

修 士 論 文

組み合わせ分割多言語署名照合の
高精度化に関する研究

平成 29 年度修了

三重大学大学院工学研究科

博士前期課程 情報工学専攻

松田 圭悟

はじめに

情報化社会におけるセキュリティ対策に必要な個人認証が広く用いられている。個人認証を行う一手法に、バイオメトリクスがある。バイオメトリクスの中でも署名照合は、署名行為が社会的に受け入れられておりデータ取得に対する拒絶感や抵抗感が少ないこと、行動的特徴であるため本人の意思確認を伴う利点があり、数多くの研究が行われている。

署名照合の方式にはオンライン署名照合とオフライン署名照合がある。オンライン署名照合は、タブレットなどを用いて取得したペンの座標情報、筆速、筆圧を利用する方法である。オフライン署名照合は、すでに書かれた署名の筆跡を利用する方法である。タブレット端末の普及に伴い、オンライン署名照合の利用環境が整ったこと、筆速、筆圧などより多くの情報を取得、利用して高精度な照合が期待できることからオンライン署名照合が実用化されつつある。オンライン署名照合に関する研究はこれまで数多く行われてきたが、そのほとんどが筆跡を時間関数と捉えて照合を行う手法であり、登録署名数の削減、さらなる照合精度の向上という課題が残されている。

これらの問題を解決するため、先行研究では多言語署名に対応した組み合わせ分割照合を提案した。組み合わせ照合は、筆記者があらかじめ登録する登録用真筆署名との照合スコアを算出する一段階目照合と、その算出したスコアを組み合わせ、識別器で真偽判定する二段階目照合から構成される。また分割照合は、登録署名が少数の場合に生じる照合性能の低下を防ぐために、フルネーム署名を分割して照合に用いる手法である。上平らは、組み合わせ分割照合法における一段階目照合で、オフライン特徴とオンライン特徴に基づく照合手法を提案し、少数の登録署名を用いた時の照合精度を改善した。オフライン特徴による照合では、フルネームの署名画像と、フルネーム署名画像を姓名で分割した姓画像との名画像との計3個からそれぞれ特徴ベクトルを抽出し、各特徴ベクトルごとにマハラノビス距離を算出する。オンライン特徴による照合では、フルネーム署名の時系列データと、フルネーム署名の時系列データを姓名で分割したデータ計3個にDPマッチングを適用する。そして二段階目照合では、3つのマハラノビス距離と3つのDPマッチングにより得られた相違度を合わせて6次元の特徴ベクトルとし、これをサポートベクターマシン（Support Vector Machine, 以下 SVM）に入力し最終的な照合を行う。しかしこの

手法には、署名分割方法が日本人署名の特徴である姓名に依存している問題がある。これを改善すべく尾木らによって提案された手法では、欧米で用いられる筆記体の署名も照合できるように画像の分割方法を重心分割に変更した。これにより中国人、オランダ人署名において高い精度で照合できることが明らかになった。しかし、個別言語ごとに性能評価を行うため、あらかじめ入力署名の言語が既知でなければならないという制約、照合精度の向上の課題、二段階目照合の識別器学習に収集コストを要する訓練偽筆が必要となる制約があった。

本研究では、組み合わせ分割照合のこれらの課題に対処した。

まず、組み合わせ分割照合法を真の多言語署名照合に適用できるようにするために、SigComp2011, SigWiComp2013 のデータセットを統合して学習と評価に用いるように改良を行った。これにより学習データが多言語化され、どの言語の署名に対しても照合することが可能になった。

次に、照合精度の向上のために、一段階目照合に新たに 2 つのオフライン署名照合手法を追加した。1 つ目は、筆圧を色相に反映したカラー署名画像から RGB 各チャンネルの署名画像を生成し、各署名から特徴を抽出しマハラノビス距離を算出する手法である。2 つ目は、筆圧を輝度値に反映したグレースケール署名画像から、重心で水平・垂直に分割した 4 つの領域の輝度ヒストグラムを抽出し、登録署名とのヒストグラム類似度を算出する手法である。提案手法の有効性を評価するため、等価誤り率 (Equal Error Rate, 以下 EER) を用いて評価実験を行った。EER とは、本人拒否率 (False Rejection Rate, 以下 FRR) と他人受入率 (False Acceptance Rate, 以下 FAR) が等しくなるようしきい値を定めたときの全体のエラー率である。実験の結果、提案した 2 手法を導入し二段階目照合の識別器にランダムフォレスト (Random Forests, 以下 RF) を用いることにより EER を 3.35% に低減できた。

最後に、二段階目照合の識別器の偽筆クラスの学習サンプルに、収集コストを要する訓練偽筆を使用せず、第三者の登録署名を用いるランダム偽筆学習を提案した。多言語署名から構成される SigComp データセットを用いた評価実験の結果、訓練偽筆を学習する場合に比べて照合性能は低下するものの、真筆のみの学習より高い精度で照合可能であることが明らかになった。

目次

はじめに	i
第 1 章 緒言	1
1.1 研究の背景	1
1.2 署名照合	2
1.2.1 偽筆	2
1.2.2 オフライン情報とオンライン情報	3
1.2.3 従来研究	3
1.3 組み合わせ分割照合	5
1.4 偽筆なし学習	6
1.5 研究の目的	6
1.6 本論文の構成	7
第 2 章 提案手法	8
2.1 照合手順	8
2.2 時系列点座標データの取得	9
2.3 画像生成手法	10
2.3.1 グレースケール画像生成	10
2.3.2 カラー画像生成	12
2.4 濃度こう配特徴を用いる照合手法	14
2.4.1 重心を用いた画像分割	14
2.4.2 特徴抽出	14
2.4.3 照合	17
2.5 輝度ヒストグラムを用いる照合手法	18
2.5.1 特徴抽出	18
2.5.2 照合	18
2.6 DP マッチングを用いる署名照合	20

2.6.1	重心を用いた署名データ分割	20
2.6.2	データの正規化	20
2.6.3	DP マッチング	21
2.6.4	照合	21
2.7	組み合わせ照合	22
2.7.1	SVM (Support Vector Machine)	22
2.7.2	RF (Random Forests)	24
2.7.3	ランダム偽筆学習	26
第 3 章	実験	27
3.1	実験データ	27
3.2	EER (Equal Error Rate)	28
3.3	組み合わせ照合実験	29
3.3.1	実験条件	29
3.3.2	実験結果	29
3.3.3	考察	31
3.4	ランダム偽筆を用いた照合実験	32
3.4.1	実験条件	32
3.4.2	実験結果	33
3.4.3	考察	35
第 4 章	結言	41
4.1	まとめと今後の課題	41
付録 A	ランダム偽筆のサンプリング枚数における性能調査	43
A.1	実験条件	43
A.2	実験結果	43
付録 B	署名データリスト	45
付録 C	プログラムソースリスト	46
付録 D	発表資料	48
謝辞	51

第 1 章

緒言

1.1 研究の背景

情報化社会の発展により，電子商取引が一般化している．一方で，クレジットカード番号やパスワードの盗用，忘却が問題となっており，セキュリティ対策の強化が重要な課題である．セキュリティ対策には個人認証が広く用いられている．個人認証には，(1) パスワードや暗証番号といった記憶を用いた認証，(2) IC カードや印鑑など所有物を用いた認証，(3) 指紋や虹彩など身体的特徴を用いた認証，(4) 署名やまばたきといった行動的特徴を用いた認証がある．

(1) と (2) は，紛失・忘却により本人であるにも関わらず認証不能となる，盗難によって第三者に不正利用される危険性がある．これらに対し，(3) と (4) はバイオメトリクス [1] と呼ばれ，紛失・忘却・盗難のリスクが非常に小さく，他人による代行が困難であるため注目されている．主なバイオメトリクスとその特徴を表 1.1 に示す [2]．ここで，普遍性，唯一性，永続性とはそれぞれ，誰もが持っている特徴であること，本人しか持っていない特徴であること，時間経過によって特徴が変化しないことを意味する．また，本人拒否率，他人受入率とはそれぞれ，本人をなりすましと判別してしまう割合，なりすましを本人と判別してしまう割合である．表 1.1 から，指紋を用いた認証が最適に思われるが，犯罪捜査に利用されているため，登録において心理的な抵抗があると言える [3]．

表 1.1: バイオメトリクスの比較 [2]

情報	特徴量	普遍性	唯一性	永続性	コスト	本人拒否率 (%)	他人受入率 (%)	データ量 (Byte)
指紋	手指の指紋特等量	◎	◎	◎	◎	1.0	0.01	250
掌形	手の大きさ・長さ・比率	◎	○	○	△	0.1	0.1	10
顔	顔輪郭・目・鼻の形・配置	◎	△	△	○	5	5	2,000
静脈	手指の静脈特徴点	◎	○	○	△	1.0	0.01	500
声紋	話者の音声特徴	◎	△	△	◎	10	10	1,500
署名	字体・書き順・筆速	◎	△	△	○	5	5	1,000

[普遍性, 唯一性, 永続性] △ < ○ < ◎ の順で高い, [コスト] △ > ○ > ◎ の順で低い

バイオメトリクスはそれぞれに一長一短があり, 要求されるセキュリティレベルや利用方法に応じた選択が必要となる. 本研究で取り上げる署名照合には, 次のような特徴がある.

- 本人確認のための署名は, すでに社会的に受け入れられているため, 他のバイオメトリクスと比べ利用者の受容性が高い.
- 行動的特徴による個人認証であるため, 本人の意志確認を伴う.
- 永続性の高い身体的特徴の登録データとは違い, 流出や漏洩が生じても登録データ変更が可能である.
- 電子文書への貼り付けが可能である.

このような特徴を持つ署名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 [4].

1.2 署名照合

署名照合とは, 本人確認を行う際に署名を求め, その署名が本人によって書かれたものか, あるいは第三者がなりすまして書いたものかを判別する作業である.

1.2.1 偽筆

署名照合の精度評価を行うには本人の筆記した署名以外に, 偽筆と呼ばれる署名が必要となる. 偽筆は本人以外の人物が本人になりすまして筆記した署名である. 偽筆は以下の4つに分類することができる [5].

- ランダム偽筆 (Random Forgery) 本人署名に関する情報が与えられない状況で筆記した署名。形状も任意（ランダム）の偽筆である。一般には，取得が容易な他人の登録署名をランダム偽筆として用いることが多い [6]。
- 単純偽筆 (Simple Forgery) なりすまし者が署名の筆記内容を知っているが，署名の形状は知らない状況で筆記した署名。文字認識を行えば，同じ署名として認められる可能性はあるが，署名形状と名前に関連がない場合はランダム偽筆と同じである。
- 模写偽筆 (Simulated Forgery) 本人署名を上からなぞって作成した偽筆である。署名形状は本人署名に非常に似ている。従って，オフライン署名認証では判別の難しい署名である。
- 訓練偽筆 (Skilled Forgery) 本人署名の形状や書き順，筆記方法等の情報を入手したなりすまし者がオンライン情報を含め真似るための練習を行った後に筆記する署名であり，オンライン署名認証においては最も脅威となる偽筆である。

1.2.2 オフライン情報とオンライン情報

署名照合方式には，オフライン署名照合とオンライン署名照合がある。オフライン署名照合とは，すでに書かれた署名から得られる筆跡の形態情報を利用する方法である。一方，オンライン署名照合とは，専用機器より取得できる筆順，筆圧，筆速といった署名時の筆記運動情報を利用する方法である。この場合，専用機器としてタブレットや電子ペンが必要になるが，タブレット PC，PDA (Personal Digital Assistant)，スマートフォンなど，ペン入力可能な機器の普及が進み，利用環境が整いつつある。また，オフライン情報よりも得られる情報が多く，筆速や筆圧，ペンの傾きは，他人が容易に真似ることができないと考えられる。現実においてどちらの情報が利用できるかは状況次第であるが，オンライン情報の取得環境が整いつつあることと，得られる情報の多さより，本研究ではオンライン署名照合を研究対象とする。

1.2.3 従来研究

オンライン署名照合に関して，署名文化のある欧米で研究が盛んに行われている [5, 7]。しかしそのほとんどが，筆跡の動的情報であるペンの座標，筆速，筆圧，傾きの時系列データに対して DP (Dynamic Programming) マッチングを行う手法 [9]，セグメントごとの曲率や向きなどの局所的特徴の時系列を HMM (Hidden Markov Model) により照合を行う手法 [10] のように，筆跡を時間関数と捉えて照合するものが主流である [4]。

最近では，ユーザ共通の Fusion モデルにより複数の距離を組み合わせることで照合

を行う手法 [11] が提案されている。この手法では、アルファベットベースの署名を含む公開オンライン署名データベースの MYCT[12] と BIOMET[13] を用いて性能評価を行っている。MCYT をデータセットとして性能評価を行った場合、3.73%の EER を達成したと報告されている。また、BIOMET をデータセットとして性能評価を行った場合、4.55%の EER を達成したと報告されている。文献 [17] では、最長共通部分列マッチングを用いた署名の回転角度表現に基づくオンライン署名照合手法を提案している。公開オンラインデータセットである SVC2004[19] と SUSIG[18] を用いた結果、SVC2004 においては 0.5%、SUSIG において 5.33%の EER を達成したと報告されている。文献 [20] では、ルジャンドル多項式を用いて関数近似を行い、登録署名との相違度を特徴量とし SVM で識別する手法が提案されている。これは、本研究と同じ SigComp2011[7] の公開データセットを用いており、中国語署名に対しては 8.93%、オランダ語署名に対しては 5.44%の EER を達成したと報告されている。さらに、SigComp2011 のコンペティションの結果、Namirial 社が販売している統計的経験モデルを用いたオンライン署名照合ツール xyzmo[21] が中国語署名に対しては 6.83%、オランダ語署名に対しては 3.65%の EER を達成したと報告されており、コンペティションベストである。

また、日本語署名を対象にした研究も数多く発表されている。文献 [14] では、筆跡の動的情報であるペンの座標、筆圧、傾きの時系列データを用いる手法が提案されている。この手法では 8 氏名から採取した 293 個の真筆、540 個の偽筆を実験データとし、登録署名数を 7~11 個として性能評価を行った結果、2.4%の FRR、1.3%の FAR を達成したと報告されている。文献 [15] では、動的情報の時間変化を離散ウェーブレット変換によりサブバンド分解した信号を特徴とし、適応信号処理技術を応用して照合する手法が提案されている。この手法では 4 氏名から採取した 98 個の真筆、200 個の偽筆を実験データとし、登録署名数を 5 個として性能評価を行った結果、約 5%の EER を達成したと報告されている。文献 [16] では、タブレットを必要とせず、高感度の電子ペンによって得られる 1 次元の筆圧情報に基づく署名照合手法が提案されている。この手法では、31 氏名から採取した 775 個の真筆と 1085 個の偽筆に対して個人認証実験を行い、11.1%の EER を達成したと報告されている。また、日本語署名の公開データセットである SigWiComp2013[8] のコンペティションの結果、日本語のオンライン署名に対し文献 [22] で提案された動的な時間伸縮法を用いることで 27.45%の EER を達成しておりコンペティションベストである。また、SigWiComp2013 を用いた研究として、文献 [23] では、話者照合でよく用いられる i-ベクトル空間モデルにおけるクラス内変動を用いた手法により 8.75%の EER を達成したと報告されている。

