

文書鑑定のためのコンピュータ支援システム

上田 勝彦[†] 松尾 賢一[†] 中村 善一[‡]

奈良工業高等専門学校 〒639-1080 大和郡山市矢田町 22

E-mail: [†] {ueda, matsuo}@info.nara-k.ac.jp, [‡] nakamura@elec.nara-k.ac.jp

概要 種々の文書中の筆跡や印影から当該文書の正当性確認や既知文書の作成者との異同判別を行う、いわゆる文書（筆跡）鑑定は、犯罪捜査や裁判における証拠を提供するために頻繁に行われている。この鑑定作業のほとんどは鑑定者の手作業によって行われており、鑑定者の時間的・労力的負担が非常に大きいという問題がある。著者らは、これらの問題点を克服するために、従来の専門家の鑑定作業を支援する対話型のコンピュータ支援システムの開発を目指している。本報告では、このシステムの実現への第一歩として、文書からの筆跡や印影の抽出、および比較対照する2つの文書に存在する同一字種の筆跡を専門家が詳細に分析するための比較チャートの作成を、市販のOCRを援用しながら自動・半自動処理により、効率的に行うシステムを提案する。このシステムをWindows2000のもとでBorland C++Builderを用いて構築した。このシステムでは操作性の向上のために、全ての処理はメニュー選択とマウス操作によって行われる。そして実際の文書を用いて評価実験を行ったところ、従来の鑑定作業と比較して十分有用であることが確認できた。

A Computer-Based System to Support Forensic Examination of Japanese Documents

Katsuhiko UEDA[†] Ken-ichi MATSUO[†] Yoshikazu NAKAMURA[‡]

Nara National College of Technology 22 Yata-cho, Yamatokoriyama, Nara, 639-1080 Japan

E-mail: [†] {ueda, matsuo}@info.nara-k.ac.jp, [‡] nakamura@elec.nara-k.ac.jp

Abstract Handwritten and stamped document examination is frequently used in criminal investigations and justices. In Japan, most of the handwriting and seal imprint examinations have been performed by expert visual inspection. Due to lack of objective measurements and reproducible decisions, it is difficult that traditional methods provide the decision with the persuasion. It is effective that the traditional document examination is supported by an interactive computer system. Based on this idea, we propose a computer-based system to support forensic examination of Japanese documents. This system consists of three major processing modules: (1) Handwriting and seal imprint extraction module, (2) Handwritten character recognition module for classifying handwritings into each character category, (3) Chart preparation module, which displays characters belonging to the same character category in both questioned and known documents for easy comparison. All modules work interactively with user and are menu and mouse driven. The experimental system has been implemented using Borland C++Builder on Windows2000 operating system. Through the experiment, it was confirmed that this system is faster and effective than traditional manual method. We will improve the usability of this system in the future.

1. まえがき

文書中の筆跡や印影から当該文書の正当性の確認や既知文書の作成者との異同を判定するいわゆる文書鑑定は、犯罪捜査や裁判にお

ける証拠を提供するために日常多く行われている。文書鑑定が必要とされる状況から考えて、鑑定結果は極めて重要な意味を持つため、十分

訓練された専門家による慎重な検査が行われている。しかし現状では、この作業は一部計測値を用いる場合もあるが、基本的には目視観察と鑑定者の経験に基づく主観的判断によって行われており、鑑定結果の客観性や鑑定者の時間的・労力的負担も非常に大きいという問題がある。また鑑定者がそれぞれ独自に行っており、鑑定資料や鑑定結果の系統的な蓄積が行われておらず、この分野の客観的・科学的な進歩の妨げになっている。

これらの問題点を克服するためには、従来の専門家の鑑定作業を支援するコンピュータによる半自動的・対話型システムの実現が有効であると考えられる。ヨーロッパを中心として、既にこのような目的のためのシステムが開発されており、実際の業務で使用されている[1]。例えばドイツでは、1983年にFISH (Forensic Information System Handwriting) と呼ばれる支援システムが開発され、その後種々の改良が行われて、欧米の国立法科学研究機関で日常業務に使用された[2]。またオランダでも同様のシステムが開発されている[3]。これらのシステムは、大量の参照用筆跡データに基づいて自動・半自動処理によって、文章、単語、文字の認識機能をはじめ、筆者の国籍、性別、年代などを識別しようとするものであり、文書鑑定技術の向上に貢献している。この他、専門家の鑑定作業を支援するために、比較対照する文書の同一単語・文字を並べて表示する筆跡比較チャート作成システムも開発されている[4]。

