

ペン先カメラ画像からの手書きパターンの抽出

伊東 克啓[†] 内田 誠一^{††} 岩村 雅一^{†††} 大町真一郎^{††††} 黄瀬 浩一^{†††}

[†] 九州大学大学院システム情報科学府 〒 819-0395 福岡市西区元岡 744

^{††} 九州大学大学院システム情報科学府 〒 819-0395 福岡市西区元岡 744

^{†††} 大阪府立大学大学院工学研究科 〒 599-8531 大阪府堺市学園町 1-1

^{††††} 東北大学大学院工学研究科 〒 980-8579 仙台市青葉区荒巻字青葉 6-6-05

E-mail: uchida@is.kyushu-u.ac.jp

あらまし 本研究では、日常の手書き内容を保存できるライティング・ライフ・ログの実現のための要素技術として、ペン先付近にカメラを搭載し、手書きパターンをリアルタイムに取得・復元する方法を提案する。実際に手書き内容を得るためには、動画の各フレームについてビデオモザイクを行う必要がある。すなわち各フレーム画像について位置合わせをしながら統合し、パターン全体を復元する必要がある。本手法の特徴は、紙指紋と呼ばれる紙面の微細構造を積極的に利用する点にある。具体的には紙面より SURF 特徴点を抽出し、それに基づいてフレーム間の特徴点対応を求める。この対応関係を基にフレーム間の射影変換行列を求めて位置合わせを行う。これを全フレームについて行うことで最終的に 1 枚の手書きパターン画像を生成する。幾つかの実例を通して、本手法の基本性能を検証した。キーワード ビデオモザイク、手書きパターン、SURF、RANSAC、紙指紋

Recovering Handwritings via Pen-tip Camera

Katsuhiro ITO[†], Seiichi UCHIDA^{††}, Masakazu IWAMURA^{†††}, Shinichiro OMACHI^{††††}, and
Koichi KISE^{†††}

[†] Graduate School of Information Science and Electrical Engineering, Kyushu University

^{††} Faculty of Information Science and Electrical Engineering, Kyushu University
Motooka 744, Nishi-ku, Fukuoka-shi, 819-0395 Japan

^{†††} Graduate School of Engineering, Osaka Prefecture University, Sakai-shi, Osaka, 599-8531 Japan

^{††††} Graduate School of Engineering, Tohoku University, Sendai-shi, 980-8579 Japan

E-mail: uchida@is.kyushu-u.ac.jp

Abstract Toward realization of “writing-life-log”, a camera-based handwriting pattern acquisition system is proposed. The camera is attached around the tip of a popular pen. It captures frame images around the pen tip continuously. Our problem is video-mosaicing of those frame images by perspective registration of consecutive frames. A key idea is to use precise structure of paper surface, called paper fingerprint, for the registration. Specifically, perspective transformation is estimated by using correspondence of SURF feature points extracted on paper surface. Since the precise structure can be captured stably as the SURF feature points from the pen-tip camera, thus it is possible to expect accurate registration of video frames.

Key words video mosaicing, handwriting, SURF, RANSAC, paper fingerprint

1. ま え が き

「紙とペン」による手書き(手書きコンテンツ)は我々人類にとって最も歴史ある情報の生成・記録メディアである。ペーパレス化が叫ばれ、様々な情報革新がありながら、未だに紙の手帳を愛用する人も多く、講義メモも紙のノートに取る人も多い。

そこでこうした手書きコンテンツをデータベースに取り込むことができれば、日常の手書き内容を保存できるライフログ、謂わばライティング・ライフ・ログ(“writing-life-log”)を実現できる。

本研究では、ペン先にカメラを搭載することで、手書きパターンをリアルタイムに取得することを目的とする。こうして

得られた手書きパターンを画像として自動的にデータベース登録できれば、上述のライティング・ライフ・ログを実現できる。

実際に手書き内容を得るためには、動画の各フレームに断片的に撮影された手書きパターンからその全体像を復元する必要がある。これはいわゆるビデオモザイクング [1] と呼ばれる技術である。具体的には、現在フレーム画像と直前フレーム画像との隣接する 2 フレーム間で特徴点の対応関係を求め、それに基づいてフレーム間の姿勢変化をを求める。この姿勢変化を全ての隣接フレームに渡って求めれば、全フレーム分の姿勢変化系列が把握できる。そしてそれらを用いて各フレーム画像を貼り合わせることで、手書きパターンの全体像を復元した 1 枚の画像を生成できる。

