

動画像からの三次元復元による湾曲した紙面のビデオモザイクング

佐藤 智和^{†,††} 池谷 彰彦^{††} 池田 聖[†] 神原 誠之^{†,††}
中島 昇^{††} 横矢 直和^{†,††}

Video Mosaicing for Curved Documents Based on Structure from Motion

Tomokazu SATO^{†,††}, Akihiko IKETANI^{††}, Sei IKEDA[†], Masayuki KANBARA^{†,††},
Noboru NAKAJIMA^{††}, and Naokazu YOKOYA^{†,††}

あらまし 本論文では、見開きの冊子のような湾曲したドキュメントを対象とする、ビデオ映像のみを用いた画像モザイクング (ビデオモザイクング) 手法を提案する。従来提案されているモザイクング手法の大半は、平面または平面に近似できるもののみを対象としており、湾曲した紙面のような非平面に対して手法を適用した場合には、生成画像上に幾何学的な歪みや複数画像の合成位置のずれに起因するブレが発生するという問題があった。提案手法では、画像上の見かけの動きを用いた三次元復元 (structure from motion) によって撮影対象の形状と画像撮影時のカメラ位置・姿勢パラメータを決定し、紙面形状を推定することで紙面上の文書や写真を平面状に展開した高解像モザイク画像を生成する。本研究では、実際に手法を試作システム上に実装し、湾曲した紙面上の文章・写真から展開モザイク画像を作成する。また、生成された画像上の歪みを評価することで手法の有効性を示す。

キーワード ビデオモザイクング, structure from motion, 形状推定, 特徴点追跡, カメラ外部パラメータ推定

1. はじめに

携帯電話に代表される持ち運び可能な小型電子機器の普及やユビキタスネットワークインフラの充実に伴い、現実中存在する書籍・紙面上に記載された文章・写真などの情報をこれらの機器を用いて手軽に電子化し、保存または遠隔地に伝送したいという要求が高まっている。現在、一般的な紙面情報の電子化にはフラットベッドスキャナ (以下、スキャナ) と呼ばれる画像センサを紙面の上で機械的に等速に動かす装置が用いられるが、機器が大きいために、手軽に用いるという用途で利用することは難しい。また、書籍などのような厚い冊子をスキャナ上に置いて電子化すると図 1(a) に示すような画像歪みや影が発生するため、一般的には書籍をスキャナ本体に押し付ける必要があるが、これ

により書籍が損傷してしまうという問題がある。このようなスキャナにおける紙面の平面化の問題を、物理的ではなく画像の幾何的変換によって解決しようという試み [1], [2] も行われているが、本研究では手軽に持ち運び可能な携帯機器での利用を想定し、一般的なカメラ画像を入力として扱う手法について検討する。

一般的なカメラを用いた書籍の電子化において最も単純な方法は、一枚の静止画像として紙面を撮影することであるが、一般的には図 1(b) に示すような透視投影歪みが発生するため、スキャナの場合と同じように紙面の平面化を行う必要がある。このようなカメラ画像を対象とした紙面平面化の手法は、撮影対象の三次元形状を計測し平面化に用いる手法と、紙面に対する事前知識を用いる手法に分類できる。Pilu ら [3] や Brown ら [4] は、スリット光の投影による三次元復元手法によって対象を三次元計測し、伸縮可能なメッシュモデルによって表現される紙面モデルを計測データに当てはめることで、任意に湾曲した紙面を平面状に展開する手法を提案している。また、Yamashita ら [5] は、固定されたステレオカメラによって測定される奥行きデータに対して NURBS 曲面を当てはめることで紙面形状を推定し、平面状に展開する手法を提案し

[†] 奈良先端科学技術大学院大学 情報科学研究科 〒 630-0192 奈良県生駒市高山町 8916-5

Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0192 Japan

^{††} NEC インターネットシステム研究所 〒 630-0101 奈良県生駒市高山町 8916-47

Internet System Lab., NEC, 8916-47, Takayama-cho, Ikoma-shi, Nara, 630-0101 Japan

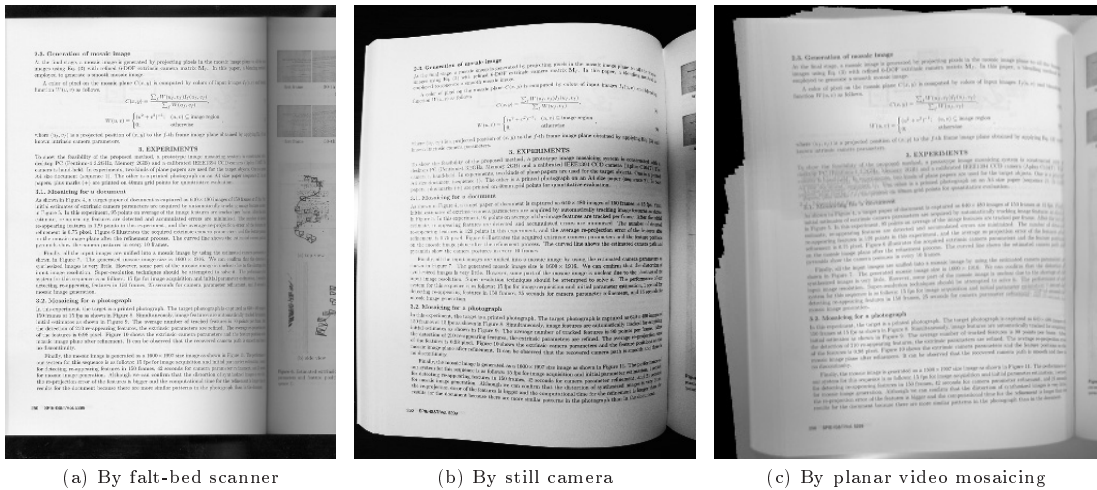


図 1 湾曲した紙面の電子化における歪みの発生例
Fig. 1 Examples of image distortion for curved document.

