，長さであり，逆に個人性が現れにくいと考えられる特性値の種類は字画と始筆・送

筆・終筆部の曲直，始筆部の方向と平均筆記速度である．また，新たな特性値である筆圧順位は，従来の平均筆圧よりも個人識別力が高いことも明らかにした．

(d) ユークリッド距離に基づく筆者照合・識別実験を行うことで，これら特性値を用いた筆者認識の可能性を検討した．その結果，識別率 96.5% ，FRR 3.21% ，FAR 3.17% であり，これら特性値を用いた筆者認識の可能性を確認した．

4. 特性値の偏差の度数分布に基づく，偏差度数比識別器，偏差度数識別器を提案し，その有効性について検討した．提案した識別器と，特徴パラメータを用いた筆者照合で用いられるユークリッド距離に基づく方法，多数決法とを比較したところ，提案識別器は EER 0.82~0.92% ，FRR を 0% としたときの FAR 3.63~8.69% であり，他の手法より良い照合性能を示し，提案手法の有効性が確認できた．
5. 偏差度数比識別器と偏差度数識別器の偽造筆跡に対する照合性能を検討するために，「奈良永遠」の十分に練習を積んだ偽造筆跡を収集し，筆者照合実験を行った．その結果，両識別器の EER は 0.86% ，FRR を 0% としたときの FAR 1.41% ~4.24% であり，偽造筆跡に対しても有効であることが確かめられた．
6. 「奈良永遠」以外の漢字文字列に対しての照合性能を評価するため，漢字文字列の選択基準を満たす「大和郡山」というデータを被験者 20 人から収集し，筆者照合実験を行った．その結果，両識別器の EER は 2.47% ~2.71% ，FRR を 0% としたときの FAR は 6.55% ~10.20% であった．この結果と「奈良永遠」の結果を総合すると，提案手法の照合性能は，EER 2% 程度，FRR を 0% としたとき FAR 10% 以下であり，目標とした数値を達成できた．
7. 以上の結果を総合すると，想定した，システムが設定した共通の漢字文字列をタブレットよりオンラインで入力する筆者照合システムは，以下のようない特徴を持つ．
 - (a) 利用者が 10 個程度の参照筆跡を登録することで，文字列ごとに照合に有効な全利用者共通の特性値が選択できる．

- (b) 使用する文字列の選択基準を示すことで、4文字中2~3文字で画数の統一が可能となり、提案手法が有効に働き、従来の日本語署名を用いたオンライン署名照合と同程度の性能を示す。
- (c) 静的特性値と動的特性値を総合的に用いているため、訓練された偽造筆跡に対しても有効に機能する。

今後の課題としては、提案手法に適した文字の画数や字画の形状などについて検討し、本論文で示した漢字文字列の選択基準をさらに明確にすることがあげられる。

さらに、本研究で得られた知見や手法を用いたより一般的な漢字文字列を利用した筆者照合システム、例えば登録漢字文字列と入力漢字文字列が異なっても同一部首を利用することで筆者照合を行うなど、の提案と構築を目指して行く。

謝辞

本研究を遂行するにあたり，終始懇切丁寧なご指導とご助言を賜りました奈良先端科学技術大学院大学情報科学研究科教授 木戸出正繼先生に心より深謝いたします。先生の暖かい励ましとご助言がなければ，この論文は完成しなかったと思います。本当にありがとうございました。

本論文をまとめるにあたり，有益なご助言とご指導を賜りました奈良先端科学技術大学院大学情報科学研究科教授 千原國宏先生，三重大学工学研究科教授 木村文隆先生，奈良先端科学技術大学院大学情報科学研究科助教 波部齊先生に心より感謝いたします。

本研究を進めるにあたり，奈良先端科学技術大学院大学情報科学研究科知能情報処理学講座において，有益なご討論とご助言を頂いた准教授 浮田宗伯先生，関西学院大学理工学部教授 河野恭之先生，大阪市立大学工学部講師 上野敦志先生に感謝いたします。

本研究を進めるにあたり，熱心にご討論頂くとともに，データ収集に協力して下さった知能情報処理学講座の卒業生の皆様，また，お世話になりました歴代秘書の皆様に感謝いたします。