本報告では、各フレーム画像から特徴点を検出する際、紙指紋 [2] に着目する。紙指紋とは、紙表面の微細構造のことである^(注1)。紙面を接写すれば、紙を形成する植物繊維の絡み具合がランダム状の模様を生成していることがわかる。この模様中の例えばコーナー点やエッジ点を検出できれば、それを特徴点として対応付けに利用できる。こうした紙面に自然に存在する特徴点は従来あまり利用されることはなかった。従って本研究の特徴の一つであると言える。

以下、本論文では関連研究を述べた後に、カメラ付きペンから手書きパターン全体を得るための原理について述べる。次に紙指紋からの特徴点検出について述べ、さらにその特徴点を手がかりにフレーム間の姿勢変化の推定を行う方法を述べ、最後にビデオモザイクングによりパターン全体を 1 枚の画像として得る方法を説明する。そして実験結果を示し、最後に今後の課題について述べる。

2. 関連研究

本研究に類似の手法として、現在開発されているアノトペン [3] が挙げられる。これはスウェーデンの Anoto 社が開発したデジタルペンで、ペン先に付いたカメラを用いて専用紙上のペン先の絶対位置を計測できる。この専用紙には、小さなドットが埋め込まれており、そのドットパターンが紙面上の位置情報を表現している。ペンが動けば観測されるドットパターンが変わるので、ペンの動き、すなわち手書きパターンを取得できる。本研究では特別な用紙を必要としないという点で、手書きパターン取得に限って言えばアノトペンよりも優位であると言える。

他のカメラ付きペンの例として、Arai らによる PaperLink [4] が挙げられる。これは、ペン型の小型カメラと蛍光ラインマーカーを組み合わせたもので、紙の文書をハイパーテキストのように扱うことができる。具体的には、ペンのボタンを押しながら蛍光ラインマーカーで紙の上をなぞり、その部分のパターンをカメラで切り出しておく。そして切り出された領域に対する

(注1)：この紙指紋という名称は、その模様により紙の同一性を検証する技術に由来する。例えば、富士ゼロックスが開発した紙指紋照合技術 XAYA [2] では、紙表面の模様すなわち紙指紋をスキャナで光学的に読み取ってデータベースなどに記録しておき、識別時には入力された紙画像の紙指紋と照合する。高精度で用紙を識別できるとしている。

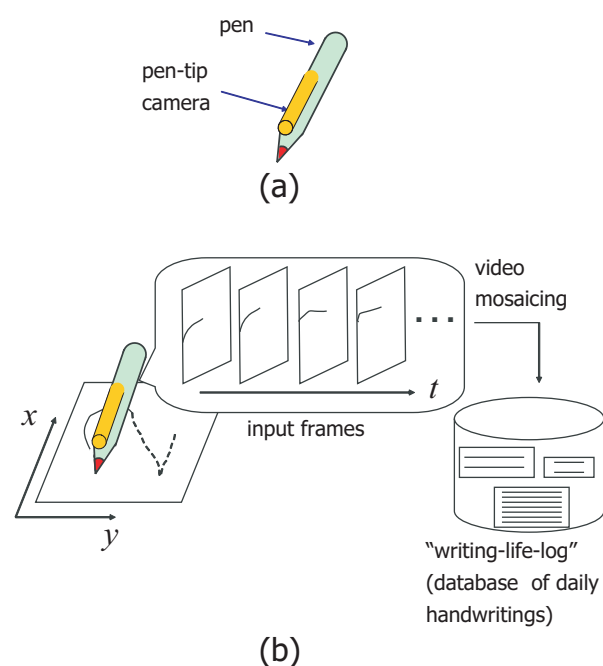


図 1 (a) カメラ付きペンおよび (b) ビデオモザイクングによる writing-life-log

アクションを定義しておけば、後でその領域を撮影することで定義しておいたアクションが起動される。

本研究の技術は文献 [5] で提案されている「情報埋め込みペン」と組み合わせる利用ができる。この情報埋め込みペンでは、紙への筆記と同時に微小インクドットの塗布により様々な情報 (例えば URL や筆者 ID など) を手書きコンテンツに埋め込むことができる。本研究で得られる手書きコンテンツの全体形状と埋め込んだ情報をペアにしてライティング・ライフ・ログに登録しておくことで、手書きコンテンツにサイバーメディア的機能を付加することが可能となる。