ている。これらの手法は任意に湾曲した紙面を扱えるが、特殊な三次元計測装置が必要であり一般のユーザが利用することは難しい。また、メッシュの局所的な当てはめ誤差が蓄積するため、結果として生成画像に大域的な歪みが発生しやすいという問題がある。

一方、撮影対象である紙面に関する事前知識を用いることで、一枚の画像のみから平面化した画像を生成する手法も提案されている。Cao ら [6] や Liang ら [7] は、事前知識として文字のベースライン情報を用い、それらを平行化することで紙面を平面化している。また、Brown ら [8] は紙面の輪郭情報を用い、それらを直線化・直交化する手法を提案している。これらの手法は、一枚の画像のみを入力とするため特殊な機材を必要としないが、写真のように文字のベースラインが存在しない場合や、高精度な輪郭抽出が難しい場合には手法が適用できない。また、画像上での 2 次元な補正を基礎としているため、透視投影に起因する幾何学的な歪みや紙面上の陰影を正しく取り除くことが原理的に難しいという問題がある。

他方、カメラを用いた紙面の高解像電子化手法として、複数枚の画像を用い、それらを高精度に位置合わせすることで合成された一枚の高解像画像を生成する画像モザイクング (入力が動画の場合はビデオモザイクング) と呼ばれる手法が研究されている [9] ~ [13]。これらの手法では、対象を平面または平面に近似できるものと仮定し、Homography と呼ばれる平面射影変換のパラメータを推定することで画像の位置合わせを

実現しているものが多い。しかし、Homography を用いる手法では原理的に紙面の傾きを検出できず、基準フレーム撮影時の画像平面が紙面に対して正対していない場合には、生成されるモザイク画像に歪みが残されるという問題があった。我々はこのような問題に対して、Homography の代わりにカメラ外部パラメータを利用することで、傾きの影響を自動的に除去する手法を提案してきた [13]。しかし我々が提案した手法を含め、平面を撮影することを前提としている従来のモザイクング手法では、湾曲した紙面を対象とした場合には、図 1(c) に示すように、生成されるモザイク画像上に大域的な幾何学的歪みが発生し、また画像を正しく位置合わせできないために局所的な画像のブレが発生する。非平面を対象としたモザイクング手法についても既に研究例はあるが、対象の形状を計測するためにアクティブカメラやスリット光投影装置などの特殊な装置が必要であったり [14]、手動による形状測定が必要である [15] という問題が残されている。

そこで本論文では、平面を対象とする従来のビデオモザイクング手法 [13] を拡張し、新たに撮影対象の三次元形状を推定することで、湾曲した紙面を平面状に展開するビデオモザイクング手法 [16], [17] を提案する。本研究では、まず動画を撮影可能なカメラを用いて紙面を撮影し、動画画像からの三次元復元 (structure from motion) によって紙面上に存在する特徴点の三次元位置と、画像撮影時のカメラの位置・姿勢を推定する。次に、得られた三次元点群を用いて紙面形状の

推定を行い、この形状情報とカメラパラメータを用いて紙面を平面状に展開したモザイク画像を生成する。また、光源と紙面に対する簡単な仮定をおくことにより、紙面上の陰影を取り除き、光学的にも正しく展開されたモザイク画像を生成する。本手法は、カメラ以外の特殊な機器や紙面上のドキュメントの内容に関する事前知識を必要とせず、動画のみから高解像なモザイク画像を獲得することができるという特長があり、原理的には携帯電話や PDA 等に内蔵されたカメラのみを用いた紙面電子化への応用が可能である。ただし、本論文では紙面の湾曲は一軸方向のみとし、紙面の撮影中において照明条件は変化しないものとする。また、カメラの内部パラメータを既知かつ撮影中固定とし、カメラのシャッタースピードは撮影画像上で動きブレが発生しない程度に高速に設定されているものとする。

以下、本論文では 2. で、湾曲した紙面に対するビデオモザイク手法の手順について詳述する。3. では、試作システムを用いて提案手法を評価し、定量的な実験結果を示す。最後に 4. で、まとめと今後の課題について述べる。

2. 湾曲した紙面に対するビデオモザイク

本章では、特徴点追跡による三次元復元に基づいて紙面の形状を推定し、平面状に展開したビデオモザイク画像を生成する手法について詳述する。図 2 に提案手法の処理の流れを示す。本手法では、まず、動画に対して特徴点追跡を行い、structure from motion の原理により、特徴点の三次元位置とカメラの位置・姿勢パラメータを逐次推定する (A)。次に、推定されたパラメータを最適化し、三次元点群に対して多項式近似による曲面・曲線当てはめを行うことで紙面形状を推定する (B)。最後に、紙面形状とカメラパラメータから、平面状に展開したモザイク画像を生成する (C)。以下では、まずカメラ外部パラメータと三次元復元に用いる誤差関数の定義を行い、続いてステージ (A) ~ (C) について詳述する。

2.1 カメラ外部パラメータと誤差関数の定義

ここではまず、三次元復元に用いる座標系と誤差関数について定義する。本研究では、紙面が存在する実環境で張られる世界座標系における任意の点 $S_p = (x_p, y_p, z_p)$ が、第 f フレームの画像座標 $x_{fp} = (u_{fp}, v_{fp})$ に投影されるものとする。このとき、第 f フレームのカメラ外部パラメータを表す 6 自由度

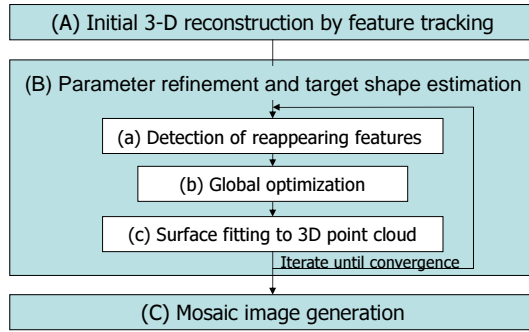


図 2 提案手法の処理の流れ
Fig. 2 Flow diagram of the proposed method.

の 3×4 行列 M_f を用いれば、三次元座標 S_p と画像上の二次元座標 x_{fp} の間には以下の関係が成り立つ。

$$(au_{fp}, av_{fp}, a)^T = M_f(x_p, y_p, z_p, 1)^T \quad (1)$$

ただし、 a は媒介変数である。本論文では記述の簡単化のために、 x_{fp} を焦点距離 1 の投影面上に張られるレンズ歪みを考慮しない理想カメラ座標とするが、実際には、実画像上の座標 $\hat{x}_{fp} = (\hat{u}_{fp}, \hat{v}_{fp})$ を、既知のカメラ内部パラメータ (レンズ歪み, 焦点距離, アスペクト, 投影中心) を用いて理想カメラ座標 x_{fp} に変換する必要がある。

次に、三次元復元に用いる誤差関数について定義する。一般に、撮影時の標本化や、特徴点抽出の誤差、およびカメラ外部パラメータ・特徴点の三次元位置の推定誤差によって、第 f フレームにおける特徴点 p の検出座標 $x'_{fp} = (u'_{fp}, v'_{fp})$ と、式 (1) によって計算される S_p の画像上での座標 x_{fp} は一致しない。そこで、本論文では、第 f フレームにおける特徴点 p に関する誤差 E_{fp} を以下のように定義する。

$$E_{fp} = |x_{fp} - x'_{fp}|^2 \quad (2)$$

2.2 特徴点追跡による三次元復元

本節では、動画の中の特徴点を追跡することでカメラ外部パラメータ M_f と特徴点の三次元位置 S_p を逐次推定する手法について述べる (ステージ (A))。なお、ここで用いる三次元復元手法は、文献 [18] の手法を拡張したものである。文献 [18] の手法では、初期フレームを含むキーフレームにおいて三次元位置が既知の基準点を与えることで、三次元復元処理における蓄積誤差を最小化し、基準点によって張られる座標系におけるカメラの位置・姿勢と特徴点の三次元位置を推定す