日本でも、筆跡や印影の鑑定に関する研究は古くから行われており[5]、コンピュータによる筆跡・印影特徴の計測の試みもある[6],[7]。また犯罪捜査や裁判などの実務において証拠提供のために日常相当数の鑑定作業が行われている。しかし前述のようにその作業のほとんどは、鑑定者の手作業によって行われており、欧米のような支援システムの開発研究の例は

ない。また日本の文書・筆跡は欧米のそれとは異なるため、欧米の研究成果を直接日本の文書には直接適用できない。

他方、文書の真偽判別を自動的に行おうとする、いわゆる署名照合や印鑑照合の自動化に関する研究は、それらを要求する分野が広いこともあり、世界的に多くの研究が行われている[8],[9]。しかし照合の問題は、ある限られた条件のもとでの1対1比較であり、短時間に自動的に正しい異同判別結果を出力するのが重要である。これに対して、鑑定の問題は制限不可能な種々の資料を様々な観点から詳細に検査を行って、異同の判定を行う。ここでは判定結果の正しさはもちろん、その根拠を資料に基づいて示すことが非常に重要である。従って文書鑑定の問題には全自動処理は馴染まないと思われる。

以上の諸背景を踏まえて、著者らは、日本の文書を対象として、専門家の文書鑑定作業を支援する対話型システムの開発を行っている。本報告では、このシステムの実現への第一歩として、文書からの印影・筆跡の抽出、および比較対照する2つの文書に存在する印影や同一字種の筆跡を専門家が詳細に分析するための比較チャートの作成を、自動・半自動処理により、効率的に行うシステムを提案する。

2. 提案システムの概要

このシステムは、図1に示すように、1) 印影・筆跡抽出、2) OCR 支援による筆跡の各字種カテゴリへの分類、3) 比較チャート作成、の3つの主な部分から構成される。本システムが対象とする文書は、手紙やメモなどの無背景の一般的な文書、および小切手や領収書などのように種々の背景模様を有する文書である。

まず処理対象文書は、イメージスキャナから解像度 300dpi (対象文書により可変)、階調数 8bit の濃淡画像または RGB 各 8bit のカラー

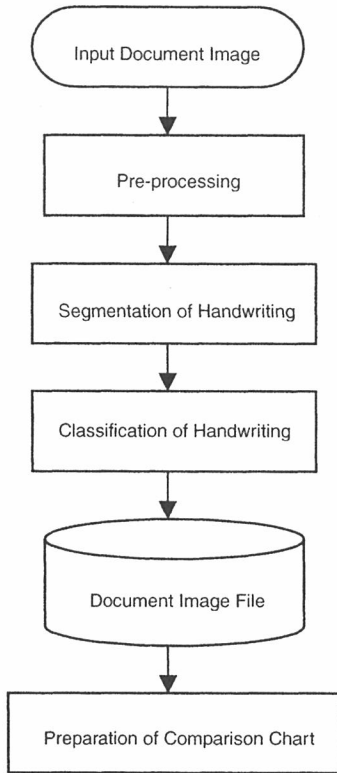
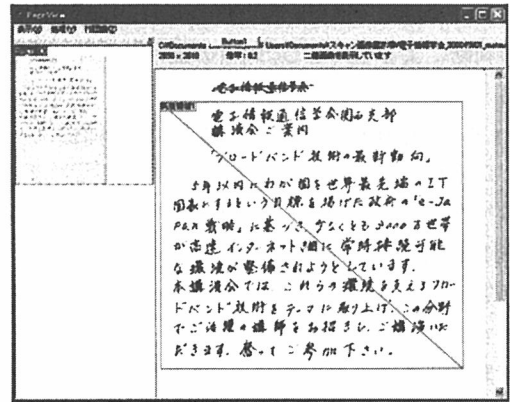


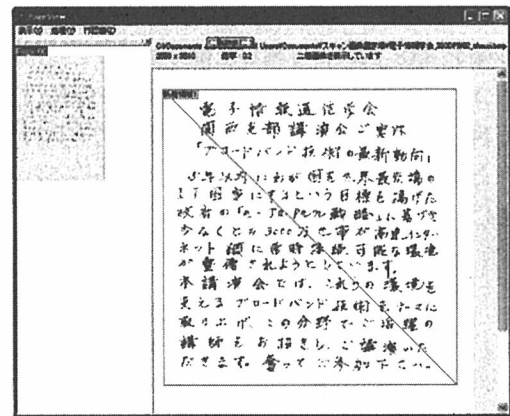
図1. 提案システムの処理の流れ.