1.3 組み合わせ分割照合

署名照合では，入力署名と登録署名との相違度を 1 つ照合手法で算出し，しきい値によって処理して真筆か偽筆かを判断する手順が主流である．しかし，近年では，複数の照合手法から求められた特徴を組み合わせる研究が報告されている．文献 [28] では，タブレットに書かれたオンライン署名に対して，得られた時系列データを用いたオンライン署名照合と，タブレットから得られた署名画像を用いたオフライン照合を組み合わせることで照合精度の向上を実現した．文献 [29] では，複数の特徴から粒子群最適化 (PSO) を用いて特徴選択し照合する手法を提案し性能向上を実現した．また先行研究 [24] では，上平らはオフライン特徴とオンライン特徴に基づく組み合わせ分割照合法を提案し，少数の登録署名を用いた場合の照合精度を改善した．組み合わせ照合は以下のように二段階の照合を行い，真偽判定を行う．

1. 一段階目照合

各照合手法を用いて，入力署名と筆記者があらかじめ登録する登録用真筆署名とを照合して照合スコアを算出する．

2. 二段階目照合

一段階目照合で算出した評価値を組み合わせ，あらかじめ学習した識別器で，真筆か偽筆かを判断する．

また分割照合は，少数の登録署名の際に照合性能の低下を防ぐために，日本語の署名において姓名で分割した姓画像と名画像とフルネーム署名画像の 3 枚の署名を用いる手法である．

上平らは，組み合わせ分割照合法の一段階目識別手法として以下の手法を導入した．オフライン特徴による照合では，フルネームの署名画像とそれを姓と名に分割した姓画像と名画像の計 3 個からそれぞれ特徴ベクトルを抽出し，各特徴ベクトルごとに登録署名とのマハラノビス距離を算出する．オンライン特徴による照合では，フルネーム署名の時系列データとフルネーム署名の時系列データを姓名で分割したデータ計 3 個に DP マッチングを適用する．そして，3 つのマハラノビス距離と 3 つ DP マッチングにより得られた相違度を合わせて新たに 6 次元照合スコアベクトルとし，二段階目照合で識別器 (SVM) に入力し最終的な照合を行う．しかしこの手法の問題点は，署名分割方法が姓名分割であるため日本人署名に限定される点である．これを改善すべく尾木らによって提案された手法では，欧米で用いられる筆記体を含む複数の言語の署名を照合できるように画像の分割方法を重心分割に変更した [25]．これにより中国人，オランダ人署名を高い精度で照合できるようになった．

1.4 偽筆なし学習

先行研究 [25] では、第二段階照合の識別器の学習に、真筆および訓練偽筆それぞれの照合スコアベクトルを学習用データとして用いた。しかし、特定の筆記者に対する訓練偽筆署名の収集は時間とコストを要するため、一般に困難である。この問題への対策として、真筆署名に対して変動を加えて生成した署名で識別器を学習する手法 [26, 27] が提案されているが、筆記者内、外の適切な署名変動パラメータ設定が困難である問題がある。

本研究では、登録された第三者の真筆署名どうしを比較して求めた照合スコアをランダム偽筆照合のスコアベクトルと定義して、偽筆クラスの学習用データに用いる。しかし、全登録署名を第三者署名と比較すると、真筆に対して偽筆の数が大きくなり、偏った学習用データによる性能低下が生じる恐れがある。このようにクラス間のサンプル数に偏りがあるデータセットを用いた識別器の学習は *imbaranced learning* と呼ばれ、多数の手法が提案されている [30]。代表的なアプローチは、学習用データセットに対するアンダーサンプリング手法である。最も簡単なアンダーサンプリング手法はランダムサンプリングであるが、ランダムな要素を含むアンダーサンプリング手法は試行のたびに照合性能に変動が生じるため、安定性と結果に対する説明性が要求される認証用途には適さない。より安定かつ効果的なサンプリング手法として、クラスタリングに基づくサンプリング手法 [31], *informed undersampling* [32], 重複データのクリーニングに基づく手法 [33] などが提案されている。

1.5 研究の目的

本研究では、以下に示すとおり、従来手法に存在した 3 種類の課題それぞれに対して改良を行った。

まず、組み合わせ分割照合法を真の多言語署名照合に適用するために、SigComp2011, SigWiComp2013 のデータセットを統合して学習と評価に用いるように改良を行った。従来手法では、個別言語ごとに性能評価をしており、あらかじめ入力署名の言語が既知でなければならないという制約があった。本研究では、この改良により学習データが多言語化され、どの言語の署名に対しても照合することが可能になった。

次に、署名照合の性能向上のために、一段階目照合のオフライン照合を 2 手法提案した。先行研究では、筆圧を反映したグレースケール画像から特徴を抽出し、登録署名とのマハラノビス距離を求めるオフライン照合手法と、DP マッチングを用いるオンライン照合手法を組み合わせていた。本研究では、以下の 2 手法を新たに導入した。1 つ目は、筆圧を色相に反映したカラー署名画像から RGB 各チャネルの署名画像を生成し、各署名か

ら特徴を抽出しマハラノビス距離を算出する手法である。2つ目は、筆圧を輝度値、筆速を署名の線幅に反映させたグレースケール画像の輝度ヒストグラム類似度を用いる手法である。

最後に、二段階目照合の識別器の学習に訓練偽筆を要する課題に対して、第三者の真筆を用いるランダム偽筆学習を提案する。そして、学習用データセットにおいて多数クラスである偽筆クラスに対し、安定にサンプリングできる手法として **One-Class SVM** と **k-means** クラスタリングを用いたサンプリング手法を提案する。提案手法はクラスタリングを用いる点で [31] の手法と類似するが、提案手法は手法にランダムな要素を含まない点が特徴である。

1.6 本論文の構成

本論文の2章では、本研究における署名照合手法の流れと、各処理の詳細について述べる。さらに第3章では、提案手法の照合実験を実験署名データ、条件とともに結果と考察について述べる。最後に第4章で、本研究のまとめと今後の課題について述べる。

第 2 章

提案手法

本章では，多言語署名のための組み合わせ分割照合法について述べる．

2.1 照合手順

提案手法による署名照合処理の流れを図 2.1 に示す．タブレット PC を用いて入力署名の時系列点座標データを取得する．入力署名として座標値と筆圧の時系列データを受け取り，得られた時系列点座標データの筆圧を輝度値に，筆速を線幅に反映させたフルネームグレースケール署名画像を生成する．さらに，筆圧を色相に反映させたカラー署名画像も生成し，RGB 分離し各チャンネル画像を生成する．生成したそれぞれの画像の重心を求め，左右に分割して部分画像（左署名画像と右署名画像）を得る．同じく，フルネームの時系列点座標データを重心で分割し，部分データ（重心左側データと重心右側データ）を得る．そして，一段階目照合として以下の 4 手法を行う．

1. 生成したグレースケール画像から濃度こう配特徴を抽出し，登録署名とのマハラノビス距離を算出するオフライン照合手法
2. 時系列署名データに DP マッチングを適用して，登録署名との相違度を算出するオンライン照合手法
3. 生成した RGB チャンネル画像から濃度こう配特徴を抽出し，登録署名とのマハラノビス距離を算出するオフライン照合手法
4. 生成したグレースケール画像から輝度ヒストグラム特徴を抽出し，登録署名との類似度を算出するオフライン照合手法

そして各手法で算出した，12 個のマハラノビス距離と 3 つのヒストグラム類似度，3 つの相違度を合わせて 18 次元の特徴ベクトルとし，二段階目照合で識別器（SVM, RF）により識別する．この識別器の学習に用いられるのは登録署名と真筆を比較して求めた真筆照

合スコアと、登録署名と第三者が模倣して書いた訓練偽筆と比較して求めた偽筆照合スコアを学習に用いる．また，訓練偽筆照合スコアを用いない場合は，第三者の登録署名から総当たりで4つの手法で作成したランダム偽筆照合スコアを学習に用いる．各手法や処理の詳細については，続く節にて述べる．

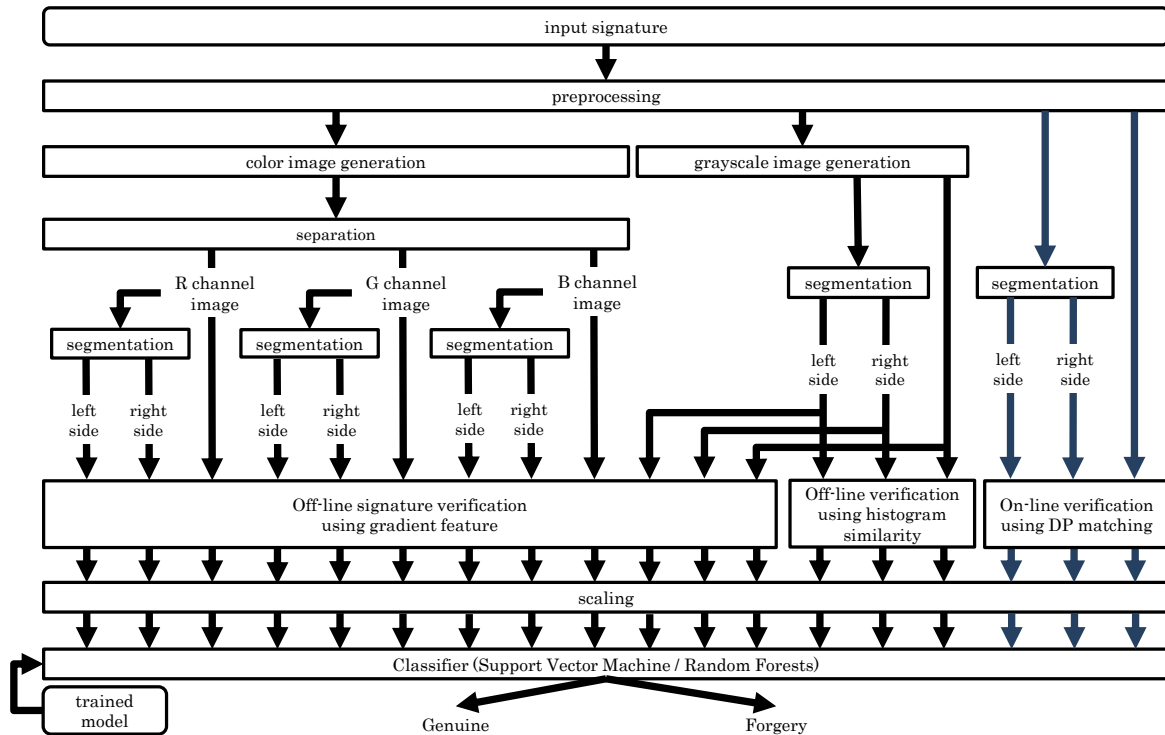


図 2.1: 提案手法の流れ

2.2 時系列点座標データの取得

署名の時系列データは，サンプリングレート 200Hz，解像度 2000line/cm の条件下で取得された．データ収集に使用された機器は WACOM Intuos3 である．取得されるデータは，タブレット上のペンの位置情報 x, y 座標データと奥行き（押し込み具合，筆圧に相当する） z 座標データである [7][8]．

2.3 画像生成手法

2.3.1 グレースケール画像生成

時系列点の座標と筆圧（ z 座標）、筆速 v を用いて、署名画像を生成する。筆圧の大きい点ほど輝度値を小さく、さらに筆速の大きい点ほど線幅が細い画像を生成する。

1. 任意の署名の i 番目のサンプル点の座標値を次式で正規化する。

$$x'_i = (x_i - x_{\min}) \frac{S}{W} \quad (2.1)$$

$$y'_i = (y_i - y_{\min}) \frac{S}{W} \quad (2.2)$$

ここで、 x'_i, y'_i はそれぞれ正規化後と正規化前における i 番目のサンプル点の x 座標、 x_{\min} は全サンプル点の x 座標の最小値、 W は筆記者の登録署名の平均横幅サイズ、 S は正規化後の登録署名の平均横幅サイズ（ $S = 700$ ）である。式 (2.2) は式 (2.1) の x を y に置き換えたものである。

2. 筆圧により輝度値を定義する。筆圧により輝度値を 0～255 に変換し筆圧が大きければ輝度値を小さく、筆圧が小さければ輝度値を大きくする

$$f(x_i, y_i) = 255 \times \frac{p_{\max} - p_i}{p_{\max} - p_{\min}} \quad (2.3)$$

3. 筆速を単位時間に対するサンプル二点間の距離を用いて算出する。

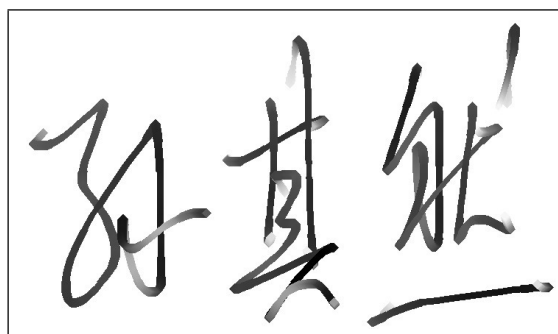
$$v_{xi} = \frac{\sqrt{(x_i - x_{i-1})^2}}{t_i - t_{i-1}}, v_{yi} = \frac{\sqrt{(y_i - y_{i-1})^2}}{t_i - t_{i-1}} \quad (2.4)$$

ここで、 t_i は i 番目のサンプル点の取得時刻である。

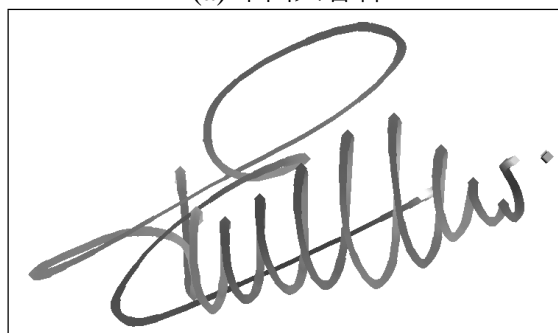
4. 線幅を以下の式によって定義する。

$$w_i = w_{\min} + (w_{\max} - w_{\min}) \times \left(1 - \frac{1}{1 + e^{-\alpha(v_i - v_{\text{ave}})}}\right) \quad (2.5)$$

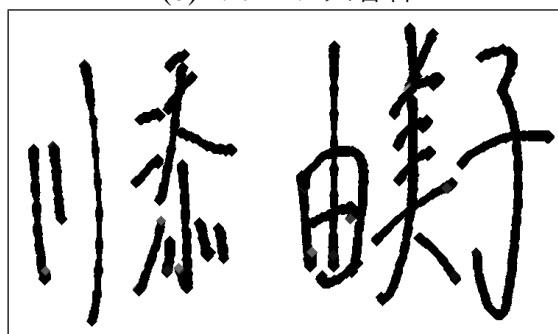
ここで $v_i = \sqrt{v_{xi}^2 + v_{yi}^2}$ は i 番目のサンプル点の筆速である。また、 $v_{\text{ave}} = \frac{v'_{\max} + v'_{\min}}{2}$ とし、 v'_{\max}, v'_{\min} は登録署名の筆速の最大値、最小値である。さらに、 $w_{\max} = 7$ 、 $w_{\min} = 2$ 、 $\alpha = \frac{5}{v_{\max} - v_{\min}}$ とし、 v_{\max}, v_{\min} は入力署名の筆速の最大値、最小値である。作成した署名画像の一例を図 2.2 に示す。



(a) 中国人署名



(b) オランダ人署名



(c) 日本人署名

図 2.2: 提案手法による生成署名

2.3.2 カラー画像生成

本研究では，新たに筆圧を色相に反映したカラー画像生成手法を提案する．筆圧情報を形状特徴に反映するためにカラー画像を生成し RGB チャンネルごとに分離して，グレースケール画像を得る．これにより，署名から筆圧情報に応じてストローク抽出を行い，筆圧情報を形状特徴に変換することができる．RGB チャンネルごとに分離するため，抽出する際の筆圧のしきい値は個別に設定する必要がある．以下に時系列点の座標と筆圧 z ，筆速 v を用いて，筆圧を色相に，筆速を線幅に変換するカラー署名画像を生成する処理を示す．

