

ビデオによる古文書の効率的画像入力法と自動接続処理

柴山 守

(大阪国際大学経営情報学部)

星野 聰

(京都大学大型計算機センター)

ビデオ動画像による古文書の入力とワークステーション(WS)上での入力画像の自動接続・復元処理について述べる。分割入力された古文書の部分画像を瞬時に接続・復元するために、入力画像の濃度ヒストグラムによる特徴抽出とストロークに依存した量子化法による高速マッチング法を工夫した。さらに、本高速マッチング法を組み込んだビデオ動画像による古文書入力システムの開発とビデオ動画像から一定時間間隔でサンプリングした静止画像を動画像の表示速度に対応して自動接続する実験結果について報告する。

An Effective Automatic Connection of Ancient Documents by Video Image

Mamoru SHIBAYAMA

Satoshi HOSHINO

Faculty of Management and
Information Science,
Osaka International UniversityData Processing Center,
Kyoto University3-50-1 Sugi, Hirakata,
Osaka 573-01, JapanYoshida Honmachi, Sakyo,
Kyoto 606, Japan

An effective image input method for ancient documents using video camera by connecting images with overlapped parts is described. In order to reduce the necessary time for matching two overlapped images, the features have been extracted from the histogram for each character string in overlapped parts. And a high-speed matching has been devised based on the quantization derived from the strokes of character on the documents. In the result of the experiment, it was found that the time spent for the connection was about one second.

【1】はじめに

古文書・古絵図等の史料の計算機入力では、主に画像処理を目的にした計算機システムに接続されるイメージ・スキャナやドラム・スキャナ、高精細な据置型CCDカメラ等の周辺装置が用いられ、概してその規模も大きい。近年、ビデオカメラやVTR等の高性能化と低価格化が進み、一方ワークステーション（WS）やパソコンにおいてビデオ・ボードやフルカラー表示を可能にするハードウェアや動画画像を扱うソフトウェアが数多く出現し、小規模でパーソナルな画像処理システムが容易に実現できるようになった。特に、民生用のビデオカメラの利用によって古文書等が直接にカラーのデジタル画像として計算機入力できれば入力作業は大いに軽減され、且つ経済的である。しかし、一般に民生用のビデオカメラによる画像入力においては、充分満足する解像度や画質を得ることは難しい。

古文書・古絵図は、一般的に巻紙のように横に長い、あるいは掛け軸のように縦に長く、また手書きであり、微妙な濃淡表現や色調も重要である。ディスプレイ上でこれらの古文書等の画像データがカラー画像で、且つ充分判読できる解像度や画質、色調で扱えるようにするためには、分割入力が考えられる。筆者は、古文書等のビデオカメラによる分割入力を、WS上で自動的に接続・復元する実験を行ってきた[1]。これらの入力法を実現するためには、入力された部分画像に対する一般的な画像の変換や画質改善の処理に加えて、画像データの接続・接合処理（原画像復元）が必要になる。画像データの接続・接合処理は、隣接する画像の重複部分の画像を利用して、位置合わせと色合わせを行った後、隣接する画像を接合するが、（1）位置合わせ及び座標系の統一、（2）濃度補正、（3）接合点探索、（4）接合点周辺の平滑化などの処理が必要とされる[2]。さらに、ビデオ動画画像を再生・表示しながら、動画画像からサンプリングした静止画像をサンプリングの時間間隔内で自動接続し、復元するためには、コンピュータに負荷をかけないような簡単なアルゴリズムで古文書の文字列特徴を抽出し、マッチングする機能が必要になる。

本報告では、縦横共に長い古文書をビデオによるカラー画像として分割入力し、WS上で隣接画像の重複部分に基づく自動接続・復元を考える。ここで分割入力とは、動画画像から一定時間間隔でサンプリングした静止画像をも含む。接続処理においては、入力画像から得られた画像データを2値化処理し、重複部分の文字列の特徴量による量子化に基づいた特徴抽出と高速マッチング法について述べ、従来から採用してきた残差逐次検定法（SSDA法）とピラミッド探索法にヒストグラムによる特徴抽出を併用した方法との比較を行う。また、実際にビデオカメラで撮影した古文書のビデオ動画画像から、一定時間間隔でサンプリングした静止画像を自動接続し、復元するシステムについて報告する。