る。しかし、本研究では、対象となる湾曲した紙面に対して三次元位置が既知の基準点をあらかじめ設定することができないため、初期フレームにおいて基準点を与える代わりに一定の仮定の下でカメラの位置・姿勢および特徴点の三次元位置を初期化し、新たに部分区間に対する最適化処理を加えることで初期値と真値のずれを自動で補正する。また、再出現特徴点を検出し、全体最適化処理を行うことで、累積的な推定誤差の発生を抑える。

はじめに、初期フレームにおける処理について述べる。本研究では、初期フレームにおいて、ユーザが持つカメラが撮影対象である紙面に対しておおむね正対しているものと仮定し、 M_f の回転成分を単位行列、平行移動成分を 0 で初期化する。また、初期フレームの画像内で検出された各特徴点 p についても、この仮定に基づき三次元座標 S_p を $(u_{1p}, v_{1p}, 1)$ で初期化する。なお、初期値として与えられる特徴点の三次元位置およびカメラ位置・姿勢は、逐次処理における部分区間での最適化処理によって自動で補正されるため、ユーザは上記の仮定を厳密に満たす必要はないが、初期値が真値から大きく離れる場合には、局所解に陥ることで正しい解が得られないことがある。

次に、後続のフレーム ($f > 1$) での処理について述べる。各フレームのカメラ外部パラメータ M_f は、以下のステップを最終フレームまで繰り返すことによって逐次推定される。

(特徴点の追跡) Harris のインタレストオペレータ [19] とテンプレートマッチングにより、前フレームの各特徴点について、現フレームにおける対応点を探索する。更に、RANSAC アルゴリズム [20] を用いて、特徴点の誤対応を排除する。

(カメラ外部パラメータの推定) 前ステップで決定された特徴点の画像上での座標 (u'_{fp}, v'_{fp}) と、対応する三次元座標 $S_p = (x_p, y_p, z_p)$ を用い、カメラ外部パラメータ M_f を算出する。ここでは、式 (2) で定義した誤差関数の和 $\sum_p E_{fp}$ を Levenberg-Marquardt 法によって最小化することで、カメラ外部パラメータを得る。なお、特徴点 p の三次元座標 S_p は、前のフレームまでに推定されている値を用いる。

(特徴点の三次元位置推定) 現フレームに存在する全ての特徴点 p について、特徴点ごとに $\sum_{i=1}^f E_{ip}$ を最小化することで三次元座標 $S_p = (x_p, y_p, z_p)$ を算出し、更新する。

(特徴点の追加と削除) 複数の評価尺度を用いること

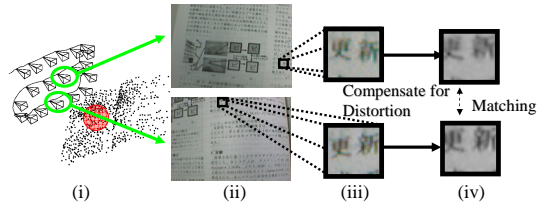


図 3 再出現特徴点の検出; (i) カメラの移動経路と特徴点の三次元位置, (ii) 入力画像中の時間的に離れたフレーム, (iii) 各画像上における同一特徴点のパターン, (iv) 投影歪みの影響を取り除いたパターン

Fig. 3 Detection of re-appearing features; (i) camera path, posture and feature position on mosaic image plane, (ii) sampled frames of an input image sequence, (iii) templates of a feature in different images, (iv) templates after removing distortion.

で、カメラパラメータ推定に用いる信頼度の高い特徴点の組をフレーム毎に自動更新する [18]。

(部分区間での最適化) バンドル調整の枠組みを用い、現フレームから一定フレーム前までの部分区間に対して、再投影誤差の総和を最小化することで特徴点の三次元位置とカメラパラメータを最適化する。

以上のステップを繰り返すことにより、各フレームにおけるカメラ外部パラメータ M_f と、特徴点の三次元位置 S_p を逐次的に推定する。

2.3 パラメータの最適化と紙面形状の推定

本節では、特徴点追跡による三次元復元によって得られたカメラパラメータと特徴点の三次元位置を最適化し、曲面当てはめによって、紙面形状を推定する手法について述べる (ステージ (B)). 本手順では、図 2(B) に示すように、まず、時間的に不連続な画像上に存在する再出現特徴点に対応付け (a)、これを用いたカメラパラメータおよび特徴点三次元位置の最適化を行う (b)。次に、特徴点の三次元点群に対する曲面当てはめによって紙面形状を推定する (c)。これらの処理 (a)~(c) を、式 (2) に示した再投影誤差の総和が変化しなくなるまで数回繰り返すことで、最終的に高精度なカメラ外部パラメータと紙面形状を得る。以下では、処理 (a)~(c) について述べる。

2.3.1 再出現特徴点の検出

カメラを動かしながら紙面全体を撮影しようとするとき、カメラは必然的に、図 3(i) に示すように、複数回の折り返しを伴って移動することになる。その結果、画像上の特徴点の中には、数フレームにわたって追跡されてから、一度フレームアウトした後に、再度フ

フレームインしてくるものが存在する．本研究ではこのような再出現特徴点を検出し，同一の特徴点として対応付ける．これにより，時間的に不連続なフレームのカメラパラメータ間に制約が付加され，後段の全体最適化処理によって累積的な誤差の影響を効果的に排除することが可能となる．ただし，前節で述べた逐次処理では，時間的に離れたフレームの画像上において必ず同一の自然特徴点が検出されているとは限らず，また本来同一の特徴点であっても推定された三次元位置に累積的な推定誤差が含まれているため，ICP アルゴリズム [21] に代表される特徴点の三次元位置のみを利用した位置合わせ手法を用いて再出現特徴点を検出・対応付けることは難しい．

そこで本研究では，時間的に離れたフレーム間で類似のパターン画像を持つ特徴点を検出することによって再出現特徴点の検出を行う．その際，図 3(iii) に示すように，紙面上の同一の特徴点であっても，カメラ運動の影響による投影歪みによってパターンが変化してしまうことが問題となるため，ここでは繰り返し処理の後段で推定される紙面形状を用い，画面上の特徴点周辺のパターンを紙面形状の対応する箇所に投影することで，投影による歪みの影響を排除する．次に，三次元空間中での距離が一定の閾値以下の特徴点の組を抽出し，両者のパターン類似度を正規化相互相関の尺度で評価する．ここでは，スケールの異なる複数テンプレートを用いて相関計算を行い，全てのスケールにおいて相関の高い特徴点の組を再出現特徴点とする．これにより，紙面上に同じ文字が多数存在する場合においても正しく再出現特徴点を対応付けることができる．ただし，本処理では，投影歪みの排除に紙面形状が必要となるため，処理 (a) ~ (c) の繰り返しの初回には，再出現特徴点の検出を行わない．