1. i 番目のサンプル点の色相 $h(x_i, y_i)$ を以下の式で定義する．

$$h(x_i, y_i) = C + R \times \frac{p_i - p_{\min}}{p_{\max} - p_{\min}} \quad (2.6)$$

ここで p_i は， i 番目の筆圧， p_{\min} と p_{\max} は登録署名の筆圧の最大値，最小値の平均である．今回， C は色相の初期値， R は色相範囲であり， $C = 180$ ， $R = 240$ とした．

2. i 番目のサンプル点の線幅を式 (2.5) で算出する．図 2.4 に作成したカラー画像を示す．
3. カラー画像を RGB 各チャンネル画像に分離する．図 2.5 に作成した各チャンネルの画像を示す．

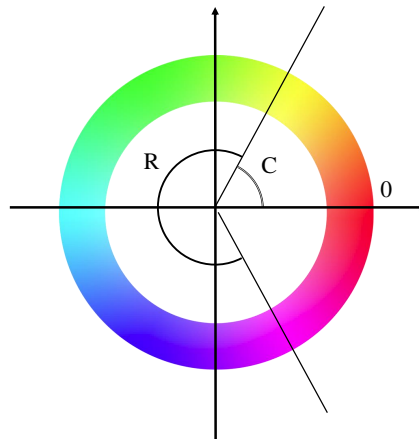


図 2.3: HSV 色空間

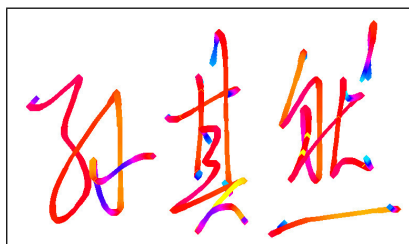
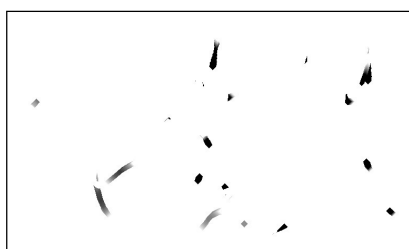
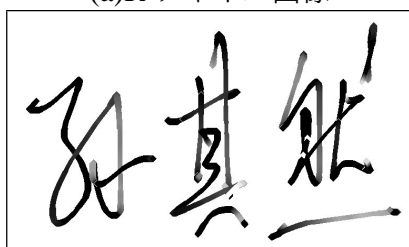


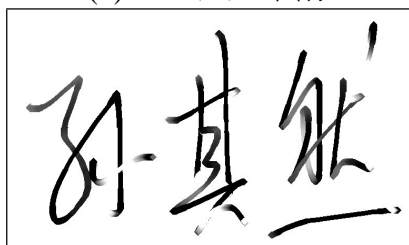
図 2.4: 作成カラー署名画像



(a)R チャンネル画像



(b)G チャンネル画像



(c)B チャンネル画像

図 2.5: RGB 分離画像

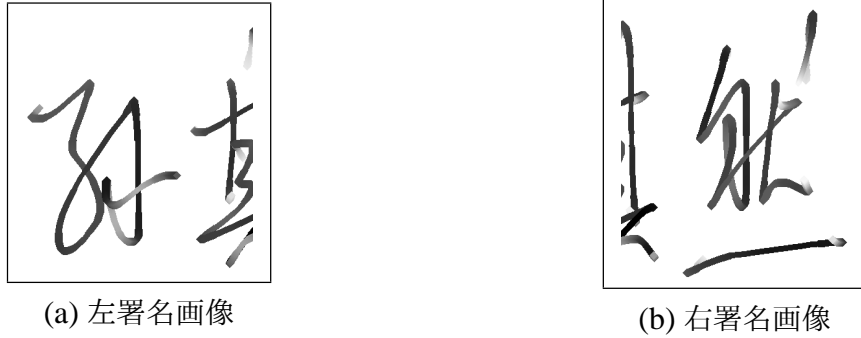


図 2.6: 分割後画像

2.4 濃度こう配特徴を用いる照合手法

濃度こう配特徴を用いる照合手法の流れを以下に示す．2.3.1 で作成したグレースケール画像及び，2.3.2 で作成したカラー画像から分離して求めた RGB 各チャネルのグレースケール画像を用いる．

2.4.1 重心を用いた画像分割

フルネームの署名画像をストロークの重心位置で，左署名画像と右署名画像に分割する．分割処理の手順は以下の通りである．

1. フルネーム署名画像をストロークの外接矩形で切り出す．
2. 次式を用いてフルネーム署名画像の重心 GC を求める．

$$GC = \left(\frac{\sum_i x_i f(x_i, y_i)}{\sum_i f(x_i, y_i)}, \frac{\sum_i y_i f(x_i, y_i)}{\sum_i f(x_i, y_i)} \right) \quad (2.7)$$

ここで， x_i ， y_i はそれぞれ画像の横軸縦軸の座標値， $f(x_i, y_i)$ は画素 (x_i, y_i) の濃度値である．

3. GC を通る垂直線で画像を左右に分割する．

図 2.2(a) の署名画像を重心位置で左右署名画像に分割した画像を図 2.6 に示す．

2.4.2 特徴抽出

特徴抽出処理では，生成した署名画像から濃度こう配特徴の抽出を行う．濃度こう配特徴とは，各画素における濃度こう配の方向別ヒストグラム特徴である [35][36]．本研究では，フルネーム画像と各分割後画像をブロック化し，576 次元の特徴ベクトルを抽出した．以下に濃度こう配特徴の処理手順を示す．

1. 入力画像（図 2.7 (a)）をストロークの外接矩形で切り出し，重心を求める．外接矩形を重心で 4 領域に分け，重心が出力画像（正規化後画像）の中心となるように 4 つの小領域ごとに位置・大きさを正規化する．正規化画像（図 2.7 (b)）のサイズは 792×352 である．
2. 正規化画像の，全画素に 2×2 の平均値フィルタ処理を 5 回行い，実数値を取る濃淡画像（図 2.7 (c)）を得る．
3. 濃淡画像の濃度値の平均が 0，分散が 1 となるよう画像を正準化する（図 2.7 (d)）．
4. 正準化後の画像に対して Roberts フィルタを適用し，各画素ごとにこう配の向きと強度を算出する．こう配の強度 $f(x, y)$ と向き $\theta(x, y)$ はそれぞれ次の式で求められる．

$$f(x, y) = \sqrt{(\Delta u)^2 + (\Delta v)^2} \quad (2.8)$$

$$\theta(x, y) = \tan^{-1} \frac{\Delta v}{\Delta u} \quad (2.9)$$

$$\Delta u = g(x+1, y+1) - g(x, y) \quad (2.10)$$

$$\Delta v = g(x+1, y) - g(x, y+1) \quad (2.11)$$

ここで， $g(x, y)$ は画素 (x, y) における画素値である．

5. 式 (2.9) から求めたこう配の向きを $\pi/16$ 刻みの 32 方向に量子化する．
6. 正規化した画像を横 17 ブロック \times 縦 7 ブロックに分割し，各ブロック内で量子化した方向別にエッジ強度の値を加算して局所ヒストグラムを得る．
7. 各ブロックにおいて，32 方向ヒストグラムのこう配強度を 1 次元加重フィルタ [1 4 6 4 1] によって平滑化し，水平右方向から 1 方向おき ($\pi/8$ 刻み) にサンプリングすることで向きを 16 方向に削減する．
8. 各方向ごとに， 17×7 ブロックのこう配強度を 2 次元加重フィルタによって平滑化し，奇数行かつ奇数列にあるブロックのこう配強度をサンプリングすることでブロック数を 9×4 に削減する．

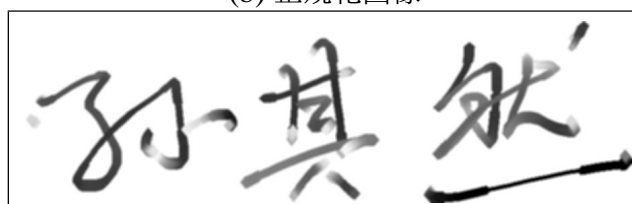
入力画像（図 2.7 (a)）に対し，抽出した濃度こう配を擬似カラー表現した画像を図 2.7 (e) に示す．この図では，こう配の向き，強度をそれぞれの色相，明度によって表現している．



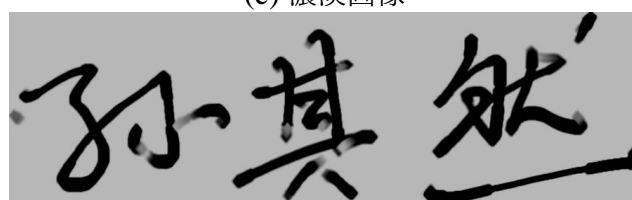
(a) 入力画像



(b) 正規化画像



(c) 濃淡画像



(d) 正準化画像



(e) こう配画像

図 2.7: 特徴抽出の流れ

2.4.3 照合

前述の特徴抽出処理により得られた特徴ベクトルに対して登録済みの真筆署名との照合を行う。照合には入力署名の特徴ベクトルと登録署名との正則化マハラノビス距離を用いる。マハラノビス距離は次式で定義される。

$$g_i(\mathbf{X}) = (\mathbf{X} - \mathbf{M}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{X} - \mathbf{M}_i) \quad (2.12)$$

ここで \mathbf{X} は入力署名の特徴ベクトル, \mathbf{M}_i は筆記者 i の平均特徴ベクトル, $\boldsymbol{\Sigma}_i$ は筆記者 i の併合級内共分散行列である。級内共分散行列 $\boldsymbol{\Sigma}_i$ のランクは (学習データ数 - クラス数) を超えない。特徴ベクトルの次元数に対して学習データが少ないと $\boldsymbol{\Sigma}_i$ が正則にならず逆行列が存在しない。そのため, $\boldsymbol{\Sigma}_i$ の代わりに次式で正則化された ${}_r\boldsymbol{\Sigma}_i$ を用いるものを正則化マハラノビス距離とする。

$${}_r\boldsymbol{\Sigma}_i = (1 - \alpha)\boldsymbol{\Sigma}_i + \alpha \frac{\text{trace}\boldsymbol{\Sigma}_w}{n} \mathbf{I} \quad (2.13)$$

ここで $\boldsymbol{\Sigma}_w$ は全筆記者の級内共分散行列の平均, $\text{trace}\boldsymbol{\Sigma}_w$ は行列 $\boldsymbol{\Sigma}_w$ の対角要素の和, \mathbf{I} は単位行列である。

2.5 輝度ヒストグラムを用いる照合手法

従来手法では局所領域の形状特徴を用いた照合を行っていたが、本研究では、署名全体の情報を活用した照合手法を新たに提案した。輝度ヒストグラムを用いる照合手法の流れを図 2.8 に表し、処理手順を以下に示す。2.3.1 で作成したグレースケール画像を用いて 2.4.1 で分割された 2 枚の分割署名とフルネーム署名に対し処理を行う。

2.5.1 特徴抽出

登録署名と入力署名に対してそれぞれヒストグラムを算出する。

1. グレースケール署名画像から式 (2.7) を用いて重心 GC を求める。
2. 重心 GC を通る垂直線及び水平線で署名画像を 4 分割する。
3. 分割した 4 つの分割署名画像から輝度ヒストグラム $H_i(x)$ をそれぞれ算出する。

2.5.2 照合

算出した登録署名と入力署名の輝度ヒストグラムの類似度を、次式のヒストグラムインターセクション D_H を用いて求める。

$$D_H = \sum_{i=1}^4 \sum_{x=0}^{255} \min(H_i(x), H'_i(x)) \quad (2.14)$$

ここで、 $H_i(x)$ と $H'_i(x)$ は入力署名、登録署名の各分割画像の輝度ヒストグラムである。図 2.8 に示すように、登録署名と真筆署名の類似度は高いためインターセクション値は高い、それに対し、登録署名と偽筆署名とのインターセクション値は低くなる。

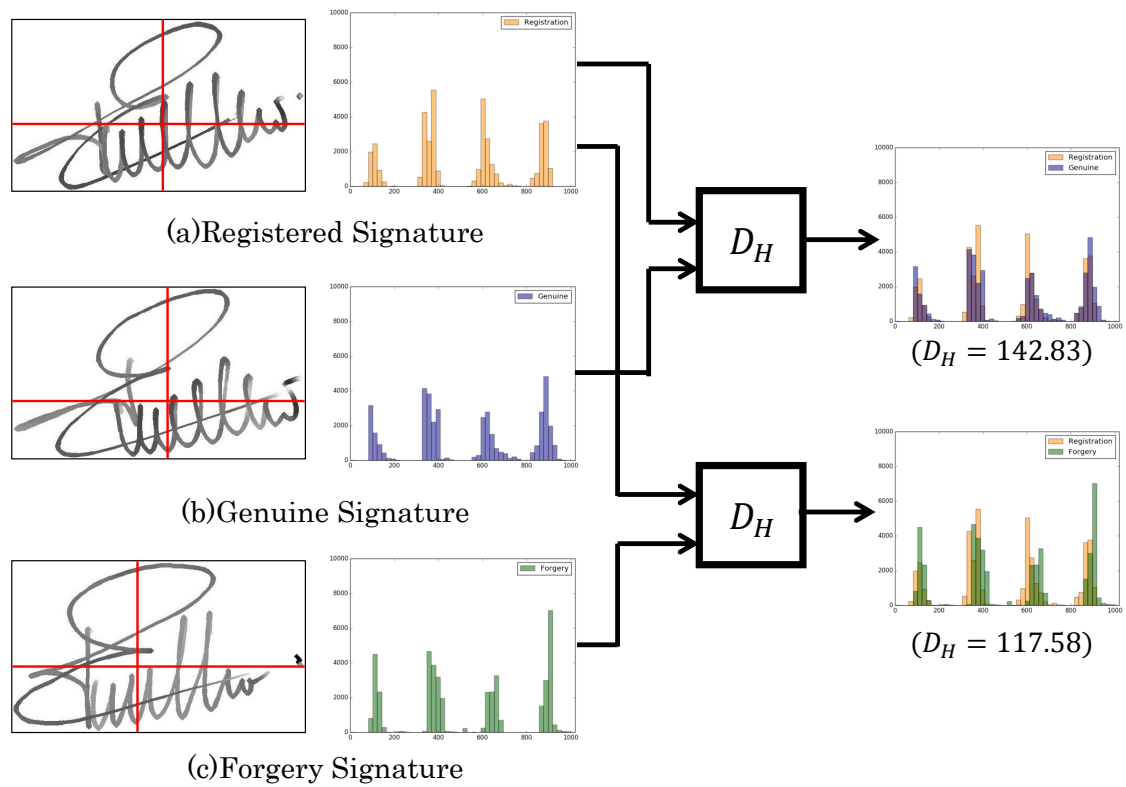


図 2.8: ヒストグラム類似度計算

2.6 DP マッチングを用いる署名照合

DP マッチングによる署名照合の流れを以下に示す.

2.6.1 重心を用いた署名データ分割

フルネームの署名データをストロークの重心位置で、左署名データと右署名データに分割する. 分割処理の手順は以下の通りである.

1. フルネーム署名座標を式 (2.15),(2.16) を用いて変換する.

$$x_i = (X_i - X_{\min}) \quad (2.15)$$

$$y_i = (Y_i - Y_{\min}) \quad (2.16)$$

ここで, $x_i \cdot X_i$ はそれぞれ変換後と変換前における i 番目のサンプル点の x 座標, $y_i \cdot Y_i$ はそれぞれ変換後と変換前における i 番目のサンプル点の y 座標, X_{\min} は全サンプル点の x 座標の最小値, Y_{\min} は全サンプル点の y 座標の最小値である.