【2】古文書・古絵図画像の接続処理

ビデオカメラ、あるいはスチルカメラなどを入力装置として古文書の入力を行うには、充分な解像度を得るために古文書の分割入力を、入力後に計算機上で隣接画像の接合・接続処理を行い、復元しなければならない。この隣接画像の接合・接続のためのマッチング手法として、筆者は隣接画像の重複部分内におけるテンプレート画像と対象画像の類似度を調べ、接合位置を求めるテンプレートマッチングを用い、古文書の特徴に従ってテンプレート画像の探索範囲を減少することによってマッチングの高速化を行ってきた[1]。この実験システムによる4分割入力した古文書画像のマッチング結果とスクロール機能を付加したウィンドウの表示例を図1に示す。実験に使用した古文書は、茨城県土浦における江戸期（天保年間）の古文書（筑波大学岩崎宏之教授提供）である。

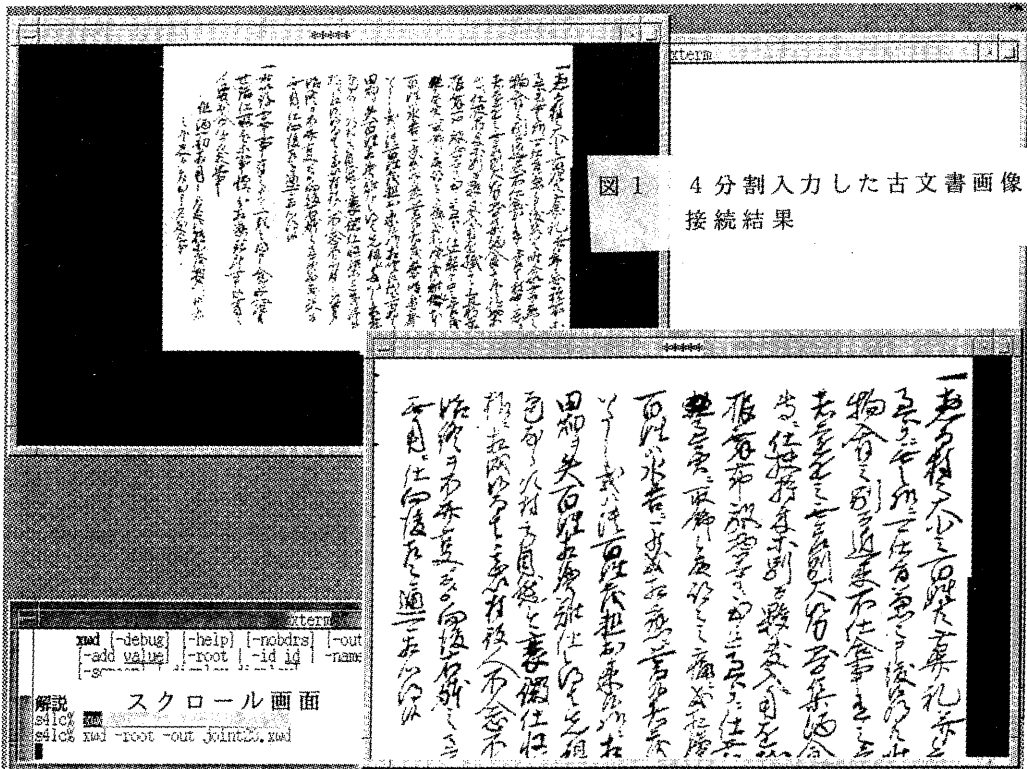


図 1 4分割入力した古文書画像の接続結果

このマッチング法の基本的操作は、つぎのとおりである。

[1] 残差逐次検定法 (SSDA法: Sequential Similarity Detection Algorithm)