本論文は，筆者が奈良工業高等専門学校で行っていた研究を奈良先端科学技術大学院大学情報科学研究科の博士後期課程に社会人入学し発展させまとめたものである。この間，先にお礼を述べさせて頂いた方々以外に多くの人々に支えられ助けて頂きました。

長年に亘り研究活動のご指導とご助言を賜りました大阪大学名誉教授 豊田順一先生に深く感謝いたします。

奈良先端科学技術大学院大学博士後期課程への社会人入学を勧めて下さった前

奈良工業高等専門学校長 一岡芳樹先生，研究活動にご理解とご鞭撻を頂いた奈良工業高等専門学校長 冷水佐壽先生に感謝いたします。

筆者がパターン情報処理の分野に進む切っ掛けを与えて下さった奈良工業高等専門学校名誉教授 上田勝彦先生に深く感謝いたします。

奈良工業高等専門学校での教育と研究活動を暖かく支えて頂くとともに，数々の有益なご教示を頂いた奈良工業高等専門学校名誉教授 木村伊一先生，同校名誉教授 成田紘一先生に心より感謝いたします。

奈良先端科学技術大学院大学での研究活動にご理解とご支援を頂いた，奈良工業高等専門学校電気工学科教職員ならびに退職された教職員の皆様に感謝いたします。

データ収集にご協力頂いた，奈良工業高等専門学校の在學生と卒業生ならびに事務職員の皆様に感謝いたします。

最後に，今まで私を支えてくれた妻と両親に感謝の意を捧げる。

参考文献

- [1] 小特集バイオメトリクスセキュリティ , 電子情報通信学会誌, Vol. 89, No. 1, pp. 26–55 , 2006.
- [2] サインの筆跡携帯で送受信 , 日経産業新聞 , 2007.10.24.
- [3] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification –The State of the Art–", *Pattern Recognition*, Vol. 22, No. 2, pp. 107–131, 1989.
- [4] F. Leclerc and R. Plamondon, "Automatic Signature Verification: The State of the Art–1989-1993", *Int. Journal of Pattern Recognition and Artificial Intelligence*, Vol. 8, No. 3, pp. 643–660, 1994.
- [5] W. Nelson, W. Turin and T. Hastie, "Statistical Methods for On-line Signature Verification", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 8, No. 3, pp. 749–770, 1994.
- [6] L. Yang, B. K. Widjaja and R. Prasad, "Application of Hidden Markov Models for On-line Signature Verification", *Pattern Recognition*, Vol. 28, No. 2, pp. 161–170, 1995.
- [7] L. L. Lee, T. Berger, and E. Aviczer, "Reliable On-Line Human Signature Verification Systems", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 18, No. 6, pp. 643–647, 1996.
- [8] V. S. Nalwa, "Automatic On-line Signature Verification", *Proceedings of The IEEE*, Vol. 85, No. 2, pp. 215–239, 1997.
- [9] Q. Wu, I. Jou and S. Lee, "On-line Signature Verification Using LPC Cepstrum and Neural Networks", *IEEE Trans. on Systems, Man, and*

- Cybernetics*, Vol. 27, No. 1, pp. 148–153, 1997.
- [10] R. S. Kashi, J. Hu, W. L. Nelson and W. Turin, "On-line Handwritten Signature Verification Using Hidden Markov Model Features", *Proc. of 4th International Conference on Document Analysis and Recognition*, pp. 253–257, 1997.
 - [11] J. G. A. Dolfig, E. H. L. Aarts and J. J. G. M. van Oosterhout, "On-line Signature Verification with Hidden Markov Models", *Proc. of International Conference on Pattern Recognition*, Vol. 2, pp. 1309–1312, 1998.
 - [12] A. K. Jain, F. D. Griess, and S. D. Connell, "On-line Handwritten Signature Verification", *Pattern Recognition*, Vol. 35, pp. 2963–2972, 2002.
 - [13] H. Heng and C. Wah, "On-line Handwritten Signature Verification Using a New Extreme Points Warping Technique", *Pattern Recognition Letters*, Vol. 24, pp. 2943–2951, 2003.
 - [14] L. Bovino, S. Impedovo, G. Pirlo and L. Sarcinella, "Multi-Expert Verification of Hand-Written Signatures", *Proc. of 7th International Conference on Document Analysis and Recognition*, pp. 932–936, 2003.
 - [15] J. Lee, H. Yoon, J. Soh, B. T. Chung and Y. K. Chung, "Using Geometric Extrema for Segment-to-segment Characteristics Comparison in Online Signature Verification", *Pattern Recognition*, Vol. 37, pp. 93–103, 2004.
 - [16] A. Kholmatov and B. Yanikoglu, "Identity Authentication Using Improved Online Signature Verification Method", *Pattern Recognition Letters*, Vol. 26, pp. 2400–2408, 2005.
 - [17] J. Fierrez, J. Ortega-Garcia, D Ramos and J. Gonzalez-Rodriguez, "HMM-based Online Signature Verification: Feature Extraction and Signature Modeling", *Pattern Recognition Letters*, Vol. 28, pp. 2325–2334, 2007.
 - [18] L. Nanni and A. Lumini, "A Novel Local On-line Signature Verification System", *Pattern Recognition Letters*, Vol. 29, pp. 559–568, 2008.
 - [19] Y. Sato and K. Kogure, "Online Signature Verification Based on Shape, Motion, and Writing Pressure", *Proc. of 6th International Conference*