2. 次式を用いてフルネーム署名データの重心 GC_D を求める.

$$GC_D = \left(\frac{\sum_i x_i p_i}{\sum_i p_i}, \frac{\sum_i y_i p_i}{\sum_i p_i} \right) \quad (2.17)$$

ここで, x_i, y_i はそれぞれ i 番目のサンプル点の署名データの座標値, p_i は i 番目のサンプル点の筆圧値である.

3. GC_D の横方向の座標値を基準値として, 各サンプル点の x 座標データが基準値以下ならそのサンプル点は左署名データ, 大きければ右署名データとして分割する.

2.6.2 データの正規化

同一人物であっても署名の大きさや書き始めの位置は常に一定ではない. これらの差異を緩和するため, DP マッチングによる相違度を計算する前にサイズの正規化を行う.

1. 学習データ, 評価データ共に文字外接枠での筆跡サイズと時間が最大値 100 となるよう線形変換する.
2. それぞれの重心点を算出し, 重心点を原点とする座標値に置き換える.

2.6.3 DP マッチング

DP マッチングとは動的計画法を用いるパターンマッチング手法であり、長さが異なる 2 つの系列の類似性を比較する方法 [37] である。

各系列中の 2 サンプル点間の相違度を

$$d(i, j) = \alpha \{ (x_i - x'_j)^2 + (y_i - y'_j)^2 \} + \beta (z_i - z'_j)^2 + \gamma (v_i - v'_j)^2 + \delta (t_i - t'_j)^2, \quad (2.18)$$

で定義する。ここで、 $x_i, y_i, z_i, v_i, t_i, (x'_j, y'_j, z'_j, v'_j, t'_j)$ はそれぞれ順に評価用 (学習用) データの i または j 番目の x, y, z, v 座標データと経過時間、 $\alpha, \beta, \gamma, \delta$ は重み係数である。各系列の始点間を $g(0, 0) = d(0, 0)$ とし、以下の漸化式で系列を終点までマッチングさせた時の相違度 $g(n, m)$ を計算する。ここで、 n, m はそれぞれの系列のサンプル点である。

$$g(i, j) = \min \{ g(i-1, j), g(i-1, j-1), g(i, j-1) \} + d(i, j) \quad (2.19)$$

2.6.4 照合

本研究では評価データ 1 個に対し全ての登録用データとの相違度を計算し、その中で最小の相違度を用いる最近傍法により照合を行う。

2.7 組み合わせ照合

2.7.1 SVM (Support Vector Machine)

SVM (Support Vector Machine) は、2 クラスの分類を行う学習機械の一種で、与えられた学習サンプルのなかで、サポートベクトルと呼ばれるクラス境界近傍に位置する学習サンプルと識別面との距離であるマージンを最大化するように分離超平面を構築し、クラス分類を行う。従来の分類手法と比べて、高い汎化能力を持ち、2 次の凸計画問題として定式化されるため、学習により最適解を得ることができる特徴を持つ。

SVM の識別関数は以下の式で表される。

$$g(\mathbf{x}) = \sum_{i=1}^d w_i x_i + b \quad (2.20)$$

ここで w_i は重みと呼ばれるパラメータで、ベクトル表示した \mathbf{w} を重みベクトルと呼ぶ。 b はバイアス項と呼ばれるパラメータである。この識別器の $g(\mathbf{x}) = 0$ を満たす点の集合（識別面）は、 $d - 1$ 次元の超平面となる。

図 2.9 に SVM の概念図を示す。SVM は超空間内で 2 クラス分離する超平面の内、マージンが最大となるような超平面を求め、 $\mathbf{w} \cdot \mathbf{x} + b = 0$ を最終的な分類決定境界とする。また $\mathbf{w} \cdot \mathbf{x} + b = 1$ および $\mathbf{w} \cdot \mathbf{x} + b = -1$ 上の訓練データをサポートベクトルと呼ぶ。

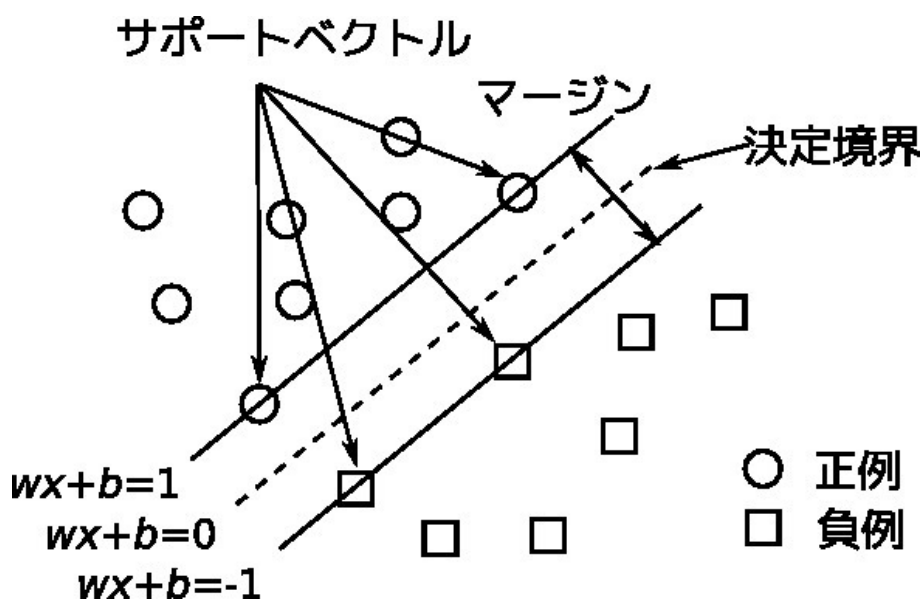


図 2.9: SVM (Support Vector Machine)

カーネルトリック

SVM では線形識別境界で分離ができない場合には，入力空間をより高次の特徴空間に写像し，そこで線形分離を行う「カーネルトリック」を用いることで非線形識別境界にも適用が可能となる．非線形識別境界の高次元への写像の様子を図 2.10 に示す．

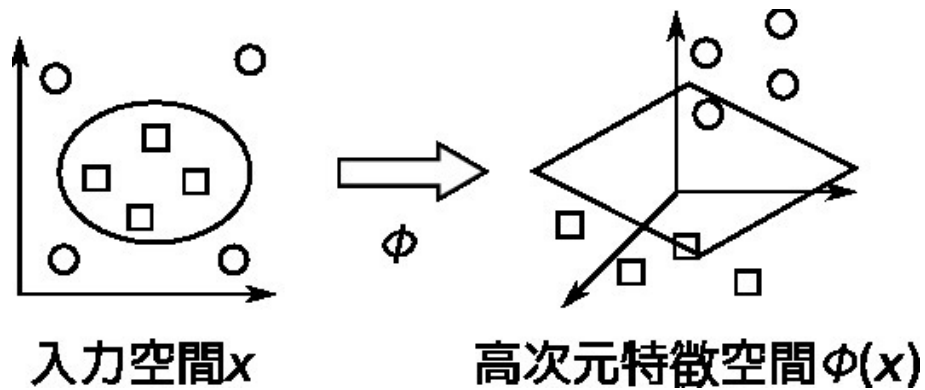


図 2.10: 非線形写像による高次元化

データのスケールリング

入力署名のそれぞれの署名照合を行い相違度や相関値を求める．SVM に入力する前に，それぞれの出力の値を線形変換する．この線形変換は学習データにおける出力の範囲が $-1 \sim 1$ となるような線形変換である．評価用データは学習用データのスケールリングを行ったパラメータでスケールリングを行う．

分類

得られた出力ベクトルに対する真偽の判定を SVM と学習用データによって学習する．SVM のカーネルとして線形，2 次多項式，3 次多項式，4 次多項式，RBF（ガウシアンカーネル）のうち RBF を用いる．評価用データに対する SVM の出力がしきい値 T 以上（以下）なら真筆（偽筆）とする．

2.7.2 RF (Random Forests)

ランダムフォレストとは、Breiman[39] によって提案された複数の決定木構造を持つマルチクラス識別器である。ランダム学習により高次元特徴でも効率的な学習が可能である。また、決定木を作成する際にデータをランダム選択するため、教師データのノイズの影響を抑制することができる。

学習

まず、全学習サンプルからランダムに選択して、ブートストラップ標本を M 個生成する。その際のデータの重複は許容される。決定木学習は、 M 個のブートストラップ標本を用いて決定木を作成していく。分岐が不可能になるまで分岐ノードを繰り返し作成し、決定木を生成する。各決定木の分岐条件は、分岐前のノードと分岐後のノードの不純度の差分である情報利得 IG_p を利用し、以下の式で定義する。そして、情報利得が最も大きくなる組み合わせを分岐ノードにする。

$$IG_p = I(S_p) - \sum_{j \in Left, Right} \frac{N_j}{N_p} I(S_j) \quad (2.21)$$

ここで、 S_p は親ノードのデータセット、 S_j は分割後の子ノード j のデータセットである。 I は不純度であり、 N_p は親ノードのサンプル総数、 N_j は子ノード j のサンプル総数、基本的に2分木を用いるため j は左ノード (*Left*) と右ノード (*Right*) の2種類である。また、不純度 I は次式で定義されるジニ不純度を用いる。

$$I(t) = - \sum_{k=1}^K p(c_k|t)(1 - p(c_k|t)) \quad (2.22)$$

ここで $p(c_k|t)$ はノード t においてクラス c_k であるサンプルの割合である。

識別

図 2.11 にランダムフォレストの識別図を示す．入力サンプル \mathbf{x} を，学習サンプルによって作成された各決定木ごとに事後確率 $P_m(c|\mathbf{x})$ を算出する． M 個の全決定木の $P_m(c|\mathbf{x})$ を用いて入力サンプル \mathbf{x} の事後確率 $P(c|\mathbf{x})$ は次式で定義される．

$$P(c|\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M P_m(c|\mathbf{x}) \quad (2.23)$$

$P(c|\mathbf{x})$ がしきい値 T 以上（以下）なら真筆（偽筆）とする．

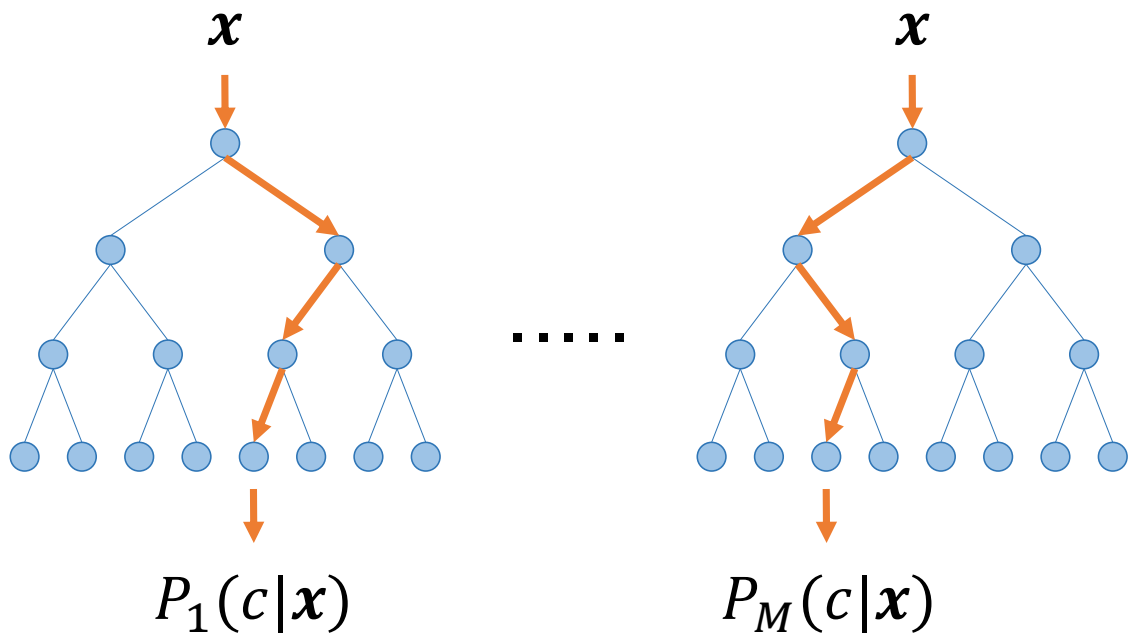


図 2.11: ランダムフォレストによる識別

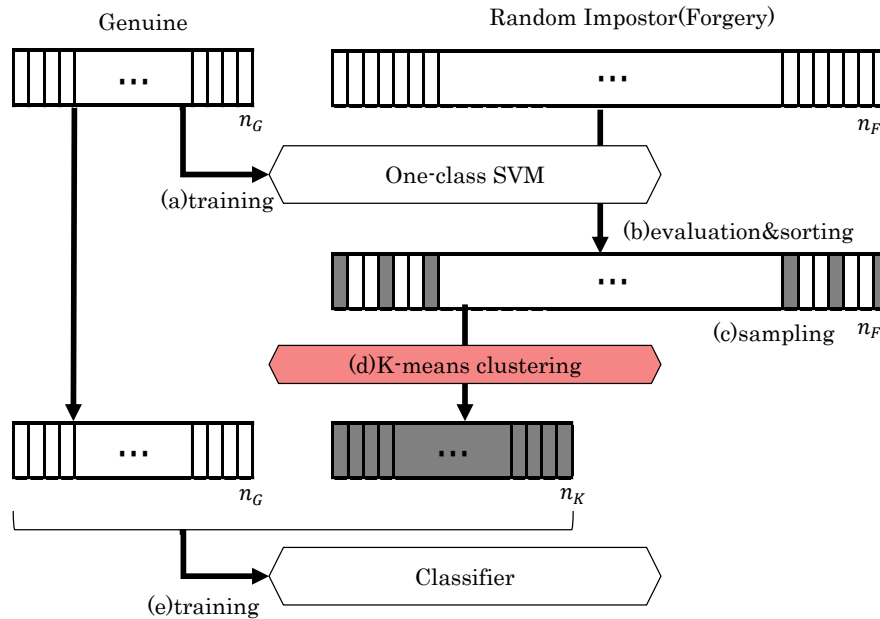


図 2.12: 識別器のランダム偽筆学習の流れ図

2.7.3 ランダム偽筆学習

本研究では、偽筆を用いない手法として、登録された第三者の真筆署名どうしを比較して求めたランダム偽筆照合スコアを偽筆クラスの学習用データに用いる。そして、学習用データセットにおいて多数クラスである偽筆クラスに対し、One-Class SVM と k-means クラスタリングを用いたサンプリングを行う。図 2.12 に提案手法によるランダム偽筆クラスのサンプリング処理の流れ図を示す。処理の流れは以下の通りである。

1. One-Class SVM に真筆署名を学習させ、ランダム偽筆照合スコアを評価する (図 2.12 (a))
2. One-Class SVM の評価値が昇順となるよう全てのランダム偽筆照合スコアを整列する (図 2.12(b)).
3. 真筆データに対する偽筆データの個数を決定し、その個数になるように整列した偽筆照合スコアを n 個おきにサンプリングする (図 2.12(c)).
4. 全てのランダム偽筆照合スコアに対し、サンプリングされたデータを初期値とする k-means クラスタリングを行い、クラスタ中心を求める (図 2.12 (d)).
5. 求めたクラスタ中心を偽筆クラスの学習用データとして用いて二段階目照合の識別器を学習する (図 2.12(e)).