2.3.2 三次元復元結果の最適化

2.2 で述べた三次元復元処理は逐次的に行えるため，各フレームに対する処理は短時間でできる反面，推定誤差が蓄積するという問題がある．ここでは，バンドル調整の枠組み [22] を用い，以下に示す動画画像全体に対する再投影誤差の総和を最小化することで，カメラ外部パラメータと特徴点の三次元位置を最適化する．

$$E_{all} = \sum_f \sum_p E_{fp} \quad (3)$$

なお，再出現特徴点については，処理 (a) によって対応付けられた特徴点の組を同一の特徴点系列とみな

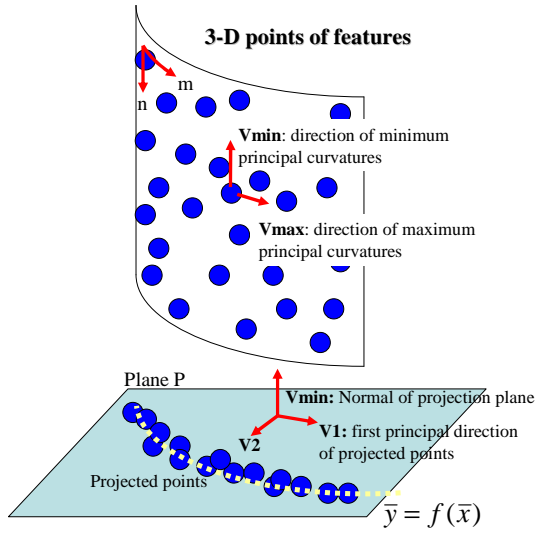


図 4 多項式当てはめによる紙面形状の推定
Fig.4 Target shape estimation by polynomial surface fitting.

し，誤差 E_{fp} を計算する．これにより，図 3(ii) のように時間的に離れたフレーム間でも整合性を保ちながら，カメラ外部パラメータと特徴点の三次元位置を最適化する．

2.3.3 紙面形状の推定

処理 (b) で最適化された特徴点の三次元点群を用い，見開きの冊子のように紙面が 1 軸方向にのみ湾曲しているという仮定の下で紙面形状の推定を行う．ここでは，まず，図 4 に示すように，三次元点群に局所的な曲面当てはめを行うことで曲率の主方向を算出し，最小主曲率方向に対して各特徴点の三次元座標を二次元に縮退させる．次に，得られた二次元座標群に対して多項式当てはめを行うことで，紙面形状の推定を行う．

最小主曲率方向は，処理 (b) によって得られた各特徴点 p の三次元位置 S_p を中心とする一定半径内の特徴点群 R_p に対して，局所的な二次曲面当てはめを行うことで算出する．ここでは，まず，特徴点群 R_p の三次元座標に対して主成分分析を行い，第一・第二・第三主成分の方向の単位ベクトルを基底とする座標系を構成する．これにより，三次元位置の分散が最小となる軸を高さ方向に取る二次曲面関数

$$z = ax^2 + bxy + cy^2 + dx + ey + f \quad (4)$$

を点群 R_p に当てはめる．次に，得られた二次曲面から，曲率が最大・最小となる方向を表す主方向および法線方向を算出する．この局所的な二次曲面当てはめ

を、全ての特徴点に対して繰り返すことで、全ての特徴点の三次元位置に対する主方向が求まる。本研究で想定する 1 軸方向にのみ湾曲する紙面では、全ての特徴点位置において理想的には最小曲率が 0 となり、その方向は紙面全体で一致する。しかし、実際には推定誤差の影響で最小主曲率の方向が一つに定まらないため、ここでは、投票によってアウトライヤを排除した後に、主方向を平均化し、曲面全体の最小主曲率方向 $\mathbf{V}_{min} = (vm_x, vm_y, vm_z)$ を決定する。

次に、得られた最小主曲率方向 \mathbf{V}_{min} を法線とする平面 $P(x, y, z) = vm_x x + vm_y y + vm_z z = 0$ 上に、各特徴点の三次元位置 S_p を投影する。ここでは、平面 P 上に投影される二次元点群の分散が最大となる軸方向の単位ベクトルを $\mathbf{V1}$ 、また、 $\mathbf{V1}$ と \mathbf{V}_{min} に直交するベクトルを $\mathbf{V2} = \mathbf{V1} \times \mathbf{V}_{min}$ とおき、以下の式を用いて S_p を $\mathbf{V1}$, $\mathbf{V2}$ を基底とする二次元空間 (\bar{x}, \bar{y}) に射影する。

$$\begin{pmatrix} \bar{x}_p \\ \bar{y}_p \end{pmatrix} = \begin{pmatrix} \mathbf{V1} \\ \mathbf{V2} \end{pmatrix} S_p \quad (5)$$

最後に、得られた二次元座標群 (\bar{x}, \bar{y}) に対して最小二乗法を用いて以下の多項式による大域的曲線当てはめを行い、紙面の形状パラメータ (a_0, \dots, a_q) を算出する。

$$\bar{y} = f(\bar{x}) = \sum_{i=0}^q a_i \bar{x}^i \quad (6)$$

ここで、幾何学的 AIC [23] を用いれば、以下の評価基準 G-AIC を最小とする次数 q を算出することで最適次数を自動で決定できる。

$$G\text{-AIC} = J + 2(N(m-r) + q + 1)\epsilon^2 \quad (7)$$

ただし、 J は式 (6) による当てはめの残差、 N は当てはめに用いる点の数、 m は観測データの次元、 r は当てはめに用いる拘束条件の数を表す。また、 ϵ は \bar{y} 軸方向に対する特徴点位置の推定誤差の平均値 (ノイズレベル) を表す。式 (6) による多項式当てはめでは、 m, r, N は当てはめに用いる次数 q に依存しないため、定数項を取り除けば、実際には以下の評価基準 G を最小とする q を選択することで最適次数を決定できる。

$$G = J + 2q\epsilon^2 \quad (8)$$

本研究ではノイズレベル ϵ として、カメラ座標系にお

ける各特徴点の奥行き距離を全てのフレームで平均した値 l の定数倍を近似的に与える ($\epsilon = Cl$)。

なお、見開きの冊子を対象とした場合のように、対象が複数の紙面を接続することで構成されている場合には、局所的二次曲面当てはめによって得られた法線方向の不連続性に基づいて複数紙面の接続線を検出し、それぞれの紙面に対して独立した形状パラメータを算出する。また、ここで得られた紙面形状は、後段の展開モザイク画像生成処理の他に、繰り返し処理 (a) における再出現特徴点の検出で、投影歪みを排除するために用いられる。