3. カメラ付きペンの概要

図 1(a) に示すように、本研究ではペン先に超小型 CCD カメラ (プラムネット ハンディミニ CCN3412Y) を搭載したカメラ付きペンを用いる。このペン先カメラから現在筆記途中の手書きパターンおよび紙面を撮影する。カメラはペンに固定されているために、視野内では常に同じ位置にある。図 2 に実際のペン先カメラからの画像例を示す。

実際に手書き内容全体を得るためには、図 1(b) に示すように、動画の各フレームに断片的に撮影された手書きパターンからその全体像を復元する必要がある。これはいわゆるビデオモザイクング処理と呼ばれる。このために、現在フレーム画像と直前フレーム画像の隣接する 2 フレーム間において、それぞれの画像上で対応する特徴点を SIFT (より詳細には SURF [7]) により検出する。そして両フレーム画像に共通する特徴点を抽出することができれば、それらの対応関係からフレーム間の位置関係、より正確には射影変換を推定できる。この対応関係 (射影変換行列) をすべての隣接フレーム間において求めることで、

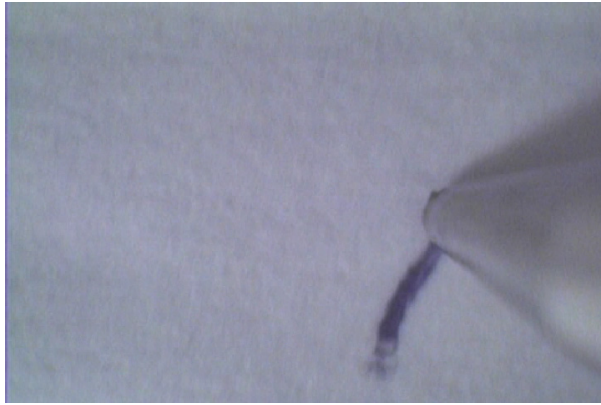


図 2 カメラ付きペンからの画像

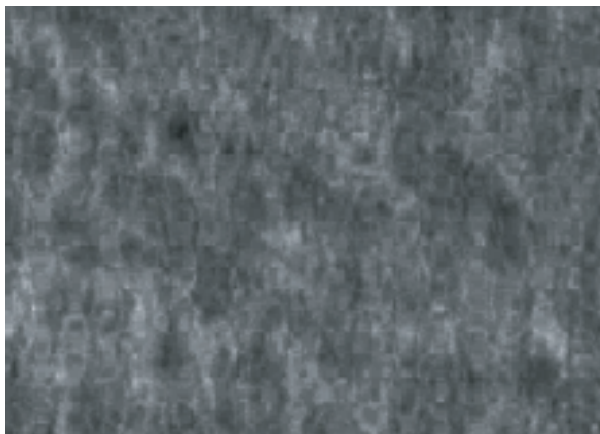


図 3 紙指紋の例．図 2 の一部を拡大したもの．ただし、見やすいように輝度を調整のうえ、コントラストを強調した．

全フレームに渡った位置関係の変化，すなわちペン先の移動を求めることができる．

ところで、カメラの位置については、様々な形態が考えられる．例えば、ペン尻付近にカメラをつければ、より広範囲の文字領域を撮影できると想定できる．従って、文字全体（さらには紙面全体）を一括して捉えられる可能性があり、その意味ではモザイクングが必要なペン先カメラより有利である．しかし、ペン尻カメラにはペンを持つ手によるオクルージョンが発生する．またペンの動きそのものが必要とされるようなアプリケーションの場合、振幅の大きなペン尻では動き推定が困難になる可能性がある．

ペン先カメラははまさに相補的な役割を為す．ペン先にカメラを搭載することで、より確実に手書きパターン付近を撮影することができる．さらに、後述するように紙指紋を有効に利用するため、詳細なペン先の動きを推定することが可能となる．このようにペン先とペン尻のカメラでは役割が異なるため、今後は平行して、もしくは組み合わせながら検討を進めていくべきと考える．