画像としてシステムに入力される. 文書入力ウィンドウの例を図2に示す. 同図は, 二人の筆者による同一内容の文書をそれぞれ入力するウィンドウの例である. 図中の矩形枠は, 入力画像から必要な処理対象領域を選択するための領域を示しており, 使用者のマウス操作により指定される. 以後の処理においては, この領域が対象となる.

入力された文書画像は, 必要に応じて雑音や罫線などの除去処理が行われる. 次いで比較対照される印影や筆跡が個別文字ごとに文書画像から切り出される. 印影や少数の筆跡の場合には, 使用者によって適当な識別コードを付された上, 文書ファイルに保存されるが, 手紙や遺言状などのように多数の筆跡からなる場合は, OCRを援用することにより, 対象筆跡を



(a) 筆者Aの文書入力



(b) 筆者Bの文書入力

図2. 文書入力ウィンドウ.

各字種ごとに分類して, 文書ファイルに保存する.

最後に, 2つの文書から切り出された筆跡とその字種情報から双方の比較対照文書に存在する同一字種の筆跡画像を提示する.

3. 筆跡・印影の抽出

本研究では, 種々の文書を取り扱う必要があるため, 処理対象となる可能性のある文書を, それらの性質を考慮して2つのグループに分類し, 各グループに適した抽出方式を開発した. すなわち銀行小切手や領収書など複雑な背景模様と色彩を有する文書, および白紙や単純な罫線を有する用紙に単色で書かれた一般的な

文書の2種類である。処理方式の選択は使用者の判断によるものとし、両者の混用も可能である。

3.1 小切手画像からの記入文字・印影の抽出

この種の文書は、文字、印影、背景がそれぞれ特有の色を有することを利用して、文字と印影を抽出する。本研究で提案する抽出処理は、図3に示すように、2つの処理部分からなる。まず文書の背景領域を除去した後、記入文字と印影をそれぞれ抽出する。

背景領域の除去には、文字・印影領域と背景領域の輝度変化の相異を利用する。まず小切手画像を8X8画素に分割した部分領域ごとに離散コサイン変換(DCT)を施し、得られたDCT係数行列を、部分領域中の輝度変化の周波数特徴とする。標準的な背景の周波数特徴を得るために、画像の四隅の部分領域を平均化した背景

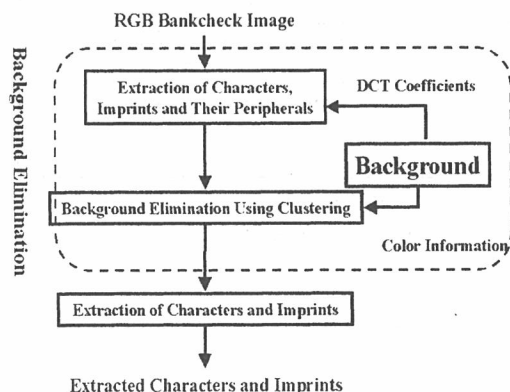


図3. 背景除去処理の流れ.

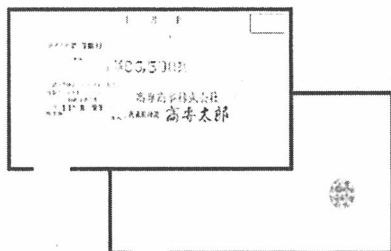


図4. 文字と印影の分離抽出結果の例.

標準画像を作成し、DCT係数行列を計算する。そして背景標準画像のDCT係数行列と、全ての部分領域との類似度を求めて、部分領域の類似度が、適当なしきい値Tよりも大きい領域を背景領域と見なして除去する。

次いで背景除去画像を、RGB色空間上に射影して代表色抽出を行い、K平均クラスタリングにより文字と印影を分離して抽出する。処理結果の一例を図4に示す。以上の処理は全て自動的に行われるが、最終的に必要な文字・印影の切り出しはマウスによる領域指定により行われる。

3.2 一般文書からの筆跡抽出

処理の簡単化のため、まず入力文書画像を白黒2値の2値画像に変換する。最初に各文字行の切り出しを行う。文書が横書きの場合は、この2値画像の垂直軸上への射影ヒストグラムを求める(図5)。文字行間で重なりがなければ、図5のようにヒストグラムは、各行に対応したピークを持つ。従って、ヒストグラムのピーク間で頻度が0になる点が行間に対応するとして、この垂直領域を文字行として切り出す。この簡単な方法では、文字行に重なりがある場合、複数行が1行として切り出されることがある。この場合は、使用者がマウス操作で自由曲線により行単位に分離する。

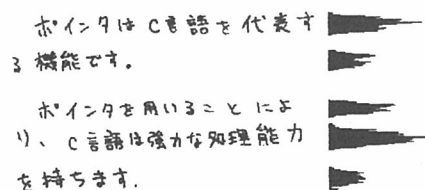


図5. 垂直軸方向への射影ヒストグラムを用いた文字行の切り出し.