重ね合わせの基準になる $m \times n$ 画素のテンプレート画像を、それより大きい $M \times M$ 画素の対象 (入力) 画像内の探索範囲 $(M-m+1) \times (M-m+1)$ 上で動かし、各座標におけるテンプレート内残差和が最小になるようなテンプレート画像の左上位置を求めて重ね合わせが行われたとする。

[2] ピラミッド探索法

画像の重ね合わせにおいて、バックトラックを必要としない探索としてピラミッド構造を採用すると、粗い解像度の最上層で概略重ね合わせを行ってから、密の解像度で詳細重ね合わせを行い、照合時間が短縮できる[3]。このピラミッド構造の画像に対して、SSDA法を用いたマッチングが効果的であるが、この方法で解像度を落とすすぎると誤探索の可能性が生じる。誤探索を防ぐために、また対象画像とテンプレート画像のサイズの大きい場合に演算時間を短縮するために、実験システムではつぎに示す特徴量空間とピラミッド構造とを用いた画像の重ね合わせ法を用いた。

[3] ヒストグラムによる特徴抽出

文書の左右部分画像の接続には、縦書きの特徴から形状の類似性によるマッチングを行うため、特徴抽出の機能を付加している。これを図2(a)に示す。この結果から、図2(b)に示すように、テンプレート画像は、入力画像の最右端から列として独立した部分 $W_T \times H_T$ 画素である。また、基準画像内の探索範囲は、画面の左端から探索して最初に見つかった列である。探索範囲は動的に決定され、ヒストグラムによって判定される列幅と高さ H (実験システムでは 200画素) の面積となる。

ここで文書の各列幅を W_{i1} 、列間の空白幅を S_{i1} とすると、ヒストグラムによる特徴を用いない場合、文字列が n 列存在するとき、X軸方向の幅 W は

$$W = \sum_{i=1}^n (W_{ii} + S_{ii})$$

となり、テンプレート画像の探索範囲は

$$\sum_{i=1}^n (W_{ii} + S_{ii}) - (W_T - 1)$$

となる。ここで、 W_T はテンプレート画像の列幅であり、 $W_T \leq W_{ii}$ である。

ここで、ヒストグラムからの特徴をマッチングに用いると、テンプレート画像の探索範囲は、つぎのようになる。

$$\sum_{i=1}^n [W_{ii} - (W_T - 1)] \quad (3)$$

この結果、抽出した文字列の特徴を利用することにより、X軸方向の探索範囲を

$$\sum_{i=1}^n S_{ii} + \sum_{i=1}^{n-1} (W_T - 1) \quad (4)$$

減少することができ、マッチングの高速化に有効であることを示した。

ビデオ動画像から一定時間間隔でサンプリングした静止画像を再生速度に遅れることなく隣接画像を接続するためには、サンプリング間隔の時間内に入力画像の特徴を抽出し、マッチングを行わねばならない。前節でのSSDA法にピラミッド探索法を併用し、テンプレート画像の探索範囲を減少した手法では実時間で画像の特徴抽出やマッチングを行うことは不可能である。

【3】量子化による高速マッチング法

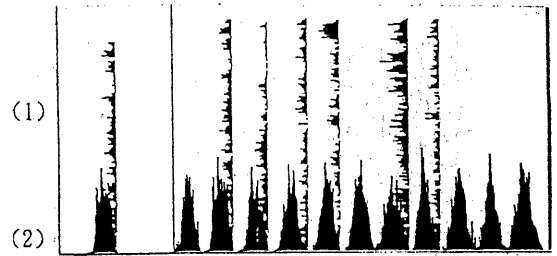
ビデオ動画像の表示速度に対応した実時間の高速マッチングを実現するためには

- (1) 古文書画像の各文字列毎の一意的特徴を抽出する。
- (2) 抽出した特徴から文字列・数値列等のパターンに変換する。
- (3) 探索範囲を出来る限り縮小し、パターン・マッチングを効果的に行う。
- (4) 位置合わせを瞬時に行って、表示する