4. 紙指紋からの特徴点検出

本研究では、紙指紋に着目する．前述のように、紙には表面上に幾何学的模様がある．この性質を用いて、各フレーム画像

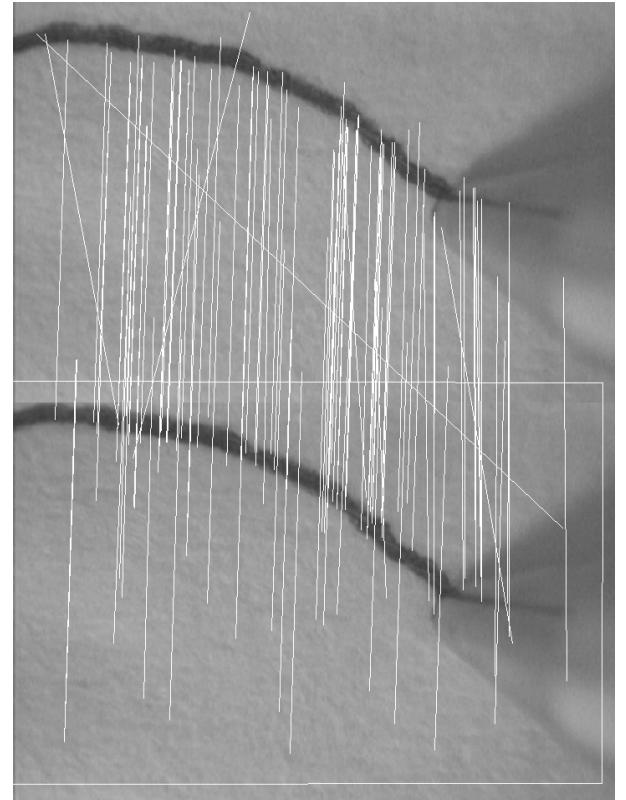


図 4 紙面上に検出された SURF 特徴点位置と隣接フレーム間でのマッチング結果

の位置関係（すなわちフレーム間移動量および方向）を把握することができれば、必ずしも手書きパターン不要である．この性質は次の二つの意味で極めて有効である．

- 第一に、ペン先による隠蔽問題の回避がある．は例えば図 2 のようにペン先が見えているとき、ペンが左から右に動いたとすると手書きパターンはペン先部分に隠蔽されて全く見えない．従って手書きパターンに注目して移動量推定をしようとしても不可能である．これに対し、紙指紋から移動量がわかれば手書きパターンが見えなかったとしても問題ない．

- 第二に、手書きパターンの開口問題回避がある．手書きパターンとして水平線を書き続けた状況を考える．この場合、画面内の手書きパターンは常に同じのものが見え、従ってペンが動いているのか静止しているのか判断できない．これは動き推定における開口問題である．水平線は極端な例であるが、文字を筆記する際にも局所的に変化のないパターンは頻繁に発生しているので、その箇所では不自然な移動量推定が発生し、結果的に手書きパターン形状は非線形に伸縮したものとなり得る．これに対し、紙指紋に着目すれば、ペンが動いている場合は紙指紋も動き、逆にペンが静止していれば紙指紋も静止しているため、この開口問題を回避できる．

特徴点としては、回転・スケール不変量かつ明るさ変化に頑強なものが望ましい．前者はペンが回転することによってフレーム画像が回転するためであり、さらにカメラ位置と紙面の距離関係も運筆によって変わるためである．こうした状況でも安定して特徴点を抽出するためには、回転・スケール不変量が