第 3 章

実験

本章では，一段階目照合のオフライン照合の 2 手法の性能評価実験，及び，二段階目照合の識別器の学習手法の有効性の検証を行う．

3.1 実験データ

本研究では，実験データとして SigComp2011 と SigWiComp2013 における多言語オンライン署名照合のデータセット [7][8] を使用した．前者は中国人署名とオランダ人署名のデータセットであり，後者は日本人署名のデータセットである．各言語のデータの内訳を以下に示す．

中国人署名データ

- 登録署名データ : 94 個 (2 個× 1 名分, 10 個× 8 名分, 12 個× 1 名分)
- 評価用署名データ : 586 個 (真筆 125 個, 偽筆 461 個)

オランダ人署名データ

- 登録署名データ : 648 個 (12 個× 54 名分)
- 評価用署名データ : 1258 個 (真筆 648 個, 偽筆 610 個)

日本人署名データ

- 登録署名データ : 66 個 (6 個× 11 名分)
- 評価用署名データ : 726 個 (真筆 330 個, 偽筆 396 個)

3.2 EER (Equal Error Rate)

署名照合には，本人拒否と他人受入の 2 種類の誤りが存在する．それぞれの発生率を本人拒否率 (FRR : False Rejection Rate)，他人受入率 (FAR : False Acceptance Rate) と呼ぶ．本人拒否率は，真筆を誤って偽筆と判定した割合，他人受け入れ率は，偽筆を誤って真筆と判定した割合である．FRR，FAR の定義を以下に示す．

$$\text{本人拒否率 (FRR)} = \frac{b}{a+b} \times 100(\%)$$

$$\text{他人受入率 (FAR)} = \frac{c}{c+d} \times 100(\%)$$

a : 真筆に対して正しく真筆と判定した数

b : 真筆に対して誤って偽筆と判定した数

c : 偽筆に対して誤って真筆と判定した数

d : 偽筆に対して正しく偽筆と判定した数

分類におけるしきい値を変化させて，FRR と FAR をプロットした DET 曲線（図 3.1）は，生体認証の評価に用いられる．DET 曲線は，原点に近い位置を通るほど照合精度が高いことを示す．図 3.1 より，本人拒否率と他人受入率はトレードオフの関係であることが明らかである．本研究では照合のしきい値 T を変化させて， $\text{FAR} = \text{FRR}$ となったときの等価誤り率 (EER: Equal Error Rate) を示す．EER は真筆と偽筆の事前確率の不均一や変化の影響を受けないため，署名照合の性能評価に適している．

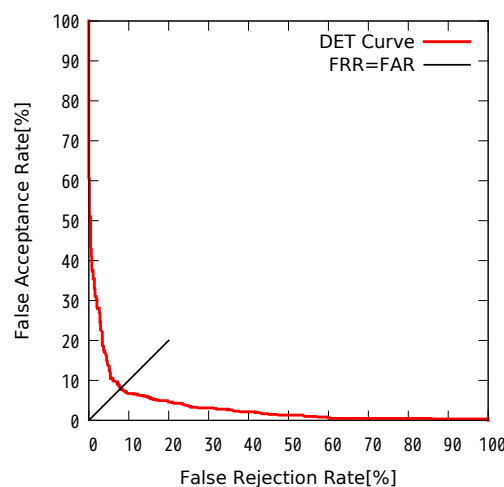


図 3.1: DET 曲線

3.3 組み合わせ照合実験

従来の組み合わせ分割照合法に今回提案した手法を追加することによる性能評価を行う。

3.3.1 実験条件

SVM のカーネルには RBF (ガウシアンカーネル) を用いた。しきい値は全筆記者共通の値とした。また、SVM と RF の学習データを中国人署名、オランダ人署名、日本人署名の 3 つのデータ全て合わせたものを用いて照合実験を行う。評価方法には、Leave-One-Out を用いる。これは、評価用データから 1 つの署名だけを除き、残りすべての署名を識別器の学習モデルにする方法である。

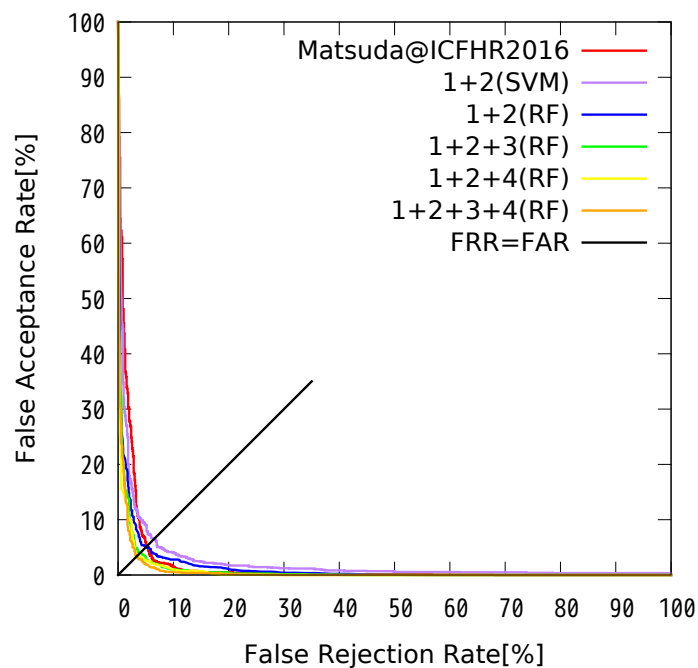
3.3.2 実験結果

表 3.1 に用いる特徴を変更したときの照合結果を示す。今回の実験における DET 曲線を図 3.2 に示す。表 3.1 より全ての 4 つの手法を組み合わせ、RF で識別した手法で EER を 3.35% まで低減することが出来た。下線部の値は、EER の最小値を表す。また、各言語において SigComp のベストスコアに対し、提案手法では全言語を同じしきい値を用いたとしても性能が良くなった。

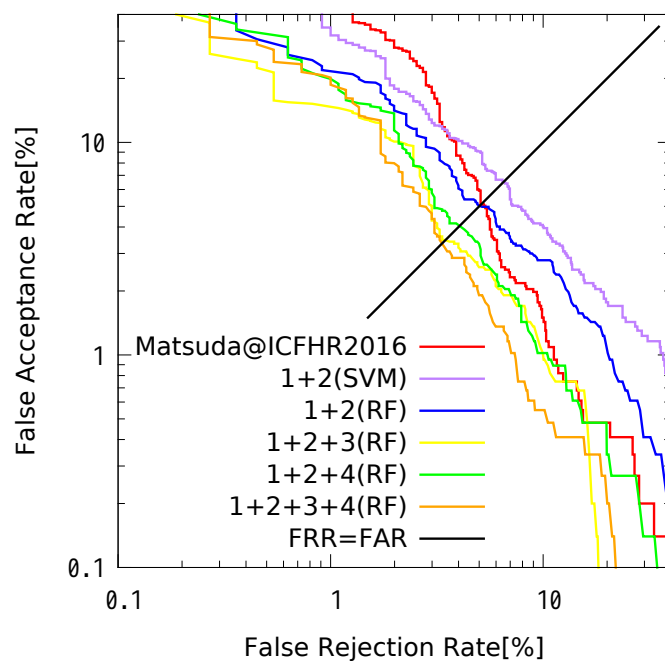
表 3.1: 各手法の組み合わせによる等価誤り率 (%)

	SigComp*	Matsuda[40]	1 + 2 (SVM)	1 + 2 (RF)	1 + 2 + 3 (RF)	1 + 2 + 4 (RF)	1 + 2 + 3 + 4 (RF)
Chinese	6.83[21]	3.24	4.10	5.63	3.75	5.46	4.27
Dutch	3.65[21]	5.72	7.23	5.01	4.37	2.62	3.10
Japanese	8.75[23]	5.51	7.30	4.55	3.72	3.31	3.03
Total		5.10	6.54	5.02	4.05	3.46	<u>3.35</u>

- 1 グレースケール署名画像から濃度こう配特徴を抽出し、登録署名とのマハラノビス距離を算出するオフライン照合手法
- 2 時系列署名データに DP マッチングを適用して、登録署名との相違度を算出するオンライン照合手法
- 3 RGB チャネル画像から濃度こう配特徴を抽出し、登録署名とのマハラノビス距離を算出するオフライン照合手法
- 4 グレースケール画像から輝度ヒストグラ特徴を抽出し、登録署名との類似度を算出するオフライン照合手法



(a) DET 曲線-線形目織



(b) DET 曲線-対数目織

図 3.2: 実験 3.3 における DET 曲線

3.3.3 考察

表 3.1 より全ての 4 つの手法を組み合わせ、RF で識別した結果が一番良い結果となった。SVM より RF が性能がいいのは、SVM では今回学習している真筆と偽筆のクラスの割合が偽筆のほうが多く、識別境界が偽筆クラス側に引き寄せられるような効果があることを示唆している。また今回、二段階目照合の識別器の学習データに、真筆クラスには偽筆クラスと酷似する署名、偽筆クラスには真筆クラスと酷似する署名がノイズとしてそれぞれ含まれている。二段階目照合の SVM では、これらの全署名を学習して識別境界を作成するため、ノイズの影響を大きく受ける。それに対し、学習データをランダムに選択し、複数の決定木を作成する RF では、必ずしも紛らわしい署名を含むとは限らず、ノイズの影響を受けず識別することができたと考えられる。

また、今回提案したカラー画像生成手法では、色相の変化は登録用真筆署名の筆圧に基づいている。そのため、真筆と登録署名とは筆圧情報が類似しており、真筆と登録署名の色相変化も似ている。それに対し、訓練偽筆と登録署名とは筆圧情報が相違しているため、訓練偽筆と登録署名の色相変化は差異が大きくなる。そして、RGB チャンネル分離により形状が大きく異なる署名を生成することができたため、真筆と訓練偽筆の識別が可能になったことが示唆される。

さらに、筆圧、筆速を輝度値、線幅に反映させたグレースケール画像の輝度ヒストグラム類似度を用いる手法を組み合わせることで、形状特徴では判別できない筆記中の情報を反映したので、さらに正確な識別が可能になったと考えられる。

3.4 ランダム偽筆を用いた照合実験

二段階目照合の識別器の学習に訓練偽筆を用いない手法について、提案手法の有効性の評価実験を行う。

3.4.1 実験条件

二段階目照合の識別器の学習に用いる偽筆の種類を変更することにより、手法の有効性を検証する。それぞれの凡例は二段階目照合の識別器の学習に用いた署名データを示し、それぞれ以下の条件である。

Skilled Forgery 真筆+訓練偽筆

One-Class SVM 真筆のみ（偽筆は用いていない）

Random impostor 真筆+全てのランダム偽筆

Random sampling 真筆+ランダム偽筆に対するランダムサンプリング

Proposed 1 真筆+ランダム偽筆に対し One-class SVM の評価値でソートを行い、真筆と同数に間引いたランダム偽筆

Proposed 2 真筆+ランダム偽筆に対し One-class SVM の評価値でソートを行い、真筆と同数に間引いたランダム偽筆を初期値として全ランダム偽筆をクラスタリングにより取得した真筆と同数のランダム偽筆

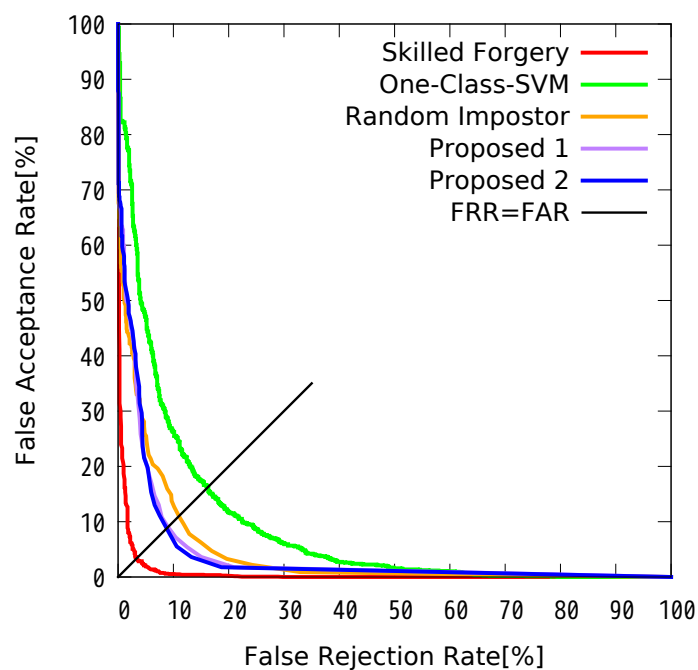
今回は 3.3 で最も性能の良かった 4 つのすべての手法を RF で組み合わせて実験を行う。

表 3.2: それぞれの実験条件における EER(%)

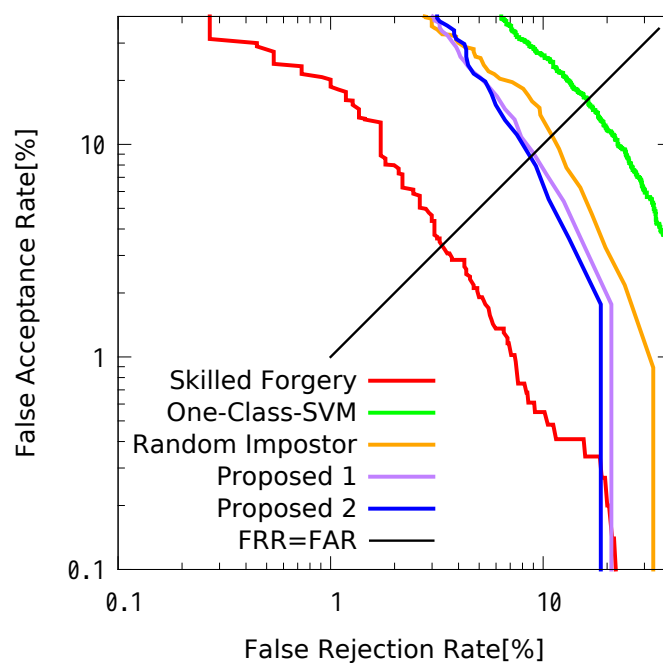
Training method	Total EER (%)
Skilled-forgery	3.35
One-class SVM	16.23
Random impostor	10.97
Random sampling	8.83
Proposed 1	9.07
Proposed 2	8.48

3.4.2 実験結果

表 3.2 はそれぞれの学習条件下での照合結果の EER を示す。6 条件中，訓練偽筆を学習用データに用いた場合に EER の最小値が得られたが，ランダム偽筆を学習に用いた場合は提案手法による EER の低減が実現された。図 3.3 に，上記の 6 条件中，Random sampling を除いた 5 条件の Detection Error Tradeoff (DET) 曲線を線形目盛 (a) と対数目盛 (b) で示す。図 3.3 より，学習用データの違いが FAR, FRR に与える影響が読みとれる。



(a) DET 曲線-線形目織



(b) DET 曲線-対数目織

図 3.3: 実験 3.4 における DET 曲線

3.4.3 考察

訓練偽筆を用いて学習した場合の性能に対して、学習用データに訓練偽筆を含まない one-class SVM では FAR, FRR の両方が増加する。これは二段階目の識別器である One-Class SVM が真筆クラスのみを用いて学習しているため、特徴空間中の偽筆クラスの分布に対して適切な決定境界を獲得できていないことを示唆する。これに対して、ランダム偽筆サンプルを学習用データに含めることで、FRR が低減される。これは第二段階照合の識別器が偽筆の分布に対して適切な決定境界を獲得できたことが理由と考えられる。一方で、全ての偽筆サンプルを学習に用いる Random impostor においては FAR が大きく増大している。この学習条件においては偽筆クラスのサンプル数が大きく、また、真筆署名と大きく異なる他人署名が含まれるため、決定境界が偽筆クラスに引き寄せられるような影響をうけることが原因と思われる。ランダム偽筆の導入による FAR の増大は、提案手法 (Proposed1, Proposed2) による偽筆クラスのサンプリングにより低減されることが確認できる。この結果は、ランダム偽筆に対して今回提案した One-Class SVM とクラスタリングを用いたサンプリング手法により効果的なデータ削減が実現されていること、データの偏りに対する性能低下を低減できることを示唆する。