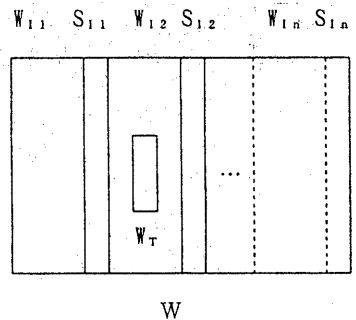
など、簡単なアルゴリズムで、CPUへの負荷を避けねばならない。以上のような理由から、つぎに示すような古文書の文字列のストロークに注目した高速マッチング法を提案する。

[1] ストロークによる特徴量抽出と高速マッチング

(1) 古文書に記載された文字列の全般的な筆跡、特にストローク方向に注目して、特徴を抽出する。これを図3に示す。図1に示した古文書は、図3の「大」に示すように左端から右端にかけて右上がりのストローク(a)、上部中心から左端に降ろすストローク(b)、中心から右端へのストローク(c)の水平方向に対する各々の角度に注目し、(2) ストロークに沿ってしきい値以上の画素を探



(a)文字列のヒストグラム



(b)テンプレート画像と探索範囲

図2 ヒストグラムによる特徴抽出

索範囲のすべてについて計数し、数値化、すなわち量子化する。例えば、左端から右端にかけて右上がりのストローク方向の量子化は、図4に示すようになる。テンプレートになる部分画像 $W_T * H_T$ の任意の高さ h に対する特徴量を Q_{Th} とするとき

$$Q_{Th} = \sum_{w=1}^{W_T} P_{Th} (\tan \theta_T \cdot w) \quad (5)$$

ここで、 θ_T は、左端から右端にかけて右上がりのストロークに対する角度、画素値は $P_{Th} (\tan \theta_T \cdot w) = \{0, 1\}$ 、 $h = 1, 2, \dots, H_T$ である。テンプレートになる画像は、基準画像の最左端にある文字列と仮定する。

入力画像 $W_I * H_I$ の任意の高さ h に対する特徴量を Q_{Ih} とするとき

$$Q_{Ih} = \sum_{w=1}^{W_I} P_{Ih} (\tan \theta_I \cdot w) \quad (6)$$

ここで、画素値は $P_{Ih} (\tan \theta_I \cdot w) = \{0, 1\}$ 、 $h = 1, 2, \dots, H_I$ である。入力画像は、一定時間間隔でサンプリングした静止画像とする。

(3) 高速マッチングは、テンプレート画像の特徴量と入力画像の各文字列の特徴量を順にマッチングし、残差が最小になるとき、マッチングされたとする。i 行目における残差を R_i とすると

$$R_i = \sum_{h=1}^{H_T} | Q_{Th} - Q_{Ih} | \quad (7)$$

ここで、 $W_T \leq W_I$ 、 $H_T \leq H_I$ 、 $\theta_T = \theta_I$ 、 $i = 1, 2, \dots, n$ である。

以上のように、このマッチング法は、古文書の各文字列を特徴量 Q_{Ih} の量子化した数値列で表し、テンプレートとなる画像の特徴量 Q_{Th} と比較しマッチングを行うもので、これはSSDA法におけるテンプレート画像の面積に等しい部分画像との残差を求める1回の作業量に等しい。すなわち、SSDA法では、各文字列において式(7)に示す作業を

$$(W_I - W_T + 1) \cdot (H_I - H_T + 1)$$

回繰り返さねばならない。

[2] 高速マッチングによる隣接画像の接続実験

特徴量の量子化による高速マッチング法を利用して、図1に示した古文書の上半分にあたる左右の隣接画像を接続する実験を行った結果を図5に示す。なお、高速マッチング法における文字列の特徴抽出は、図3(a)及び図4の方法のみを採用している。