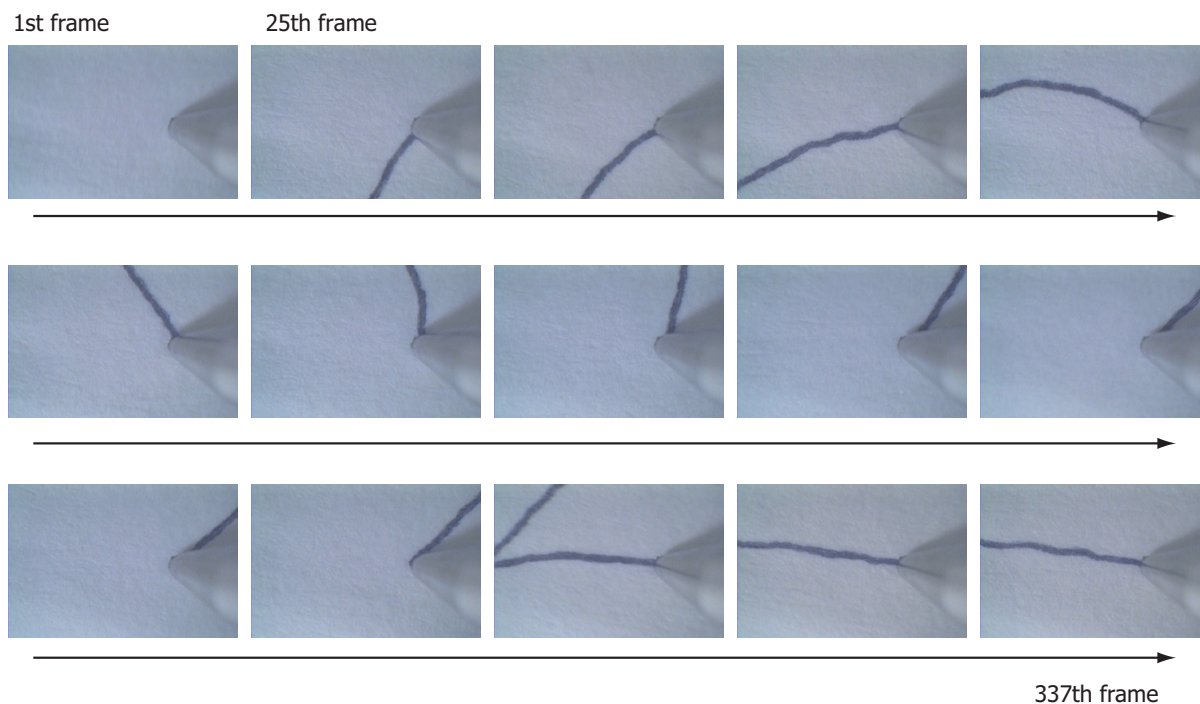


図 5 入力されたフレーム画像列。(24 フレーム毎に表示。)

望ましい。(厳密には射影変換不変量が望ましいが、隣接フレーム間での変位はそう大きくないので、回転・スケール不変量であれば、近似的に対応できるものと考え。)一方、後者は各フレーム画像の一部に現れる影の影響を排除したいためである。

本報告ではこれらの要求を満たすものとして SIFT(Scale-Invariant Feature Transform) の枠組みによる特徴点検出ならびに特徴記述を利用した。良く知られているように、SIFT で記述される特徴量は回転やスケール変化に対して不変であり、また照明変化にも頑健という性質を持っている。本報告ではその高速版である SURF(Speed-up Robust Features) [7] を用いることとした。

5. ビデオモザイクング

極めて類似した SURF 特徴を持つ点对を隣接フレーム間で複数求めることで、隣接フレーム間の姿勢変化を推定することができる。紙面が平面の場合、カメラ付きペンで撮影したフレーム画像は、互いに射影変換の関係にある。そこで、隣接フレーム間の点对関係から射影変換を推定すれば、それをを用いて隣接フレームを重ね合わせることができる。この処理をすべてのフレームにわたって行えば、結局全フレームを重ね合わせることができる。いわゆるビデオモザイクングである。

図 4 は、隣接するフレーム間に実際に求めた点对関係である。類似した SURF 特徴を持つ点を線で結んで表示している。SURF 特徴点は紙面に多数検出されるので図として見難くなっているが、良く見ると同じ紙面上の位置通しに対応しているものが多いことがわかる。

同図を見ると、全く誤った点对応を与えている場合があることもわかる。この大きく誤った点对応を含めて射影変換行列を求めた場合、手法によってはその悪影響が拡大し、誤った射影

変換行列が得られる可能性もある。紙指紋にも限界があると予想されるため、こうした誤った点对応はむしろ不可避と考えるのが妥当であろう。

このため、射影変換行列の推定には、いわゆるロバスト推定法が必要になる。そこで本報告では RANSAC [8] を利用する。RANSAC は少数の点对で射影変換行列を求め、その射影変換行列によりどの程度他の点对を説明できるかを評価する方法である。射影変換行列を求める点对の組をランダムに変えながらこの評価を行うことで、ロバストに射影変換行列を求めることが可能である。