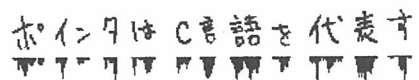


図6. 水平軸上への射影ヒストグラムを用いた文字切り出し.



図7. 文字行と個別文字の切り出し結果例.



図8. 文字認識ウィンドウの例(左:入力文字, 右:認識結果).

次いで切り出された各行の水平軸上への射影ヒストグラムを基に各文字(筆跡)を切り出す(図6)。しかし互いに接触した文字は、行の切り出しと同様、この方法では正確に切り出せないため、自動分割の後、修正が必要な場合は、マウス操作で自由曲線による分割が可能である。縦書き文書の場合も射影方向が異なるだけで同様の処理である。行・文字切り出しウィンドウ上の切り出し結果の一部を図7に示す。

4. 筆跡の字種分類

日本字の筆跡鑑定は、同一字種の筆跡について詳細な比較を行うのが原則である。このため比較対照する2つの文書から同一字種の筆跡

を検索する必要がある。現在この作業は手作業で行われているが、大きな時間と労力を要している。

一方最近、手書き文字認識技術の進歩により、OCRの認識性能が向上している。そこで本研究では、筆跡の各字種カテゴリへの分類のために市販の手書き文字OCRライブラリ(リコー、HocrKanji3.20)を援用することにした。切り出された個別筆跡がOCRによって認識される。認識結果は、シフトJISコードで出力されるが、同時に当該文字の形でディスプレイにも表示される。使用者は、認識結果を確認し、もし修正が必要な場合は、OCRから出力される第10位までの候補文字の中から正しい文字を選択する(図8)。この候補に正しい文字がない場合は、キーボードから該当文字を入力する。

最後に各筆跡画像が字種コードとともに文書ファイルに保存される。なお先に述べたように、行切り出しからOCRによる字種分類処理までは処理の容易さとOCRの制限から、全て2値画像を対象としたが、専門家による詳細な検査のために、保存される筆跡画像は、2値画像に加えて元のカラー画像や白黒濃淡画像も同時に保存する。具体的には、各文書ごとに一つのフォルダを用意して、このフォルダに元のカラー・白黒濃淡文書画像、個別筆跡のカラー画像および白黒濃淡画像、2値画像を一括してビットマップファイルとして保存する。これらの画像ファイルは、図9に示すインデックス・ファイルによって管理される。

5. 比較チャートの作成

比較する2つの文書フォルダ内のインデックス・ファイルを参照することにより、これらの文書から抽出された印影や筆跡とその字種・出現順情報に基づいて、双方に存在する同一字種の筆跡画像を提示する。図10に、比較

```

[C_00000] ;出現順位
Character=講 ;認識結果
BinaryImageFile=CB_00000.bmp
GrayImageFile=CG_00000.bmp
ColorImageFile=CC_00000.bmp
[C_00001]
Character=演
BinaryImageFile=CB_00001.bmp
GrayImageFile=CG_00001.bmp
ColorImageFile=CC_00001.bmp
[C_00002]
Character=会
BinaryImageFile=CB_00002.bmp
GrayImageFile=CG_00002.bmp
ColorImageFile=CC_00002.bmp
.
.
.

```

図9. 筆跡画像のインデックス・ファイル.

チャート作成ウィンドウの一例を示す. この例では, ディスプレイの解像度の関係で, ウィンドウ内に比較チャートの一部しか表示されていないが, 画面スクロールまたはチャートの解像度を下げる事により, チャート全体を概観することも可能である. さらに提示された印影・筆跡画像は, 比較のためにマウスでドラッグすることにより自由に移動させることができ, 両者の重ね合わせ比較も可能である.

現在, 鑑定者がこのチャートから鑑定資料を作成するために, チャート上の特定の比較対照筆跡を指定することにより, 筆跡画像上に鑑定者の注釈や意見を記入することができるエディタを作成中である.