図5において、ウィンドウの左上部の[File1]、[File2]で左右各々の部分画像を読み込み、[Start]ボタンのクリックによって接続処理を開始する。

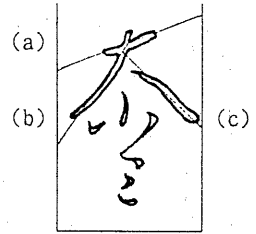


図3 文字のストロークによる特徴

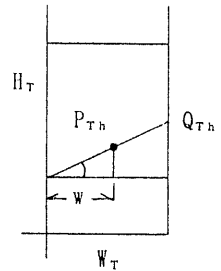
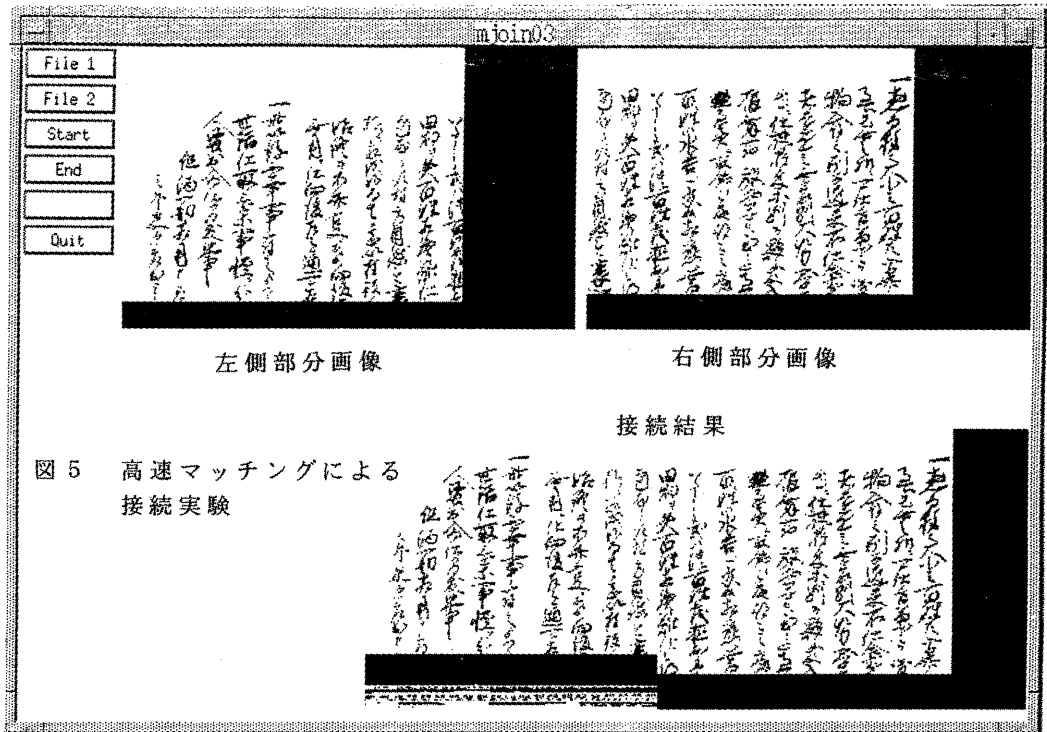


図4 右上がりのストローク方向



S S D A法とピラミッド探索法にヒストグラムによる特徴抽出を併用した図1の方法と高速マッチング法による接続に要した所要時間の比較は、表1のとおりである。

表1 隣接画像の接続に対する所要時間の比較

方式	S S D A法	高速マッチング法
所要時間	6 sec	1 sec以下

ワークステーションNWS-3150 (主記憶16MB, 37MIPS) 使用

実験結果から、高速マッチング法が優れていることが示された。

【4】ビデオによる古文書画像入力システム

前節で述べた量子化による高速マッチング法を組み込んだビデオによる古文書画像入力システムを試作した。本システムのシステム構成を図6に示す。図6において、ビデオ機器の動画像を取り込み、表示する機能を備えたビデオボード (NWB-254) を搭載したワークステーションNWS-3470にビデオ機器とのインターフェースVbox CI-1000を接続し、VboxのLANC端子とビデオカメラCCD-TR900のLANC端子を接続している。また、ビデオカメラからのビデオ出力信号をビデオボードに入力している。