フレーム間画像で特徴点の対応関係から射影変換を推定する際には、本手法特有の工夫が 3 つある。これらを以下に列挙する。

- 第一は、ペン先部分に現れる特徴点の除去である。カメラがペンに固定されているために、ペン先部分は画像内で常に一定の位置にある。従って、紙指紋部分の特徴点はフレーム間で動きを見せたとしても、ペン先部分の特徴点は静止しているように見える。このため、これらを総合して移動量を推定してしまうと、ペン先の特徴点が悪影響を及ぼし、誤った結果が得られる。このため、ペン先部分に現れる SIFT 特徴点は無視する必要がある。
- 第二は、ペン先付近の手書きパターン(黒インク)の除去である。このペン先付近の手書きパターンは、直前のフレームから現在のフレームまでの時間の間に新たに筆記された部分である。このため、直前のフレームには対応する点がなく、最悪の場合は誤対応を生じることになる。従って、この部分も、ペン先部分と同様に無視して考える必要がある。今回のペン先画像のビデオモザイクングと通常のビデオモザイクングとの違いは、このように本手法では動的に生成されつつあるパターン

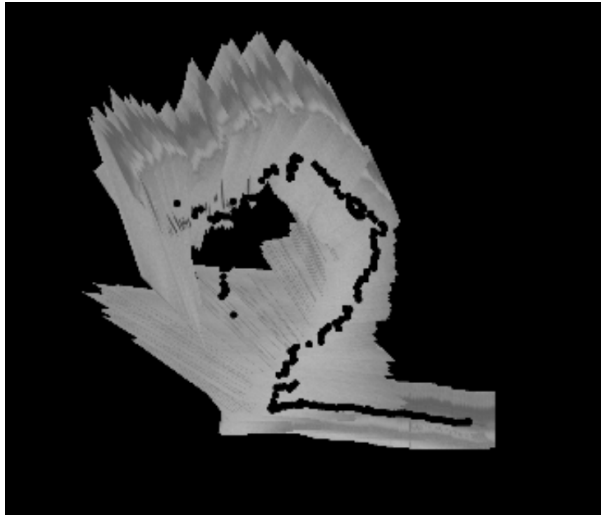


図 6 図 5 のフレーム列をモザイクした結果

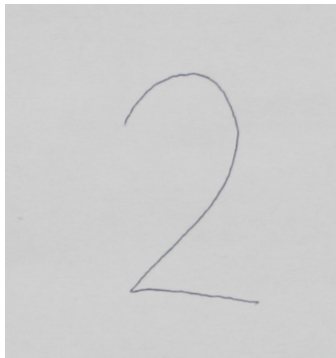


図 7 図 6 において実際に筆記された文字

を対象としている点である。

- 第三は、誤った射影変換行列の無視である。基本的に手書きは連続的に行われるので、射影変換自体も連続的なものが得られるはずである。しかし、SURF 特徴点对応の不安定性などの理由に、突発的に全く誤った射影変換行列が得られる場合がある。このような場合、現在のフレームをスキップして、一つ前のフレームと次のフレーム間で射影変換をすればよい。フレームの大部分はオーバーラップしているので、数フレームスキップしてもあまり影響はないと言える。ただし、連続スキップにより間が開きすぎると、2 フレーム間には大きな姿勢変化が発生し、それだけ SURF 特徴の対応が難しくなるので、注意が必要である。

6. 実験結果

図 5 に “2” を筆記した際のペン先画像のフレーム系列を示す。用いた紙はコピー用紙（非再生紙）であり、筆記した “2” のサイズはおよそ 3.5cm×2.5cm であった。フレーム数はおおよそ 340 フレームであった。カメラのフレームレートが 30fps であるため、これはおよそ 11 秒に相当する、これは動きボケを避けるべく、ゆっくり筆記したためである。