2.4 展開モザイク画像の生成

前節で述べた繰り返し処理によって推定される紙面形状とカメラパラメータを用いることで幾何学的な補正を行い、平面状に展開したモザイク画像を生成する。また、光学的な補正を行い、紙面上の陰影を排除する (ステージ (C))。

2.4.1 展開モザイク画像の生成

ここでは、図 4 に示すような、平面状に展開された紙面上の座標を表す (m, n) と、多項式近似曲面上の三次元座標 $(\bar{x}, f(\bar{x}), \bar{z})$ に関する以下の関係式を用い、曲面を平面に展開する。

$$(m, n) = \left(\int_0^{\bar{x}} \sqrt{1 + \left\{ \frac{d}{dx} f(x) \right\}^2} dx, \bar{z} \right) \quad (9)$$

具体的には、まず、式 (9) によって得られる展開モザイク画像上の座標 (m, n) に対応する三次元座標 $(\bar{x}, f(\bar{x}), \bar{z})$ を、以下の式によって各フレームのカメラ画像上に投影する。

$$\begin{pmatrix} au_f \\ av_f \\ a \end{pmatrix} = \mathbf{M}_f \begin{pmatrix} \mathbf{V1} \\ \mathbf{V2} \\ \mathbf{V}_{min} \end{pmatrix}^T \begin{pmatrix} \bar{x} \\ f(\bar{x}) \\ \bar{z} \end{pmatrix} \quad (10)$$

次に、投影座標の画素値を用いて展開モザイク画像を生成する。ただし、本研究では推定された紙面形状を用いて展開モザイク画像の生成を行うため、紙面形状の推定誤差は、モザイク画像上において直接画像合成位置のずれ (画像ブレ) として表れる。このような紙面形状の推定誤差に起因する画像合成位置のずれは、紙面の法線とカメラの光軸の成す角が 90 度に近づくにつれて顕著に表れるため、本研究では以下に示す、紙面の法線方向に応じた重み付き平均を展開モザイク画像 (m, n) の画素値 $I(m, n)$ とすることで、画像ブレを抑えながら展開モザイク画像を生成する。

$$I(m, n) = \frac{1}{\sum_f W_f} \sum_f W_f I_f(u_f, v_f) \quad (11)$$

$$W_f = K^{-\theta_f} \quad (12)$$

ただし, $I_f(u_f, v_f)$ は第 f フレームの画素 (u_f, v_f) の輝度値を表す. また, K は定数であり, θ_f は紙面上の三次元位置 $(\bar{x}, f(\bar{x}), \bar{z})$ における法線方向と, 第 f フレームのカメラ中心と $(\bar{x}, f(\bar{x}), \bar{z})$ を結ぶ直線の成す角 (単位: ラジアン) を表す.

2.4.2 陰影の排除

幾何学的な補正だけでは照明条件によっては陰影が不均一となるため, 人間の視覚には湾曲が正しく補正されたようには知覚されないことが多い. 本研究では, 紙面の背景が紙面上で最も明るい色 (反射率が紙面上で最大) のランバート面から構成され, 光源を平行光源と仮定することで紙面上の陰影を排除する. 本手法においては, 紙面の最小主曲率方向が展開モザイク画像の画像座標 (m, n) において縦軸方向に射影されることから, 平行光源の仮定の下では, 展開モザイク画像上の陰影は, 同一の m 座標を取る画素に対して一様に現れる. また, 同一の m 座標を取る縦につながった画素群の中に, 紙面の背景色が必ず含まれていると仮定すれば, 以下の式によって, 各画素 (m, n) の輝度値 $I(m, n)$ を補正し, 陰影を取り除いた新たな画素値 $I_{new}(m, n)$ を算出することができる.

$$I_{new}(m, n) = \frac{I_{max} I(m, n)}{\max(I(m+u, n+v); \forall (u, v) \in W)} \quad (13)$$

ただし, W は縦長 (例えば横 $5 \times$ 縦 500 画素) のウィンドウを表し, I_{max} は補正された画像上での紙面背景の輝度値 (例えば典型的には 255) を表す. なお, 紙面上に紙の素材よりも反射率が高いものが存在する場合や, 紙面にフチなしで写真が印刷されている場合など, ウィンドウ W 内の最大輝度に対応する画素が必ずしも紙面の背景と対応しない場合には, 本ステップにおける処理を適用しても正しく陰影を排除することはできない.

3. 実験

提案手法の有効性を確認するため, プロトタイプシステムを作成し, 見開きの冊子とワイン上のラベルを対象として実験を行った. 試作システムは, 一般的なデスクトップ型 PC (Pentium Xeon 3.2GHz, Memory

2GB) と小型カメラ (Aplux C104T, VGA, IEEE1394 接続) から構成される. 本実験では, カメラの内部パラメータをあらかじめ Tsai の手法 [24] によって校正し, このパラメータを撮影中固定とした. カメラのシャッタースピードは, 本システムで用いられる典型的なカメラの動き速度に対して, 画像上での動きブレが発生しないように設定し (物理的な値はカメラの仕様上不明), その他カメラで自動調整可能な項目についてすべて撮影中固定とした. 実験は複数の光源が存在する室内環境で行い, 紙面に最も近い光源までの距離は約 2m であった. また, 式 (8) による最適次数決定のための定数は経験的に $C = 0.007$ とした.

3.1 見開きの冊子に対する実験

本実験で撮影対象として用いた湾曲した紙面を図 5 に示す. 実験対象は 2 つの湾曲した紙面から構成されており, 冊子左側が文章, 右側が写真・図から構成されている. また, 定量評価に用いるために, 紙面上の 4cm の格子位置に + マークをあらかじめ印刷してある. このような紙面を対象として, 手持ちのカメラで 200 フレームから成る VGA 解像度の動画画像を撮影し, 実験に用いた. 図 6 に撮影された画像の一部と検出された特徴点位置を示す. 図中の \times 印は検出・追跡された特徴点の位置を表す. これに対して 2.2 で述べた三次元復元手法を適用することで, 図 7 に示すカメラ外部パラメータと特徴点の三次元位置を推定した. なお, 図中の曲線はカメラの移動経路を, 四角錐は 20 フレーム毎のカメラの姿勢を, 点群は特徴点の三次元位置を表す.

次に, 2.3 で述べた再出現特徴点の検出・全体最適化と曲面当てはめ処理を繰り返し, 紙面形状を推定した. 本実験では, 3 回の繰り返し処理で推定結果が収束し, 最終的な入力画像上での特徴点の再投影誤差は平均 0.59 画素となった. 図 8 に紙面形状の推定結果を示す. 本研究では多項式近似によって紙面形状を推定するが, 本実験において幾何学的 AIC [23] を用いて自動決定された多項式の次数は, 紙面左側が 5 次, 紙面右側が 4 次であった.