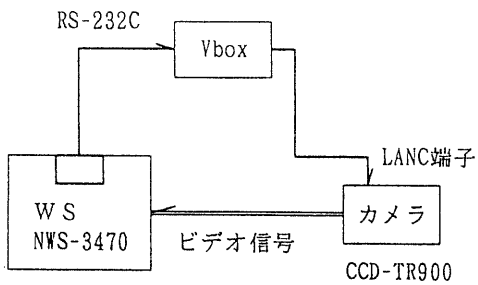


図6 ビデオ画像入力システム

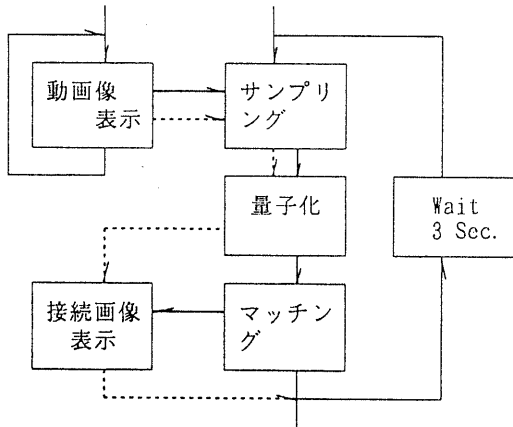


図7 ビデオ画像入力システムのプログラム構造

入力対象である古文書画像は、ビデオカメラで撮影しておきビデオテープを図6のビデオカメラに再生モードでセットしておく。

本システムのプログラム構造は、図7に示すようにビデオカメラからビデオ動画をを入力し、表示する部分と一定時間間隔でフレームメモリ上の動画を静止画像として取り込み、古文書の各文字列の特徴を抽出して、高速マッチングを行う部分からなる。プログラムは、ビデオ画像を扱うためのWS上のソフトウェアXlib、Motif、Vboxを制御するためのVISCAコマンドをC言語を用いて記述している。本システムによる古文書の接続・復元結果を図8(a)に示す。

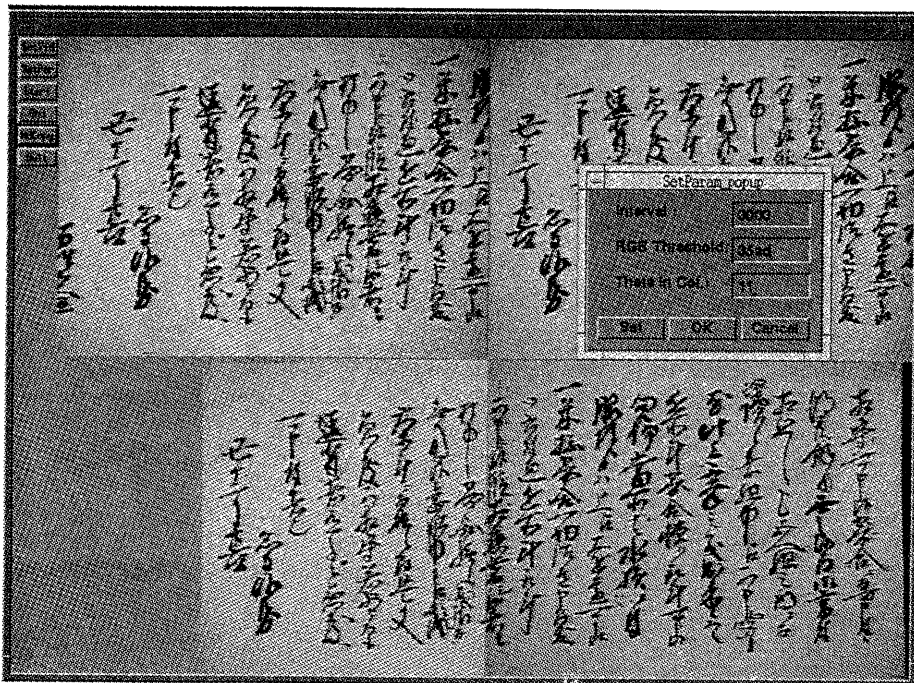


図8 古文書画像
入力システム

(b)パラメタ設定
ウィンドウ(内側)