6. 結果

以上の考えに基づいた実験システムを, Windows2000のもとでポーランド C++Build-

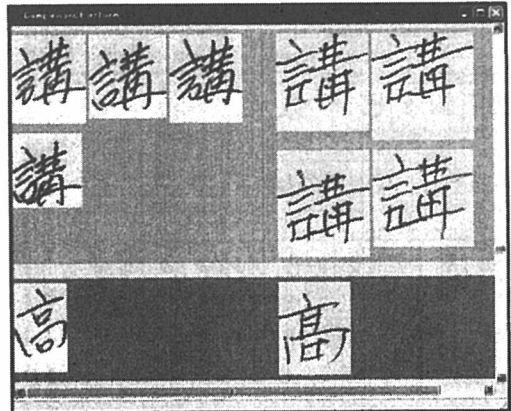


図10. 作成された筆跡比較チャートの例 (左: 文書 A (図 2(a)), 右: 文書 B (図 2(b))).

der を用いて構築した. そして本システムを評価するために, 図2に示した A4 サイズの横書き一般文書 2 枚から比較チャートを作成する一連の処理を行った. 2 枚の文書に含まれる総文字数は 408 文字 (204 字/枚) であった. その結果を表1に示す. 一連の処理に要した時間は約 19 分で, 手作業での処理に要する時間に比べて十分高速である. また OCR の援用効果についても表1から明らかのように, 94% は候補文字からの選択で処理できることから, 十分有用であることが確認できた. 特に筆跡鑑定のための手掛かりを多く含むとされる漢字に限定すると, 正認識率 (第1候補) は 87% であった. しかし OCR の認識率は, 手書き文字の品質に大きく依存するため, さらに多くの文書を用いた評価が必要である.

表1. 評価実験の結果

処理時間	19分
総文字数	408文字
(1) OCR 第1位認識文字数	267文字 (65%)
(2) OCR 第2~10位候補文字数	115文字 (28%)
(3) キーボード入力文字数	26文字 (6%)
(1) + (2)	382文字 (94%)

7. あとがき

本報告では、日本の文書を対象として、専門家の文書鑑定作業を支援するシステムとして、鑑定作業の最も基本的な部分である筆跡比較チャートを作成するシステムを提案し、その有効性を確認した。今後、本システムの有効性について、実務での使用を通じてさらに詳細な評価を行う予定である。またこのシステムに筆跡特徴の計測機能を付加して、鑑定のためのさらに詳細な参考情報を提供できるようにすることも今後の重要な課題である。

謝辞: 本研究に関して、実務分野の立場から有益なご助言を頂く兵庫県警察本部科学警察研究所の小林邦久主任研究員並びに下山昌彦博士に謝意を表す。また卒業研究課題として本システムのプログラム開発に協力してくれた本校専攻科電子情報工学専攻学生 岡本圭司君に深謝する。

なお本研究の一部は、日本学術振興会科学研究費補助金（基盤(C)13680500）の補助により行われた。

文 献

- [1] P.E.Bair, Image Processing of Forensic Documents, Proc. Of 3rd International Conference on Document Analysis and Recognition, pp.1-4, 1995.
- [2] M. Philipp, Expected Future Developments in the Forensic Information System Handwriting (FISH), 4th European Conference for Police and Government Handwriting Experts, 1994.
- [3] W.C. de Jong et al., Computer-Aided Analysis of Handwriting -the NIFO-TNO approach-, 4th European Conference for Police and Government Handwriting Experts, 1994.
- [4] G.Holcombe et al., Computer Aided Creation of Handwriting Comparison Charts in the Forensic Examination of Questioned Documents, Handwriting and Drawing Research, pp.493-507, IOS Press, Amsterdam, 1996.
- [5] 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.
- [6] 吉田公一他, 文字の外形を利用した筆跡の分類に関する基礎的研究, 科学警察研究所報告法科学編, Vol.40, No.1, pp.24-31, 1987.
- [7] 吉田公一, 印章鑑定に関する基礎的研究(1),

科学警察研究所報告法科学編, Vol.38, No.3, pp.1-5, 1985.

- [8] 吉村ミツ, 吉村功, 筆者認識研究の現段階と今後の動向, 信学技報, PRMU96-48, pp.81-90, 1996.
- [9] 上田勝彦, 印鑑照合の自動化技術, システム/制御/情報, Vol.35, No.7, pp.423-430, 1991.