次に, 推定されたカメラ外部パラメータと紙面形状を用いて平面状に展開した展開モザイク画像 (解像度: 3200×2199 画素) を生成した. ここでは, 式 (12) に示したモザイク画像生成時の画素の重み係数 K を 10^4 とした. 図 9 に, 生成されたモザイク画像について陰影を取り除く前後の結果を示す. 同図 (a) では, 幾何学的には湾曲が補正されているが, 陰影が不均一



図 5 実験対象とした湾曲した紙面
Fig. 5 Curved document as a target of the experiment.

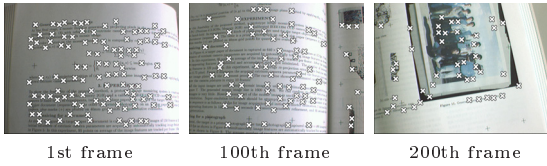


図 6 入力画像の一部 (紙面)
Fig. 6 Sampled frames of input images and tracked features (curved document).

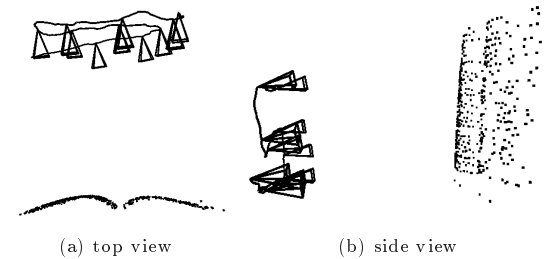


図 7 推定されたカメラの位置・姿勢と特徴点の三次元位置 (紙面)
Fig. 7 Estimated camera parameters and 3-D positions of features (curved document).

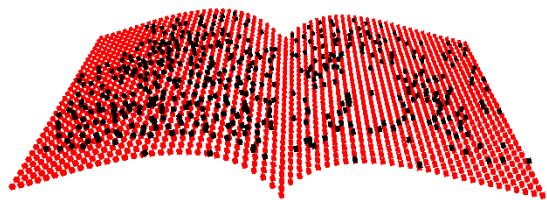
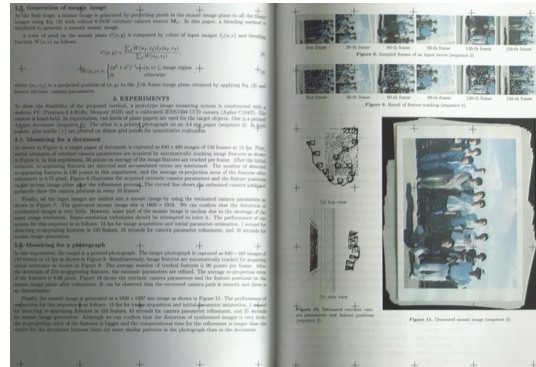
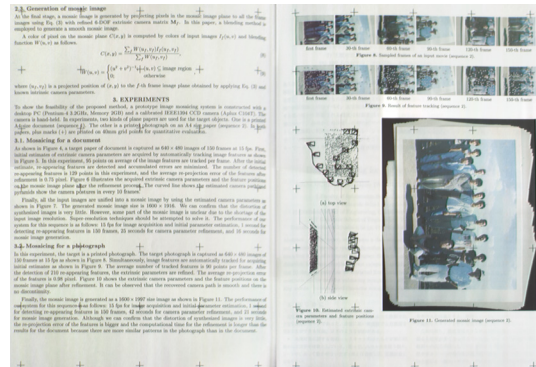


図 8 三次元点群から推定された紙面形状
Fig. 8 Fitted surface for 3-D point cloud (curved document).

であるために湾曲しているかのように知覚されてしまう。しかし、(b)では、紙面上の陰影は取り除かれ、湾曲した紙面がおおむね正しく平面状に展開されていることが確認できる。図 10 に、入力画像の一部の領域を拡大した画像 (a) と、それに対応するモザイク画像

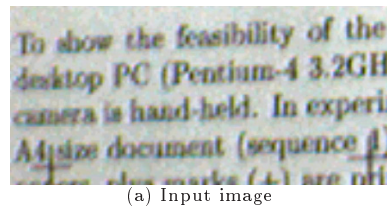


(a) before shading removal

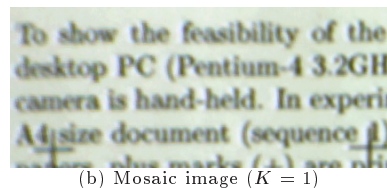


(b) after shading removal

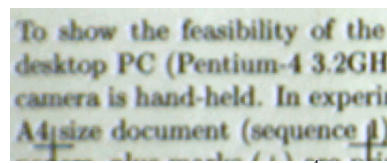
図 9 平面状に展開されたモザイク画像 (紙面)
Fig. 9 Generated un-wrapped mosaic image (curved document).



(a) Input image



(b) Mosaic image (K = 1)



(c) Mosaic image (K = 10⁴)

図 10 入力画像とモザイク画像の拡大比較
Fig. 10 Comparison between input image and un-wrapped mosaic image (curved document).

上の領域 (b), (c) を示す．なお，ここでは式 (12) で示したモザイク画像合成時の重みの効果を確認するため，重みを均一とした場合 (図中 (b), $K = 1$) および紙面の法線方向に応じた重みを加えた場合 (図中 (c), $K = 10^4$) の生成画像についてそれぞれ掲載した．図 10(a) から，入力画像上では幾何学的な歪みが存在し，文字のベースラインが平行ではないことが分かるが，(b), (c) においては歪みが取り除かれ，ベースラインが平行化されている．本実験では，動きブレの問題を回避するため比較的高速なシャッタースピードを設定したため，入力画像上で暗ノイズが発生しているが，(b), (c) では多数の画像を合成することによりこのようなノイズが低減されている．また，(a) と比較してより精細な画像が生成されており，文字の可読性が向上している．(b), (c) の比較においては，紙面の法線方向に応じた重みを加えた (c) の方が画像の合成ブレが小さいことから，重み付けによる画像合成の効果を確認できる．なお，本実験においては，ステージ (A) の逐次処理によるカメラ位置・姿勢推定に 27 秒 (平均 7.4fps)，ステージ (B) のカメラパラメータの最適化に 71 秒，ステージ (C) の展開モザイク画像の生成に 113 秒を要した．