- on Pattern Recognition*, pp. 823–826, 1982.
- [20] 田口英郎, 桐山公一, 田中永二, 藤井克彦, ”ペンの動きに着目したオンライン署名識別法”, 電子情報通信学会論文誌 D, Vol. J71-D, No. 5, pp. 830–840, 1988.
 - [21] M. Yoshimura, S. Matsuda, and I. Yoshimura, ”On-line Signature Verification Incorporating the Direction of Pen Movement”, *IEICE Trans.*, Vol. E74, No. 7, pp. 2083–2092, 1991.
 - [22] 李永鎬, 金太均, 安居院猛, 中嶋正之, ”木マッチングによる英字署名識別に関する研究”, 電子情報通信学会論文誌 D- , Vol. J75-D- , No. 1, pp. 31–38, 1992.
 - [23] 金長吉, 渡辺正晴, 川嶋稔夫, 青木由直, ”非公開パラメータに基づく署名照合”, 電子情報通信学会論文誌 D- , Vol. J75-D- , No. 1, pp. 121–127, 1992.
 - [24] P. Zhao, A. Higashi, and Y. Sato, ”On-line Signature Verification by Adaptively Weighted DP Matching”, *IEICE Trans. Inf. & Syst.*, Vol. E79-D, No. 5, pp. 535–541, 1996.
 - [25] 楊旭華, 古橋武, 小幡賢三, 内川嘉樹, ”局所改善型遺伝的アルゴリズムを用いたサインの個人性特徴発見”, 電気学会論文誌 C, Vol. 116-C, No. 5, pp. 548–555, 1996.
 - [26] 山崎恭, 小松尚久, ”カテゴリー化された筆跡情報に基づく個人性抽出手法”, 電子情報通信学会論文誌 D- , Vol. J79-D- , No. 8, pp. 1335–1346, 1996.
 - [27] 菊池真美, 赤松則男, ”高速筆記者のための高感度筆圧ペンの試作と筆者認識実験”, 電子情報通信学会論文誌 D- , Vol. J83-D- , No. 8, pp. 1763–1772, 2000.
 - [28] 山中晋爾, 浜本隆之, 半谷精一郎, ”ペンの傾きを利用した署名照合方式の改良”, 電子情報通信学会技術研究報告, ISEC2000-18, pp. 65–72, 2000.
 - [29] Y. Komiya, T. Ohishi, and T. Matsumoto, ”A Pen Input On-line Signature Verifier Integrating Position, Pressure and Inclination Trajectories”, *IEICE Trans. Inf. & Syst.*, Vol. E84-D, No. 7, pp. 833–838, 2001.
 - [30] M. Kondo, D. Muramatsu, M. Sasaki, and T. Matsumoto, ”Nonlinear