図 3.4, 図 3.5, 図 3.6, 図 3.7, 図 3.8 はそれぞれ、真筆の照合スコアの散布図, 訓練偽筆 [Skilled Forgery] の照合スコアの散布図, 図 2.12(b) で作られた第三者の真筆署名どうしを比較して求めたランダム偽筆照合全体 [Random impostor] の散布図, 2.12(b) の偽筆スコアベクトルを n 個おきにサンプリングした結果 [Proposed 1], 2.12(d) のクラスタリングを行った結果 [Proposed 2] である。全て、真筆を学習した One-Class SVM で識別した評価値によって色分けをしている。それぞれの図の原点は訓練偽筆クラス (図 3.5) の平均に設定されている。この図から、真筆クラスの分布 (図 3.4) の大きさを基準とすると、ランダム偽筆クラス (図 3.6) 中に真筆クラス付近に存在するサンプルが多数含まれており、これらを偽筆クラスとして第二段階照合の識別器の学習に利用可能であることが示唆される。全登録署名それぞれに対して第三者署名と比較すると真筆に対して偽筆の数が大きく (図 3.6)、偏った学習用データにより性能低下するため、偽筆の数を減らす必要があるが、One-Class SVM の評価値を用いることでランダム偽筆全体の分布を保ったままサンプリングできていることが明らかである (図 3.7)。さらに、安定にサンプリングできる手法としてクラスタリングを行うことでランダム性がなくなり、全ランダム偽筆の特性を保った偽筆が作成できていると考えられる (図 3.8)。