最後に，あらかじめ紙面上の格子点位置に印刷しておいた + 印を用い，生成されたモザイク画像上に生じた幾何学的な歪みの定量的評価を行った．ここでは，生成されたモザイク画像上での + 印の座標を手作業で計測し，隣接する格子点間の距離および格子点間を結ぶ直線の成す角を算出することで歪みの評価を行った．表 1 に本実験における格子点間の距離の平均，最大，最小，標準偏差を，表 2 に縦横に隣接する格子点間を結ぶ直線の成す角の平均，最大，最小，標準偏差を示す．表 1, 2 から，紙面の内容によらずおおむね一定の精度でモザイク画像を生成できることが分かる．ただし，格子間を結ぶ直線の垂直からのずれは最大 1.1 度程度であり，実際，図 9(b) においても紙面の接合部付近に幾何学的な歪みが残されている．これは，紙面の接合部付近では，紙面の法線とカメラ光軸方向の成す角が 90 度に近くなるために特徴点追跡の精度が低下し，三次元復元における推定誤差が大きくなることに起因していると考えられる．

以上のことから，本手法では湾曲した紙面の電子化において問題となる幾何学的な歪みおよび陰影の影響を効果的に排除することができ，電子化された書籍の文字を人間が読み取るといった用途において問題ない

表 1 モザイク画像上の格子間距離 [単位:画素 (括弧内:平均からのパーセンテージ)]

Table 1 Distance between adjacent grid points in the mosaic image.

| target | average | maximum | minimum | std.dev. |
|------------|------------------|------------------|-----------------|----------------|
| left side | 356.4 (100.0) | 363.0 (101.8) | 350.0 (98.2) | 2.94 (0.82) |
| right side | 355.1 (100.0) | 361.0 (101.6) | 348.0 (98.0) | 2.86 (0.81) |

表 2 モザイク画像上で縦横に隣接する格子点間を結ぶ直線の成す角 [単位:度]

Table 2 Angles formed by adjacent grid lines in the mosaic image.

| target | average | maximum | minimum | std.dev. |
|------------|---------|---------|---------|----------|
| left side | 89.8 | 90.8 | 89.0 | 0.60 |
| right side | 89.6 | 90.5 | 88.9 | 0.67 |

レベルでの画像生成を実現している．しかし，紙面の接合部付近において幾何学的な補正が十分でなく，若干歪みが残されているため，OCR での文章の自動読み取りや，精密な紙面の電子化を目的とする場合には，特徴点以外の画素値情報の利用による紙面形状推定の高精度化を行い，更に精度の高い幾何学的歪みの排除を実現する必要がある．

3.2 ワインラベルに対する実験

本実験では，図 11 に示すようなワインボトルに貼り付けられたラベルを対象とし，手持ちのカメラで 100 フレームから成る VGA 解像度の動画画像を撮影した．図 12 に撮影された画像の一部と検出された特徴点位置を示す．図中の × 印は検出された特徴点の位置を表す．前節で示した実験結果と同様に，提案手法による三次元復元処理と形状推定処理を行うことで，図 13 に示す三次元復元結果と形状の推定結果を得た．図中の四角錐は 10 フレーム毎のカメラの姿勢を表す．なお，本実験における再投影誤差は平均 0.58 画素であり，幾何学的 AIC で決定された多項式の次数は 2 次であった．同図から，比較的特徴点の少ないワインラベルに対しても形状が推定されていることを確認できる．

次に，推定されたカメラ外部パラメータと紙面形状を用いて，平面状に展開した展開モザイク画像 (解像度:3200 × 2343 画素) を生成した．図 14 に，生成されたモザイク画像を示す．本実験においては，入力画像上で鏡面反射によるハイライトの発生が確認されたため，2.4.2 で述べた陰影の排除手法は適用しなかった．このため，最終的に生成された画像上において両隅で輝度が低くなっているが，幾何学的にはおおむね



図 11 実験対象としたワインラベル
Fig.11 Wine label as a target of the experiment.



1st frame 50th frame 100th frame

図 12 入力画像の一部 (ワインラベル)

Fig.12 Sampled frames of input image sequence (wine label).

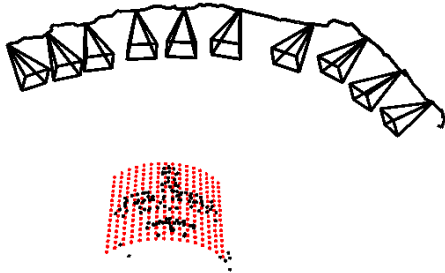


図 13 推定されたカメラパス・特徴点の三次元位置と紙面形状 (ワインラベル)

Fig.13 Estimated camera parameters and fitted surface (wine label).



図 14 平面上に展開されたモザイク画像 (ワインラベル)

Fig.14 Generated un-wrapped mosaic image (wine label).

正しく平面への展開が行われている．本来ワインラベルは円筒であるため，グラフ曲面を想定した式 (6) による 2 次多項式で正しい形状を推定することは原理的に不可能であるが，本実験のように円筒の一部分といった限られた範囲内であれば，このような近似を用いても大きな問題は発生しなかった．ただし，本実験では形状の推定に用いた特徴点が比較的少数であったことから，最小主曲率方向の推定誤差が比較的大きく，生成された展開モザイク画像が若干傾いていることが分かる．

4. まとめ

本論文では，手持ちのビデオカメラを用いて湾曲した紙面に対して平面状に展開したモザイク画像を生成する手法を提案した．本手法は，動画像からの三次元復元手法を用いることで，特殊な機材なしに自動で紙面の形状を推定する．また，1 軸方向のみの湾曲を仮定することで，推定された三次元点群に多項式近似による紙面形状を当てはめ，湾曲した紙面を平面状に展開する．更に，紙面の背景が白色で光源が無窮遠であるという仮定を用いることで，紙面上の陰影を排除し，光学的にも正しく展開されたモザイク画像を生成できることを示した．

提案手法では，紙面の形状パラメータの算出において多項式近似でのモデル化を行っているため，円柱の全周に描かれた絵画に対するモザイクのように多価関数でのモデル化が必要な形状に対して手法を適用することはできない．しかし，基本形状のモデルをユーザがあらかじめ選択するなどインタフェース面での工夫を行い，撮影対象に応じた形状モデルの当てはめを行うことで，一価関数による多項式近似が不可能な形状に対しても展開モザイク画像を生成できると考えられる．また，今後は特徴点位置の最適化だけではなく，モザイク画像上での同一の画素に対応する各フレームの画素値の一致度 (photo consistency) を考慮した三次元復元パラメータの最適化を行うことで，更なるモザイク画像の生成品質の向上を図る．