- Separation of Signature Trajectories for On-line Personal Authentication”, *Proc. of International Conference on Acoustics, Speech, and Signal Processing 2003*, pp. II-349–II352, 2003.
- [31] 中西功, 西口直登, 伊藤良生, 副井裕, ”DWT によるサブバンド分解と適応信号処理を用いたオンライン署名照合”, *電子情報通信学会論文誌 A*, Vol. J87-A, No. 6, pp. 805–815, 2004.
- [32] H. Kameya, S. Mori, R. Oka, ”A Segmentation-free Biometric Writer Verification Method Based on Continuous Dynamic Programming”, *Pattern Recognition Letters*, Vol. 27, pp. 567–577, 2006.
- [33] 吉村ミツ, 木村文隆, ”手書き文字の個性が現れる特性の機械的計測化とその解析”, *電子通信学会論文誌 D*, Vol. J63-D, No. 9, pp. 795–802, 1980.
- [34] 尺長健, 金子博, 淀川英司, ”2 次統計量の線分スペクトル分解による手書き文字の筆者認識”, *電子通信学会論文誌 D*, Vol. J67-D, No. 7, pp. 776–783, 1984.
- [35] S. Srihari, S. Cha, H. Arora and S. Lee, ”Individuality of Handwriting”, *Journal of Forensic Sciences*, vol. 47, No. 4, pp. 1-17, 2002.
- [36] 佐藤剛, 宮川道夫, 石渡正倫, ”手書き文字の直線性と曲線性に着目した筆者識別の試み”, *電気学会論文誌 C*, vol. 120-C, no. 3, pp. 397-403, 2000.
- [37] 梅田三千雄, 三好健生, 三崎揮市, ”自己想起型ニューラルネットワークによる筆者識別と照合”, *電気学会論文誌 C*, vol. 122-C, no. 11, pp. 1869-1875, 2002.
- [38] A. Shekhawat, S. Parulekar and S. Srihari, ”Individuality Studies for Online Handwriting”, *Proc. of the 11th Conference of the International Graphonomics Society (IGS2003)*, pp. 266–269, 2003.
- [39] 近藤弘文, 宮崎葵光, 山口操風, *小学校実践書写指導*, 東京書籍, 1990.
- [40] 河口至商, *多変量解析入門*, 森北出版, 1978.
- [41] 吉田公一, *文書鑑定的基础と実際*, 立花書房, 1983.
- [42] 高澤則美, ”筆跡鑑定”, *科学警察研究所報告法科学編*, Vol. 51, No. 2, pp. 43–53, 1998.

- [43] 石村貞夫, 分散分析のはなし, 東京図書, 1992.
- [44] Ed. by A. Jain, R. Bolle, and S. Pankanti, BIOMETRICS Personal Identification in Networked Society, Kluwer Academic Publishers, 1999.

研究業績

(関連論文)

[学術論文]

- [1] 中村善一，豊田順一，“書写技能に基づく筆跡に現れる個人性の抽出”，電子情報通信学会論文誌 D- ，Vol. J77-D- ，No. 3，pp.510-518，1994.（第2章と関連）
- [2] 中村善一，木戸出正継，“筆跡鑑定の知見に基づく特性値を用いたオンライン筆者照合”，システム制御情報学会論文誌，Vol.22，No.1，pp.37-47，2009.（第5章と関連）

[国際会議]

- [1] Yoshikazu Nakamura, Masatsugu Kidode, "Individuality Analysis of Online Kanji Handwriting", Proceedings of Eighth International Conference on Document Analysis and Recognition, pp.620-624, 2005.（第3章と関連）
- [2] Yoshikazu Nakamura, Masatsugu Kidode, "Online Writer Verification Using Kanji Handwriting", Proceedings of International Workshop on Multimedia Content Representation, Classification and Security, pp.207-214, 2006.（第4章と関連）

[国内発表]

- [1] 中村善一，豊田順一，“書写技能に基づく筆跡の個人性の計測”，1992年電子情報通信学会秋季大会，D-319，1992.
- [2] 中村善一，木戸出正継，“筆跡鑑定の知見に基づく漢字文字列からの個人性抽出”，電子情報通信学会技術研究報告 PRMU2004-23，pp.65-70，2004.
- [3] 中村善一，木戸出正継，“筆跡鑑定の知見に基づく特性値を用いたオンライン筆者照合”，第50回システム制御情報学会研究発表講演会講演論文集，8T1-1，pp.635-636，2006.

[その他]

- [1] 中村善一，“書写技能に基づく特性値を用いた筆者照合”，奈良工業高等専門学校研究紀要第29号，pp. 27-31，1994.