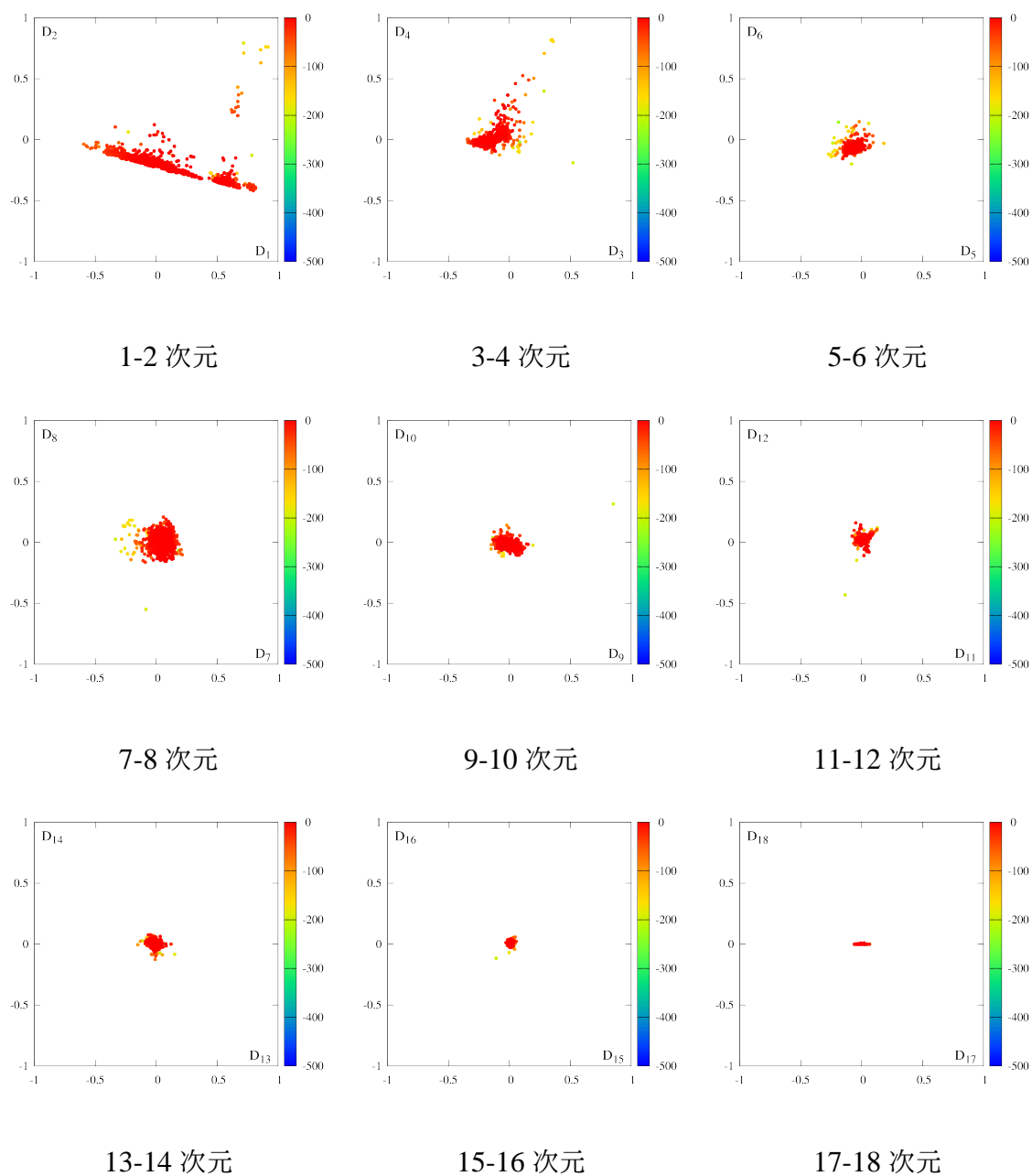


図 3.4: 真筆の照合スコアの分布

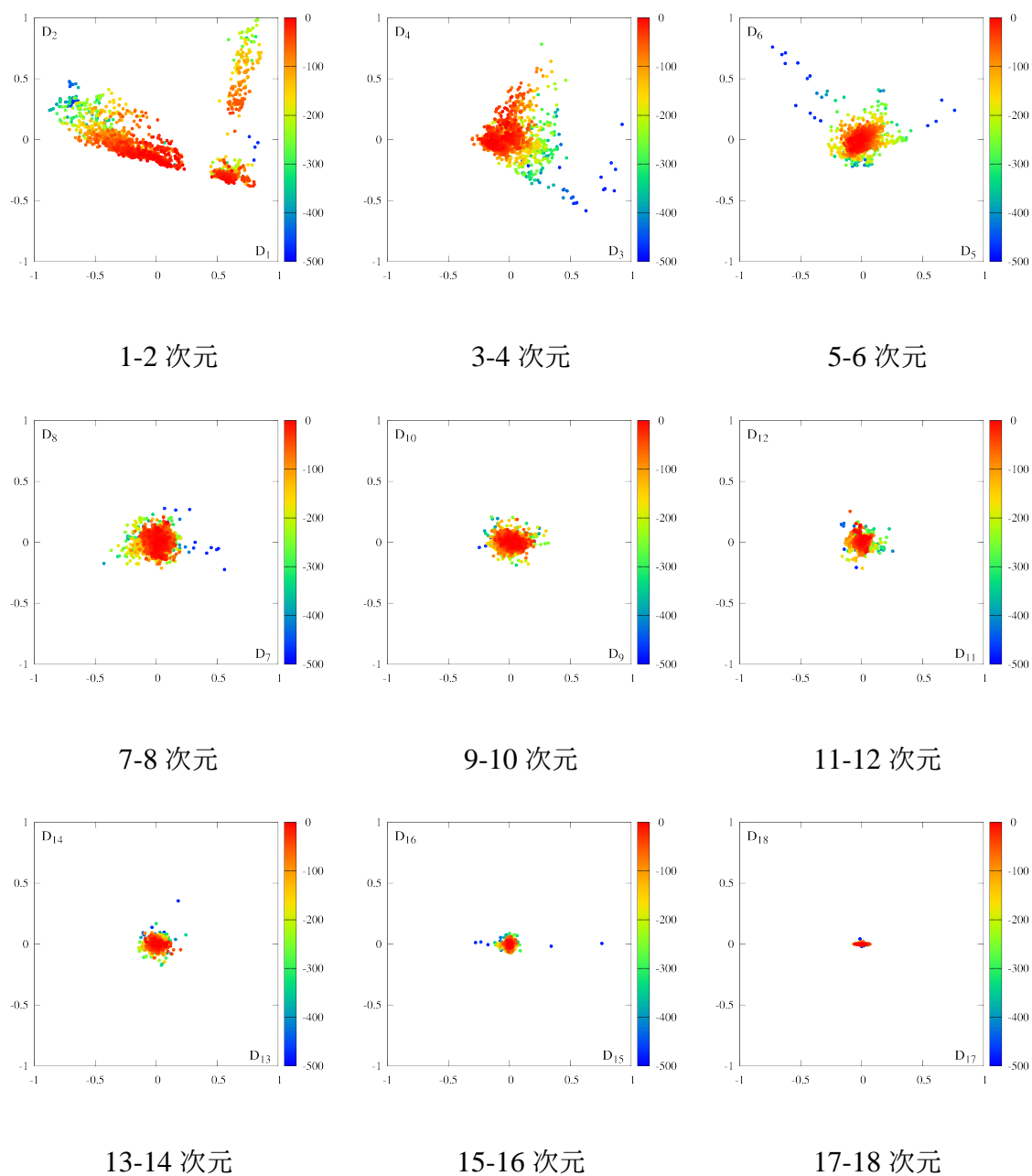


図 3.5: 訓練偽筆の照合スコアの分布

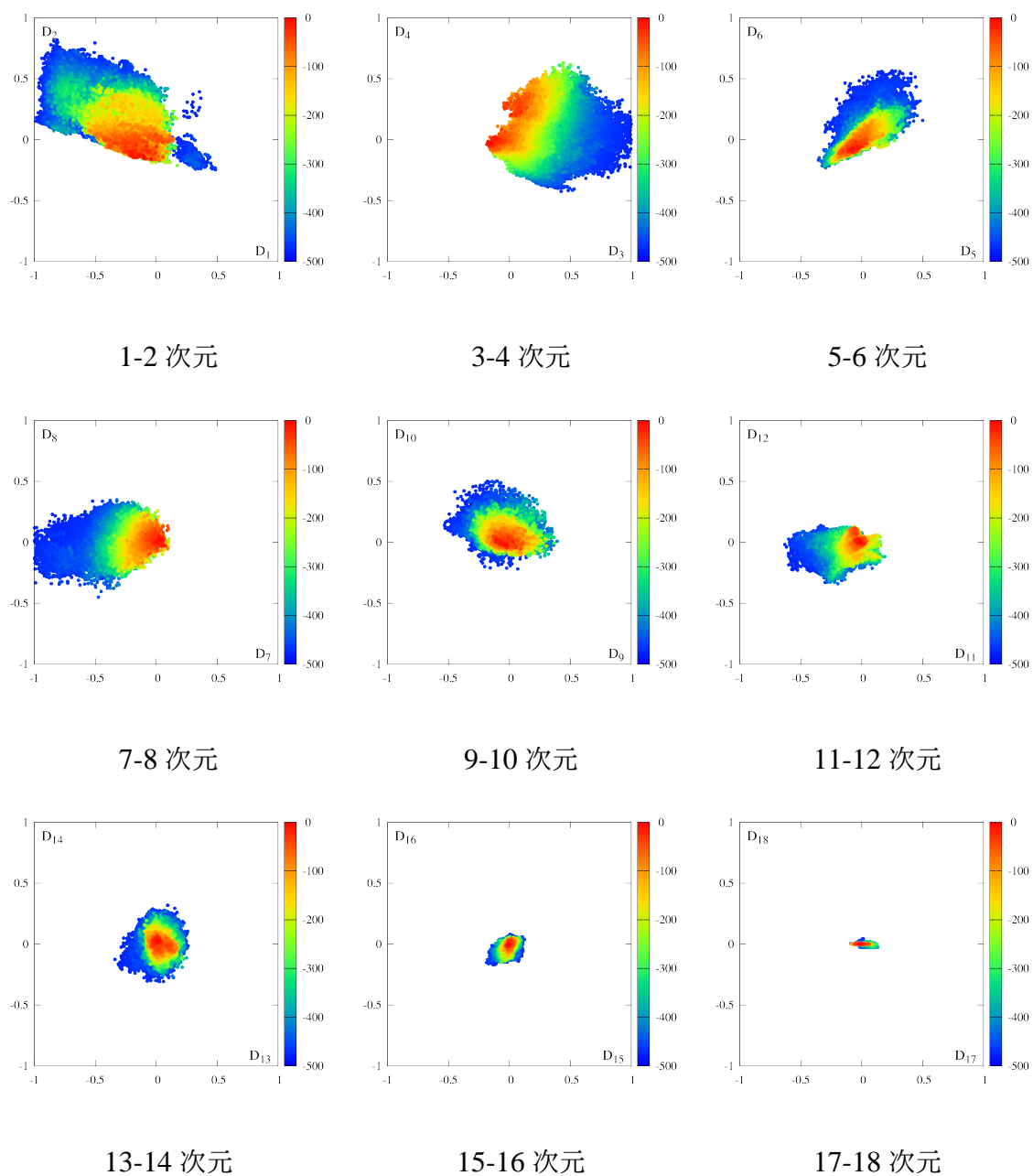


図 3.6: ランダム偽筆全体の照合スコアの分布

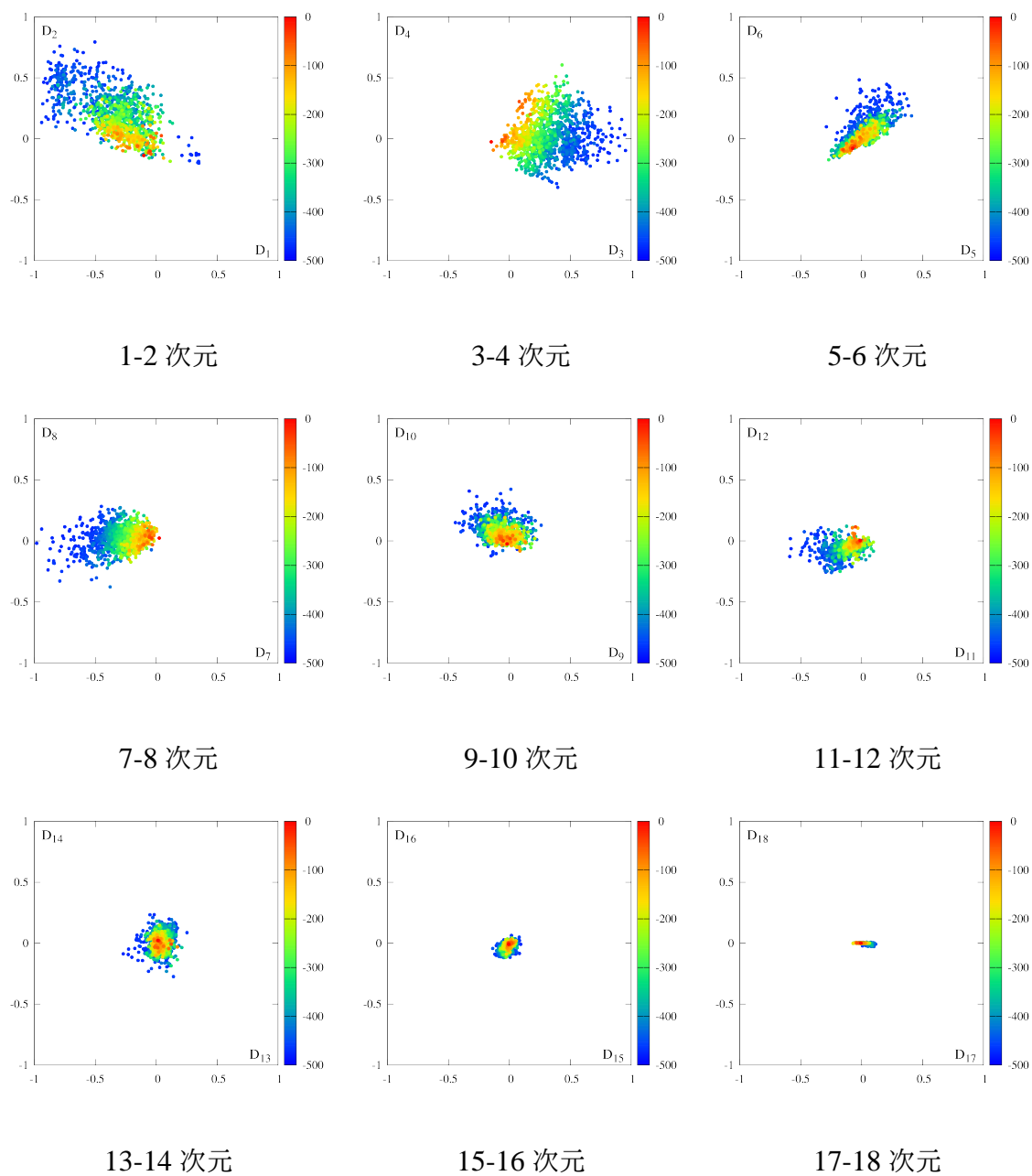


図 3.7: Proposed 1 の偽筆の照合スコアの分布

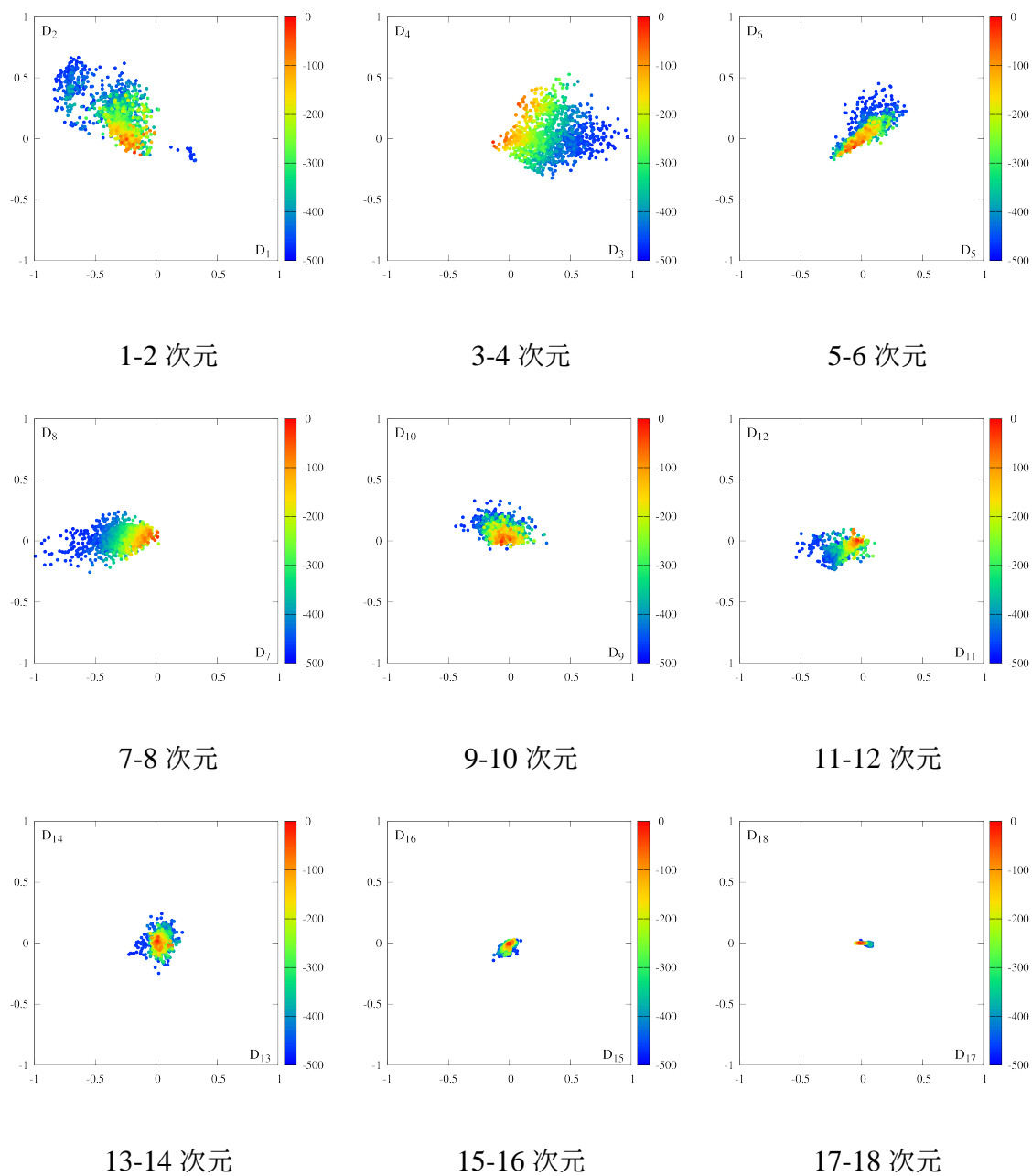


図 3.8: Proposed 2 の偽筆の照合スコアの分布

第 4 章

結言

4.1 まとめと今後の課題

本研究では，組み合わせ分割署名照合法の 3 つの課題に対して改善をした．

まず，組み合わせ分割照合法を真の多言語署名照合に適用するために，SigComp2011, SigWiComp2013 のデータセットを統合して学習と評価に用いるように改良を行った．先行研究は評価署名の言語と，識別器の学習署名言語は同一として実験を行っていたが，これでは，評価する署名言語が既知で，さらに評価署名と同一言語の学習データをあらかじめ用意する必要があった．今回，二段階目照合の識別器の学習データの多言語化を実現することにより，どの言語署名でも筆記者が真筆署名を数枚登録さえすれば，署名照合を行うことができることが明らかになった．

次に，署名照合の性能向上のために，一段階目照合のオフライン照合を 2 手法提案した．先行研究では，筆圧を反映したグレースケール画像から特徴を抽出し，登録署名とのマハラノビス距離を求めるオフライン照合手法と，DP マッチングを用いるオンライン照合手法を組み合わせていた．本研究では，以下の 2 手法を新たに導入した．1 つ目は，筆圧を色相に反映したカラー署名画像から RGB 各チャネルの署名画像を生成し，各署名から特徴を抽出しマハラノビス距離を算出する手法である．2 つ目は，筆圧を輝度値，筆速を署名の線幅に反映させたグレースケール画像の輝度ヒストグラム類似度を用いる手法である．実験の結果，提案した 2 手法の導入により EER を 3.35% に低減することが確認できた．

最後に，二段階目照合の識別器学習に，収集コストを要する訓練偽筆を使用せず，偽筆クラスの学習サンプルに第三者の登録署名を用いるランダム偽筆学習を提案した．多言語署名から構成される SigComp データセットを用いた評価実験の結果，訓練偽筆を学習する場合に比べて照合性能は低下するものの，真筆のみの学習より高い精度で照合可能であることが明らかになった．

今後の課題として，ランダム偽筆を用いた照合における性能向上が必要であり，さらに，登録真筆署名に形状変化を加えることで訓練偽筆を疑似的に作成して用いれば性能向上が望めると考える．さらに，今回提案した手法はオフライン署名照合のみであるため，HMMなどのオンライン署名照合を用いた手法を組み合わせることで EER をさらに低減することが可能であると考ええる．

付録 A

ランダム偽筆のサンプリング枚数における性能調査

2.7.3 のランダム偽筆学習手法における，真筆に対する偽筆のサンプリングの割合 N_K/N_G による性能調査を行う．

A.1 実験条件

同様に実験データとして SigComp2011 と SigWiComp2013 における多言語オンライン署名照合のデータセット [7][8] を使用した．評価方法には，Leave-One-Out を用いる．

A.2 実験結果

図 A.1 にランダム偽筆学習手法の真筆に対する偽筆のサンプリングの割合 N_K/N_G を変更したときの照合結果を示す．今回， $N_K/N_G = 2.0$ の時に EER が 8.21% である．また，真筆に対する偽筆の枚数が真筆に近い時に性能が良いことが明らかになった．これは，ランダムフォレストの決定木を作成する際のブースストップ標本において真筆と偽筆の割合が近い場合，ランダム偽筆クラスに大きく影響されることを防ぐことができると示唆する．

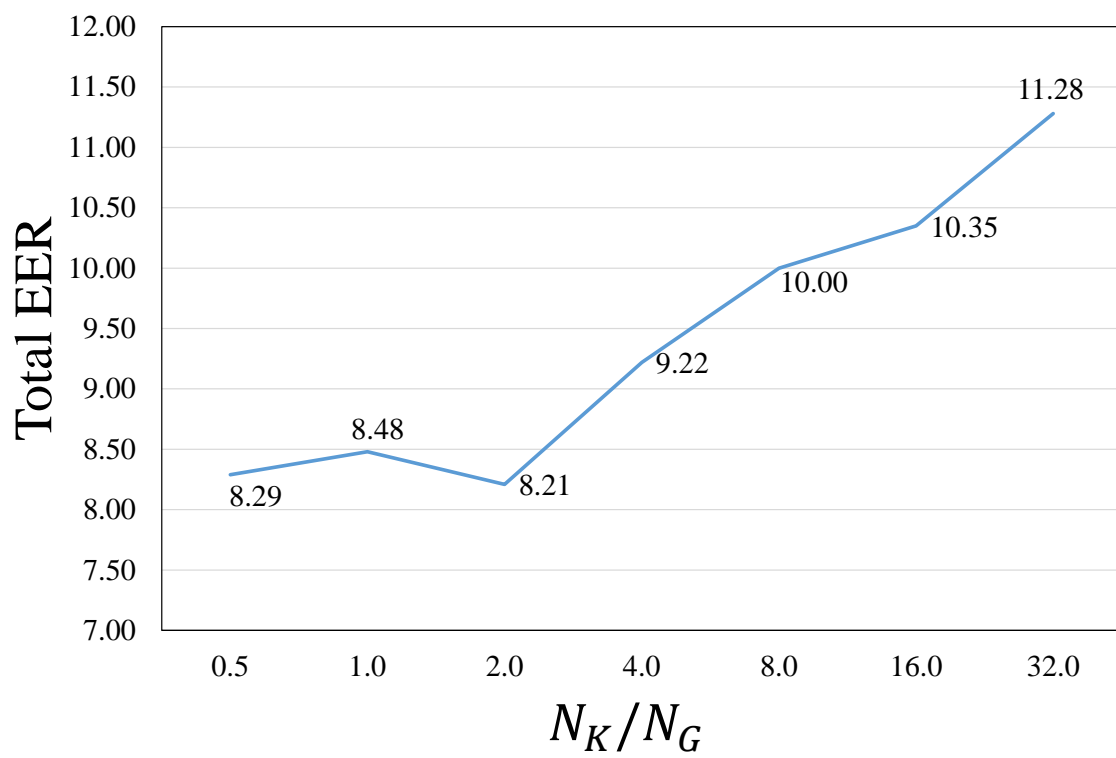


図 A.1: 真筆に対する偽筆の枚数による等価誤り率

付録 B

署名データリスト

~matsuda/Research/Classifier/Testdata/

内に、言語別に置いてある．各ディレクトリについて以下に示す．

```
|
|--Chinese
|   | # 中国人署名のデータ
|   |--Reference
|   | # 参照用署名のディレクトリ
|   |--Questioned
|   | # 評価用署名のディレクトリ
|
|--Dutch
|   | # オランダ人署名のデータ
|   |--Reference
|   | # 参照用署名のディレクトリ
|   |--Questioned
|   | # 評価用署名のディレクトリ
|
|--Japanese
|   | # 日本人署名のデータ
|   |--Reference
|   | # 参照用署名のディレクトリ
|   |--Questioned
|   | # 評価用署名のディレクトリ
```

付録 C

プログラムソースリスト

作成したプログラムを以下のディレクトリ

~matsuda/Research/Classifier/

に置く．ファイル等の詳細やコンパイル方法，実行方法はディレクトリ内の README を参照してください．

Classifier

| # 研究用ディレクトリ

|

| --DP

| # DP マッチングにより求めた相違度，照合結果を格納するディレクトリ

|

| --Dictionary

| # 学習辞書のディレクトリ

|

| --feature

| # 濃度こう配特徴ベクトルを保存するディレクトリ

|

| --Histogram

| # ヒストグラム類似度を保存するディレクトリ

|

| --Fusion

| # Classifier による各署名照合手法融合用ディレクトリ

|

```
| --images
|   # 生成した署名画像を保存するディレクトリ
|
| --preprocess
|   # 画像生成のための前処理を行うディレクトリ
|
| --Result
|   # 識別関数により求めた距離値，照合結果を格納するディレクトリ
|
| --run
|   # 実験用のプログラム，シェルスクリプトを一括実行するためのディレクトリ
|
| --Testdata
|   # 実験用データのディレクトリ
```


付録 D

発表資料

修士論文発表で用いたプレゼンテーション資料を以下に掲載する。



組み合わせ分割多言語署名照合の 高精度化に関する研究

三重大学 大学院 工学研究科 情報工学専攻
ヒューマンインターフェース研究室
416M521 松田 圭悟

バイオメトリクスにおける署名照合

署名照合

行動的特徴を用いた
生体認証





利点

- 海外で社会的受容性が高い
- 登録データの**変更が可能**
- 低コスト**で導入可能

利用例

- クレジットカード支払い
- 銀行取引
- 契約締結

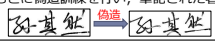


1

署名照合の技術的課題

- 高精度化
 - 本人拒否エラーと他人受入エラーの低減
- 運用コストの低減
 - 学習用**訓練偽筆**サンプルの収集は高コスト
- 多言語化
 - 入力署名の言語が未知の場合での照合

※**訓練偽筆**
本人署名の筆記情報（形状、書き順など）
をもとに偽造訓練を行い、筆記された署名



2

関連研究

- オンライン署名照合
 - 最長共通部分列マッチングを用いた手法[1](2013)
 - ルジャンドル多項式に基づく特徴抽出を用いた手法[2](2014)
 - 複数の特徴を組み合わせるオンライン署名照合
 - 複数のオンライン特徴とオフライン特徴を抽出し各特徴を多段階で組み合わせる手法[3](2015)
 - 複数の特徴から粒子群最適化(PSO)を用いて特徴選択し照合する手法[4](2017)

[1] K. Barkoula, et al., "Online signature verification based on signatures turning angle representation using longest common subsequence matching", IJDAR, 2013
 [2] M. Parodi, et al., "Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations", Pattern Recognition, 2014
 [3] J. Galbally, et al., "On-line signature recognition through the combination of real dynamic data and synthetically generated static data." Pattern Recognition, 2015
 [4] S. Dutta, et al., "An Efficient Approach for Recognition and Verification of On-line Signatures using PSO", ACPR 2017

3

本研究の成果

□ 高精度化

- 新しいオフライン照合手法の導入
 - [査読有り国際会議]K.Matsuda et al.(ACPR2017)