文 献

- [1] T. Wada, H. Ukida and T. Matsuyama: "Shape from Shading with Interreflections Under a Proximal Light Source: Distortion-Free Copying of an Unfolded Book," *Int. J. of Computer Vision*, Vol. 24, No. 2, pp. 125–135, 1997.
- [2] Z. Zhang, C. L. Tan and L. Fan: "Restoration of Curved Document Images through 3D Shape Modeling," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 10–15, 2004.
- [3] M. Pilu: "Undoing Paper Curl Distortion using Applicable Surfaces," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 67–72, 2001.
- [4] M. S. Brown and W. B. Seales: "Image Restoration of arbitrarily Warped Documents," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 10, pp. 1295–1306, 2004.
- [5] A. Yamashita, A. Kawarago, T. Kaneko and K. Miura: "Shape Reconstruction and Image Restoration for Non-flat Surfaces of Documents with a Stereo Vision System," *Proc. Int. Conf. on Pattern Recognition*, Vol. 1, pp. 482–485, 2004.
- [6] H. Cao, X. Ding and C. Liu: "A Cylindrical Surface Model to Rectify the Bound Document Image," *Proc. Int. Conf. on Computer Vision*, Vol. 1, pp. 228–233, 2003.
- [7] J. Liang, D. Dementhon and D. Doermann: "Unwarping Images of Curved Documents using Global Shape Optimization," *Int. Workshop on Camera-based Document Analysis and Recognition*, pp. 25–29, 2005.
- [8] M. S. Brown and Y. C. Tsai: "Undistorting Imaged Paint Materials using Boundary Information," *Proc. Asian Conf. on Computer Vision*, Vol. 1, pp. 551–556, 2004.
- [9] R. Szeliski: "Image Mosaicing for Tele-Reality Applications," *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 44–53, 1994.
- [10] N. Chiba, H. Kano, M. Higashihara, M. Yasuda and M. Osumi: "Feature-based Image Mosaicing," *Proc. IAPR Workshop on Machine Vision Applications*, pp. 5–10, 1998.
- [11] C. T. Hsu, T. H. Cheng, R. A. Beuker and J. K. Hong: "Feature-based Video Mosaicing," *Proc. IEEE Int. Conf. on Image Processing*, Vol. II, pp. 887–890, 2000.
- [12] U. Bhosle, S. Chaudhuri and S. D. Roy: "A Fast Method for Image Mosaicing Using Geometric Hashing," *IETE J. of Research, Special Issue on Visual Media Processing*, Vol. 48, No. 3-4, pp. 317–324, 2002.
- [13] 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", *電子情報通信学会論文誌 (D-II)*, Vol. J88-D-II, No. 8, pp. 1490–1498, 2005.
- [14] P. Grattoni and M. Spertino: "A Mosaicing Approach for the Acquisition and Representation of 3D Painted Surfaces for Conservation and Restoration purpose," *Machine Vision and Applications*, Vol. 15, No. 1, pp. 1–10, 2003.
- [15] W. Puech, A. G. Bors, J. M. Chassery and I. Pitas: "Mosaicing of Paintings on Curved Surfaces," *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 44–49, 1996.
- [16] 佐藤 智和, 池谷 彰彦, 池田 聖, 神原 誠之, 中島 昇, 横矢 直和: "動画画像からの三次元復元による湾曲した紙面のビデオモザイク", *画像の認識・理解シンポジウム (MIRU2006) 講演論文集*, pp. 98–105, 2006.
- [17] A. Iketani, T. Sato, S. Ikeda, M. Kanbara, N. Nakajima and N. Yokoya: "Video Mosaicing for Curved Documents Based on Structure from Motion," *Proc. IAPR Int. Conf. on Pattern Recognition*, Vol. IV, pp. 391–396, 2006.
- [18] T. Sato, M. Kanbara, N. Yokoya and H. Takemura: "Dense 3-D Reconstruction of an Outdoor Scene by Hundreds-baseline Stereo Using a Hand-held Video Camera," *Int. J. of Computer Vision*, Vol. 47, No. 1-3, pp. 119–129, 2002.
- [19] C. Harris and M. Stephens: "A Combined Corner and Edge Detector," *Proc. Alvey Vision Conf.*, pp. 147–151, 1988.
- [20] M. A. Fischler and R. C. Bolles: "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395, 1981.
- [21] P. J. Besl and N. D. McKay: "A Method for Registration of 3-D Shapes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239–256, 1992.
- [22] B. Triggs, P. McLauchlan, R. Hartley and A. Fitzgibbon: *Bundle Adjustment – A Modern Synthesis*, pp. 298–375, LNCS, Springer Verlag, 2000.
- [23] K. Kanatani: "Geometric Information Criterion for Model Selection," *Int. J. of Computer Vision*, Vol. 26, No. 3, pp. 171–189, 1998.
- [24] R. Y. Tsai: "An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 364–374, 1986.

(平成 xx 年 xx 月 xx 日受付)

佐藤 智和 (正員)

1999年阪府大・工・情報工卒。2003年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。現在、同大情報科学研究科助教。動画像からの三次元復元、奥行き画像推定、ビデオモザイクの研究に従事。2001年本会学術奨励賞受賞。博士

(工)。情報処理学会、IEEE各会員。

1992年奈良先端科学技術大学院大学・情報科学センター教授。現在、同大情報科学研究科教授。1990年情報処理学会論文賞受賞。2005年情報処理学会フェロー。工博。情報処理学会、日本バーチャルリアリティ学会、人工知能学会、日本認知科学会、映像情報メディア学会、IEEE、ACM SIGGRAPH各会員。

池谷 彰彦 (正員)

1997年阪大・工・電子制御機械卒。1999年同大学院博士前期課程了。同年NEC入社。以来、コンピュータビジョンの研究に従事。2004年本会学術奨励賞、PRMU研究奨励賞、FITヤングリサーチ賞受賞。

池田 聖 (正員)

2001年広島大・理・物理卒。2006年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。現在、同大情報科学研究科助教。カメラキャリブレーション、テレプレゼンスの研究に従事。博士(工)。IEEE会員。

神原 誠之 (正員)

1997年岡山大・工・情報工卒。2002年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了。現在、同大情報科学研究科助教。ステレオビジョン、拡張現実感、ウェアラブルコンピュータの研究に従事。1999年本会学術奨励賞受賞。博士(工)。

情報処理学会、日本バーチャルリアリティ学会、IEEE各会員。

中島 昇 (正員)

1990年慶大・理工・電気卒。1992年同大学院修士課程了。同年NEC入社。以来、文書画像処理の研究に従事。1996年本会学術奨励賞受賞。

横矢 直和 (正員：フェロー)

1974年阪大・基礎工・情報工卒。1979年同大学院博士後期課程了。同年電子技術総合研究所入所。以来、画像処理ソフトウェア、画像データベース、コンピュータビジョンの研究に従事。1986~87年マツギル大・知能機械研究センター客員教授。

Abstract A number of methods for video mosaicing have already been investigated. However, most of these methods often assume that the target object is flat or very far from the camera to avoid the disparity problem. This paper proposes a novel video mosaicing method for curved documents based on 3-D reconstruction by structure-from-motion. The proposed method can generate the mosaic image of geometrically restored target document, even if the document has a curved surface. Experiments on curved documents have shown the feasibility of the proposed method.

Key words video mosaicing, structure from motion, shape estimation, feature tracking, extrinsic camera parameter estimation