図 6 がビデオモザイクの結果である。モザイク画像の上に、各フレームでのペン先位置に小さな黒丸 () をプロットして

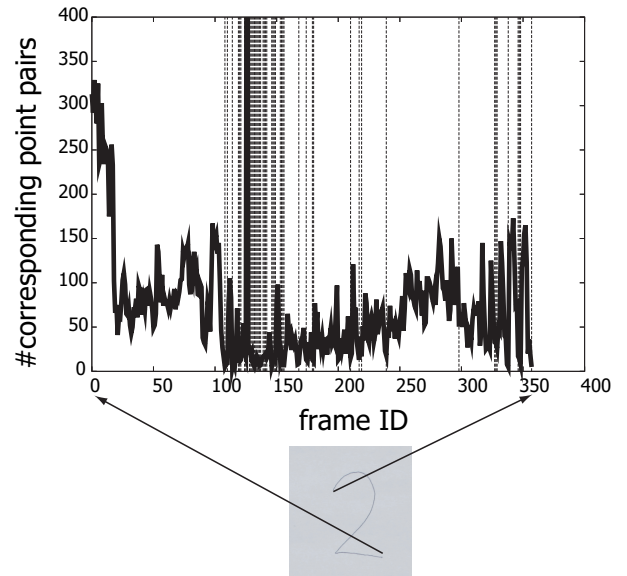


図 8 各フレームにおける特徴点对応数

いる。この黒丸の系列がすなわち復元された手書きパターンである。また図 7 がこのフレーム系列で実際に書かれた “2” をスキャンしたものである。

図 7 と比べると、図 6 の復元画像はかなりジャギーであるが、それでも “2” であるとは見て取れる。今回の場合、単に逐次的に貼り合せてモザイク画像を作ったため、射影変換の推定誤差が蓄積していく。また初期フレームが紙面に正対していなければ、それが全体に影響する。このため、原理的に形状が不安定になりやすい。それでもこの程度の復元ができているということは、紙指紋の特徴点を用いたモザイクングに見込みがあることを示している。

図 8 は、各フレームにおける SURF 特徴点对の数である。また同図の縦線は、そのフレームで不自然な射影変換行列が求めたためにスキップ (5. 参照) したことを表している。詳細な吟味は今後の課題であるが、図 6 と併せて考えると、特徴点对が少なくなり、スキップが起こる付近では、精度が落ちやすいという傾向があるように見える。具体的には “2” の屈曲点付近および上部付近においてスキップが多く見られ、特に後者付近の復元パターンはやはりジャギーになっている。

図 9 は図 10 を筆記した場合のビデオモザイクの結果である。この “2” はおよそ 1.6cm×1.1cm のサイズであり、図 6 より小さい。全体で 118 フレーム、すなわち 4 秒程度で書かれたものである。若干の非線形伸縮が見られるが、“2” であることは明瞭にわかる程度の精度は保っている。図 11 は、各フレームにおける特徴点对応数である。

今回は黒インク部分に重みを置くといったような処理は一切しておらず、紙面・インクの区別無く求めた SURF 特徴で射影変換を推定している。文献 [9] で示したように、実際には黒インク部分の重ね合わせ評価だけでもかなりの精度でモザイクングは可能である。従って、今後は黒インクがペン先に隠蔽されている場合にだけ紙指紋を使うといった工夫も可能と思われる。