(a)接続・復元結果
ウィンドウ(外側)

入力操作は、まず図 8 (a) の左上部分にある[SetVid]ボタンをクリックし、ビデオカメラの頭だしなどの設定を行う。つぎに、[SetPar]をクリックし、図 8 (b) に示すウィンドウで

- (1) 動画をサンプリングする時間間隔 (単位: ミリ秒)
- (2) 量子化の際の計数の対象とする各画素値のしきい値
- (3) マッチングのための右上がりストロークの角度 θ

を設定する。引き続いてビデオ動画が左上部分に表示される。[Start]ボタンをクリックすることによって自動接続・復元処理が開始され、図 8 (a) の左上部分に表示されたビデオ動画が取り込まれ、表示される。自動接続・復元処理が開始された先頭の画像が基準画像 (テンプレート画像) の対象となり、この画像中の最左端の文字列をテンプレート画像として抽出する。この後、一定時間間隔 (本実験では 3 秒) でサンプリングされる静止画像 (図 8 (a) の右上部分) の各文字列との高速マッチングを行って、接続・復元結果が図 8 (a) の下半分に瞬時に表示される。

【5】おわりに

分割入力した古文書の部分画像のマッチング処理において、古文書の各文字列のストロークの特徴に注目した量子化と高速マッチング法の方式について述べ、残差逐次検定 (SSDA) 法とピラミッド探索法の併用による接続処理との結果について比較した。また、量子化法による高速マッチング機能を付加したビデオ動画による古文書入力システムを試作した。実験結果から、古文書の接続処理において筆跡のストロークの特徴量を量子化することによって特徴抽出が行えること、またアルゴリズムが簡単であることから、ビデオ動画から一定時間間隔でサンプリングした静止画像の接続処理に対しても動画の再生速度に充分追従できることが判った。

特にビデオ動画からのサンプリングによる入力の場合、動画の表示速度に追従させようとするとテンプレート・マッチングでは不可能であり、サンプリング間隔の時間内における特徴抽出とより精度の高いマッチングが必要とされる。このマッチング時間の短縮のためにストロークに基づく特徴抽出の方法を改善し、マッチングの性能向上を図ること、また高速マッチングにおいて並列処理を考えねばならない。さらに、量子化の際に必要な濃度のしきい値設定、隣接点周辺の濃度の平滑化、用紙の端点の自動認識が必要である。これらは今後の課題である。また、縦横共に長い古文書・古絵図のビデオ画像入力には、被写体をカバーするためにカメラの撮影方向を変更したパンニングに対する自動認識の機能と自動接続の処理を付加しなければならない。

本研究は、科学研究費補助金試験研究 (B) 「東洋学研究における研究者用マルチメディア情報 CD-ROM の実用化」の下で進めた。

【参考文献】

- [1] 柴山 守、星野 聰: ビデオによる古文書の効率的画像入力法、情報処理学会「人文科学とコンピュータ研究会」研究報告、92-CH-14、1992.6
- [2] 高木幹雄、下田陽久編: 画像解析ハンドブック、東京大学出版会、pp. 461-467、1991.10
- [3] 曾根光男、寺田聡、板内正夫: 特徴量空間とピラミッド構造とを用いた画像の高速重ね合せ法、電子情報通信学会誌 D Vol. J71-D, No. 1, pp. 102-109, 1988.1
- [4] 星野 聰: ビデオ機器を用いた画像データベースの作成、京都大学大型計算機センター研究開発部研究発表報告集、第 8 号