□ 運用コストの低減

- 訓練偽筆を用いない学習方法の提案
 - [学術論文]松田 他, 電子情報通信学会論文誌A (Dec.2017)

□ 多言語化

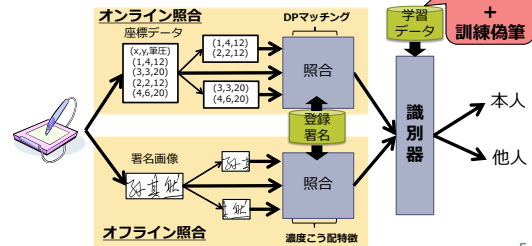
- 学習用データセットを多言語化
 - [査読有り国際会議]K.Matsuda et al.(ICFHR2016)

4

組み合わせ分割照合(伊藤,上平,尾木)

□ 2段階照合

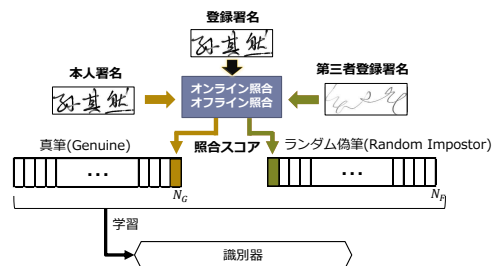
- 1段階目照合: オンライン照合, オフライン照合
- 2段階目照合: 識別器(SVM/RF)



5

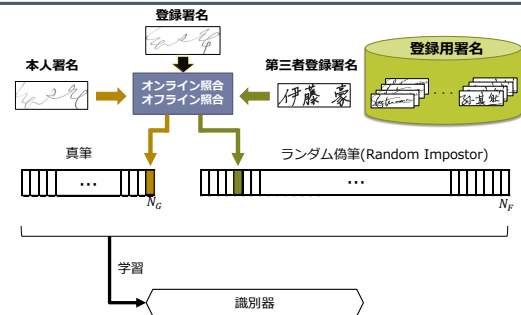
訓練偽筆を用いない学習方法

- 従来は, 訓練偽筆(Skilled Forgery)を用いて算出した照合スコアを使用



6

ランダム偽筆を用いた学習



7

評価実験用データセット

- SigComp2011[5], SigWIComp2013[6]のデータセット (多言語署名照合のためのコンテスト)

□ 登録用真筆データ

中国語署名: 94 (10人)
オランダ語署名: 648 (54人)
日本語署名: 60 (10人)



□ 評価データ

中国語署名: 586 (真筆:125, 偽筆: 461)
オランダ語署名: 1258 (真筆:648, 偽筆: 610)
日本語署名: 726 (真筆:330, 偽筆: 396)

[5] M.Liwicki, et al. "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)." ICAR2011.
[6] M.I.Malik, et al. "ICAR 2013 competitions on signature verification and writer identification for on-and offline skilled forgeries (SigWIComp 2013)." ICAR2013

8

評価方法

評価の定義

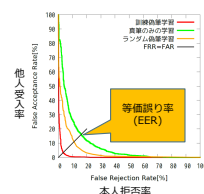
$$\text{本人拒否率} = \frac{b}{a+b} \times 100(\%)$$

$$\text{他人受入率} = \frac{c}{c+d} \times 100(\%)$$

評価基準

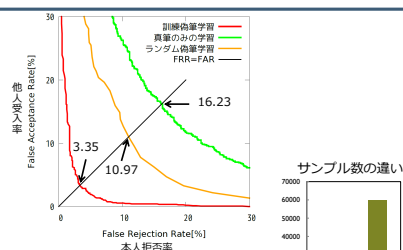
DET曲線:
本人拒否率と他人受入率を
プロットしたもの

		判定結果	
		真	偽
正解	真	a	b
	偽	c	d



9

照合結果(DET曲線)



- 真筆照合スコア(1100個)に対して偽筆照合スコア(59725個)が多すぎる
- 偽筆照合スコアにばらつきがある
- 2クラスの学習サンプル数の調整が必要

10

Imbalanced learning

- アンダーサンプリング
 - Random undersampling
 - クラスタリングに基づくサンプリング手法[7]
 - Informed undersampling[8]
- オーバーサンプリング
 - Random oversampling
 - SMOTE[9]

課題

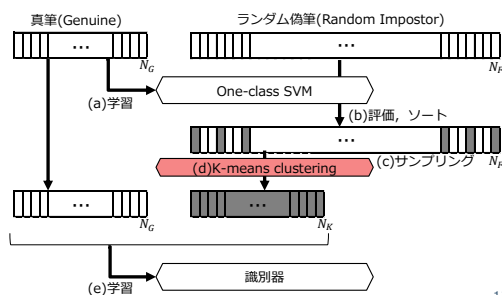
ランダム要素のない安定的なサンプリングが必要

[7] T. Jo and N. Japkowicz, "Class Imbalances versus Small Disjuncts," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 40-49, 2004.
 [8] X.-Y. Liu, J. Wu, and Z.-H. Zhou, "Exploratory Under Sampling for Class Imbalance Learning," Proc. Int. Conf. Data Mining, pp. 965-969, 2006.
 [9] Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research 16 (2002): 321-357

11

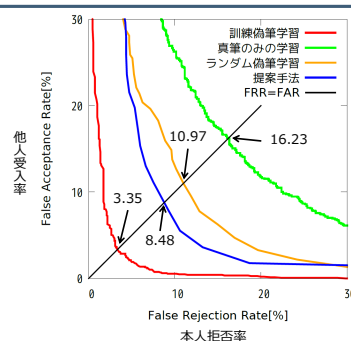
提案手法

- One-class SVMとクラスタリングを用いる手法



12

照合結果(DET曲線)



13

まとめ

- 高精度化
 - 新しいオフライン照合手法の導入
 - 等価誤り率3.35%に低減
- 運用コストの低減
 - 訓練偽筆を用いない学習方法の提案
 - ランダム偽筆による学習: 10.97%
 - 提案手法によるサンプリング: 8.48%
- 多言語化
 - 学習用データセットを多言語化
 - 中国語, オランダ語, 日本語署名で照合可能

14

謝辞

若林哲史教授には、本研究を進めるにあたり、適切な御助言、ご指導をして頂きました。深く感謝いたします。大山航助教には、研究の進め方や発表における論理的な説明のしかた、論文執筆に至るまで様々なご指導をしていただきました。また研究における課題にいつも親身に相談に乗って頂きました。心より感謝申し上げます。白井伸宙助教には、私の研究に適切な御助言、ご指導をして頂きました。深く感謝いたします。木村文隆名誉教授には、ご退官までの間、研究について親身になってご指導して頂きました。心より感謝申し上げます。三宅康二名誉教授には、毎週のディスカッションに足を運び、大変有益な助言をいただきました。心より感謝申し上げます。短期留学中に研究の助言をしてくださったミシガン大学ディアボーン校所属、Yi Lu Murphey 教授、Malayappan Shridhar 教授に大変感謝申し上げます。また、諸連絡の掲示や備品の貸し出しなど、日頃様々なお世話をして頂き研究を快適にしやすい環境を作ってくれた、ヒューマンインターフェース研究室事務員田中みゆきさん、中塚沙智子さんに深く感謝いたします。そして、本研究や研究室生活への多くのアドバイスを与えてくださった研究室の先輩、後輩の皆様、お互いの研究について議論し合い、技術の共有を行い、ともに切磋琢磨した同期の皆様に深く感謝いたします。皆様のおかげでこの学生生活が私にとって非常に楽しく、有意義なものとなりました。最後になりましたが、ここまで私の学生生活を支えてくれた家族に今一度の感謝の意を表して、本論文の結びといたします。

参考文献

- [1] 鷺見和彦. "1. バイオメトリクスセキュリティ概論 (j 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 89.1 (2006): 27-30.
- [2] 瀬戸洋一. "1. バイオメトリックセキュリティ認証技術の動向と展望 (特集 バイオメトリック認証システム)." 情報処理 47.6 (2006): 571-576.
- [3] (社) 日本自動認識システム協会. よくわかるバイオメトリクスの基礎. オーム社, 2005.
- [4] 中村善一. 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. 奈良先端科学技術大学院大学博士論文, 2008.
- [5] 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.
- [6] J. Ortega-Garcia, J. Fierrez-Aguilar, J. Martin-Rello, and J Gonzalez Rodriguez. Complete signal modeling and score normalization for function based dynamic signature verification. Springer Lecture Notes in Computer Science, Vol.2688, pp.658-667.2003.
- [7] Liwicki, M.; Malik, M.I.; van den Heuvel, C.E.; Xiaohong Chen; Berger, C.; Stoel, R.; Blumenstein, M.; Found, B., "Signature Verification Competition for Online and Offline Skilled Forgeries (SigComp2011)," in Document Analysis and Recognition (ICDAR), 2011 International Conference on , pp.1480-1484, 18-21 Sept. 2011
- [8] Malik, M.I.; Liwicki, M.; Alewijnse, L.; Ohyama, W.; Blumenstein, M.; Found, B., "ICDAR 2013 Competitions on Signature Verification and Writer Identification for On- and Offline Skilled Forgeries (SigWiComp 2013)," in Document Analysis and Recognition (ICDAR), 2013 12th International Conference on , pp.1477-1483, 25-28 Aug. 2013
- [9] 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.

- [10] J. Fierrez, J. Ortega-Garcia, D Ramos and J. Gonzalez-Rodriguez. HMM-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recognition Letters*, Vol.28, No.16, pp.2325-2334, 2007.
- [11] 村松大吾, 本郷保範, 松本隆. ユーザ共通 Fusion モデルを用いたオンライン署名認証. *信学論*, Vol.J90-D, No.2, pp.450-459, 2007.
- [12] Ortega-Garcia, Javier, et al. "MCYT baseline corpus: a bimodal biometric database." *IEE Proceedings-Vision, Image and Signal Processing* 150.6 (2003): 395-401.
- [13] Garcia-Salicetti, Sonia, et al. "BIOMET: A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities." *Audio-and Video-Based Biometric Person Authentication*. Springer Berlin Heidelberg, 2003.
- [14] Y. Komiyama, 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.
- [15] 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. *信学論*, Vol.J87-A, No.6, pp.805-815, 2004.
- [16] 野中俊宏, 菊地真美, and 赤松則男. "HIP2000-19 署名時の筆圧情報に基づく個人認証." *電子情報通信学会技術研究報告. HIP, ヒューマン情報処理* 100.139 (2000): 7-12.
- [17] K. Barkoula, G. Economou, S. Fotopoulos, "Online signature verification based on signatures turning angle representation using longest common subsequence matching", *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, pp. 261-272, 09 2013.
- [18] A. Kholmatov, B. Yanikoglu, "SUSIG: An on-line signature database associated protocols and benchmark results", *Pattern Anal. Appl.*, vol. 12, no. 3, pp. 227-236, Sep. 2009.
- [19] D. Yeung et al., "SVC2004: First international signature verification competition", *Proc. Int. Conf. Biometric Authentication*, pp. 16-22, 2004. <http://www.cse.ust.hk/svc2004/results.html>.
- [20] Marianela Parodi, Juan C. Gmez, "Legendre polynomials based feature extraction for online signature verification. Consistency analysis of feature combinations", *Pattern Recognition*, Volume 47, Issue 1, 2014, Pages 128-140, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2013.06.026>.
- [21] "Digital Transaction Management — Digital and Electronic Signature Software — Trust Services" <https://www.xyzmo.com/> (accessed 2018-1-19)

- [22] A. Kholmatov and B. Yanikoglu, "Identity authentication using im-proved online signature verification method," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2400-2408, 2005.
- [23] Zeinali Hossein, Babaali Bagher, Hadian Hossein. "Online Signature Verification Using i-vector Representation". *IET Biometrics*. 10.1049/iet-bmt.2017.0059.
- [24] 上平 裕太, 大山 航, 若林 哲史, 木村 文隆: "組み合わせ分割照合法による日本語署名照合の高精度化", *電学論 C*, Vol. 134, No. 12, pp.1809-1816 (Dec. 2014) DOI:<http://dx.doi.org/10.1541/ieejeiss.134.1809>
- [25] W. Ohyama, Y. Ogi, T. Wakabayashi, and F. Kimura, "Multilingual signature-verification by generalized combined segmentation verification." *Proc. 13th International Conference on Document Analysis and Recognition (ICDAR2015)*, pp. 811-815. Nancy, France, Aug. 2015.
- [26] B. Fang, C.H. Leung, Y.Y. Tang, P.C.K. Kwok, K.W. Tse and Y.K. Wong, "Offline signature verification with generated training samples." *IEE Proc. - Vision, Image and Signal Processing*, vol.149, no.2, pp. 85-90, Apr.2002.
- [27] M. Diaz, S. Chanda, M.A. Ferrer, C.K. Banerjee, A. Majumdar, C. Carmona-Duarte, P. Acharya and U. Pal, "Multiple generation of Bengali static signatures." *Proc. 15th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR2016)*, pp. 42-47, Shenzhen, China, 2016.
- [28] Javier Galbally, Moises Diaz-Cabrera, Miguel A. Ferrer, Marta Gomez-Barrero, Aythami Morales, Julian Fierrez "On-line signature recognition through the combination of real dynamic data and synthetically generated static data." *Pattern Recognition* 48.9 (2015): 2921-2934.
- [29] Saikat Dutta, Rajkumar Saini, Pradeep Kumar, Partha Pratim Roy "An Efficient Approach for Recognition and Verification of On-line Signatures using PSO ", 2017 4th IAPR Asian Conference on Pattern Recognition(ACPR2017), pp.882-887, Nov. 2017
- [30] H.He, E.A. Garcia, "Learning from imbalanced data." *IEEE Trans. knowledge and data engineering*, vol.21, no.9, pp.1263-1284, Sep. 2009.
- [31] T. Jo and N. Japkowicz, "Class Imbalances versus Small Disjuncts," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 40-49, 2004.
- [32] X.Y. Liu, J. Wu, and Z.H. Zhou, "Exploratory Under Sampling for Class Imbalance Learning," *Proc. Int. Conf. Data Mining*, pp. 965- 969, 2006.
- [33] I. Tomek, "Two Modifications of CNN," *IEEE Trans. System, Man, Cybernetics*, vol. 6, no. 11, pp. 769-772, Nov. 1976.

- [34] 川添 由美子, 大山 航, 若林 哲史, 木村 文隆: ”濃度こう配特徴を用いたオンライン署名照合の高精度化”, 電学論C, Vol. 130, No. 12, pp.2142-2149 (Dec. 2010)
- [35] 若林哲史, 鶴岡信治, 木村文隆, 三宅康二: ”特徴量の次元数増加による手書き数字認識の高精度化”, 信学論, D-II, Vol.J77-D-II, No.10, pp.2046-2053, 1994.
- [36] Shi, M., Fujisawa, Y., Wakabayashi, T., and Kimura, F. (2002). Handwritten numeral recognition using gradient and curvature of gray scale image. Pattern Recognition, 35(10), 2051-2
- [37] 古市唯, 大山航, 若林哲史, 木村文隆: “DP マッチングを用いた手書き署名照合の高精度化”: 平成 22 年度電気関係学会東海支部連合大会 (8 月 30-31 日)(N2-2)
- [38] 津田宏治. ”サポートベクターマシンとは何か.” 電子情報通信学会誌 83.6 (2000): 460-466.
- [39] L. Breiman “Random Forests”, Machine Learning, Vol. 45, No. 1, pp. 532, 2001.
- [40] K. Matsuda, W. Ohyama, T. Wakabayashi and F. Kimura, ”Effective Random-Impostor Training for Combined Segmentation Signature Verification,” 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016, pp. 489-494.