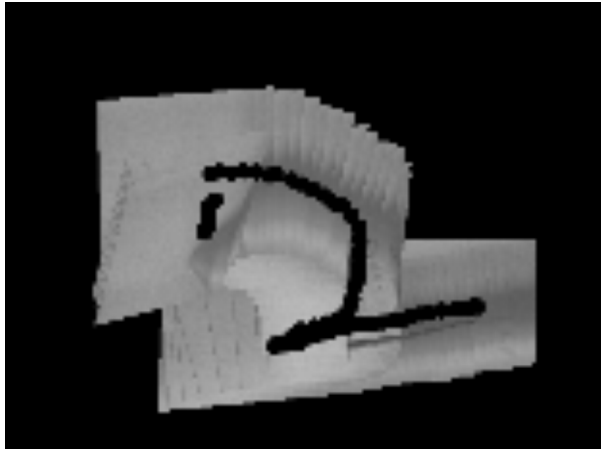


図 9 他のモザイクング結果

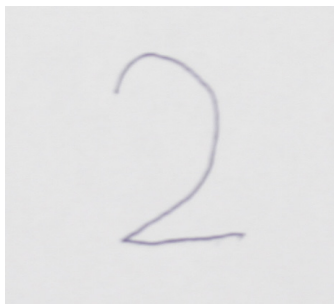


図 10 図 9 にて実際に筆記された文字

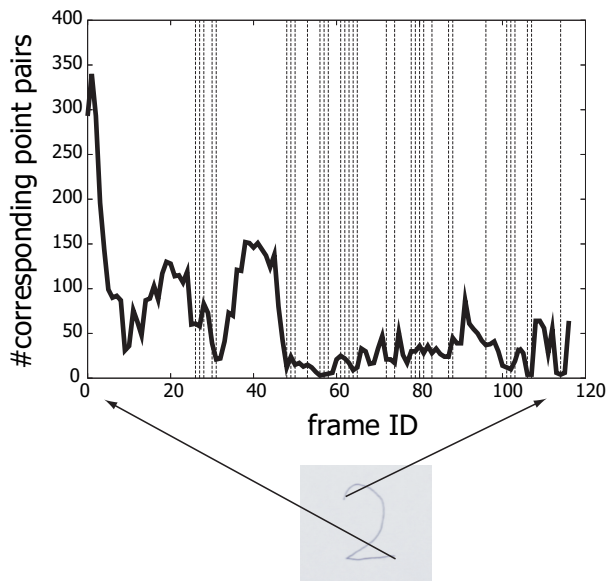


図 11 各フレームにおける特徴点对応数

7. ま と め

本報告では、ペン先に取り付けたカメラからの映像から手書きパターンを復元することを目的としたビデオモザイクング法について論じた。本手法の最大の特徴は紙面の模様 — 紙指紋 — を利用することにある。紙指紋から SURF 特徴点を抽出し、それをビデオモザイクングに利用することで、例えば手書きパターンそのものがペン先に隠蔽されているような状況であっても、ペンの動きすなわち手書きパターンを復元できる。

極めて初期的な検討段階ではあるが、このような単純な方式でも手書きパターンの概略が復元できることがわかった。

今後の課題としては以下が挙げられる。まずは、SURF 特徴点を求める前に紙指紋を強調するための画質変換を行ったり、黒インク部分の対応関係を重要視しながら紙指紋特徴点による対応と組み合わせるなど、SURF による特徴点对応の安定化を図る必要があると考える。さらに、ビデオモザイクングの方法の改良が重要と考える。本報告では隣接フレーム間の位置合わせを繰り返す手法を採ったので、後のフレームになるほど誤差が蓄積してしまい、モザイクング結果が崩れてしまう場合があった。再出現点の利用 [6] などによる安定化が必須と思われる。動きボケの除去も重要であろう。ペンの移動速度によっては、動きボケが顕著になり、SURF 特徴点を検出できなくなり、結果としてマッチングもできなくなってしまう。従って動きボケの除去が重要になって来るが、手書きが写っているフレームでは手書きが線状パターンであることを活かしたボケ除去 [10] も考えられる。

評価についても、復元精度の定量的な評価や、筆記速度に対する耐性測定が必要と考えている。また本手法では紙指紋を手がかりにモザイク画像を求めているので、紙の質の影響についても調査したいと考えている。

謝辞: 過去のカメラ付きペンおよびアノトペンについて、日立中央研究所の池田尚司氏にご教授頂いた。ここに深謝する次第である。本研究の一部は科学技術振興機構平成 18 年度可能性試験、科学研究費補助金 (基盤 (B), No.20300049)、および Microsoft Research Asia Mobile Computing in Education Theme Program によった。

文 献

- [1] M. Irani and P. Anandan, "Video indexing based on mosaic representations," Proc. IEEE, vol. 86, no. 5, pp. 905-921, 1998.
- [2] <http://www.fujixerox.co.jp/company/technical/xaya/>
- [3] <http://www.anoto.com/>
- [4] T. Arai, D. Aust, S. E. Hudson, "PaperLink: a technique for hyperlinking from real paper to electronic content," Proc. ACM Conf. Human Factors in Computing Systems (CHI'97), pp. 327-334, 1997.
- [5] 田中一弘, 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, "データ埋め込みペンに関する基礎的検討," ヒューマンインタフェース学会論文誌, vol. 10, no. 4, pp. 559-567, 2008.
- [6] 池谷彰彦, 佐藤智和, 池田 聖, 神原誠之, 中島 昇, 横矢直和, "カメラパラメータ推定による紙面を対象とした超解像ビデオモザイクング," 信学論, vol. J88-D-II, no. 8, pp. 1490-1498, 2005.
- [7] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," Proc. ECCV2006 (LNCS volume 3951), part 1, pp. 404-417, 2006.
- [8] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Comm. of the ACM, vol. 24, no. 6, pp. 381-395, 1981.
- [9] 伊東克啓, 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, "ペン先カメラ画像からの手書きパターンの復元," 電子情報通信学会 2008 年総合大会 ISS 特別企画学生ポスターセッション, ISS-P-323, 2008.
- [10] X. Y. Qi, L. Zhang, C. L. Tan, "Motion deblurring for optical character recognition," Proc. ICDAR2005, pp. 389-393, 2005.