

LI-003

カメラパラメータ推定による超解像ビデオモザイク Super-Resolved Video Mosaicing Based on Camera Parameter Estimation

池谷 彰彦†
Akihiko Iketani

佐藤 智和†‡
Tomokazu Sato

池田 聖†
Sei Ikeda

神原 誠之†‡
Masayuki Kanbara

中島 昇†
Noboru Nakajima

横矢 直和†‡
Naokazu Yokoya

1. はじめに

カメラ付き携帯電話などの低解像度の撮像デバイスを用いて、手軽に高精細な文書画像を取り込める技術が求められている。手持ちの低解像度カメラで紙面を撮影した動画画像から、紙面の高解像度画像を生成する手法として、ビデオモザイクが提案されている[1], [2], [3]。これらの手法では、はじめに連続する2フレーム間で平面射影変換パラメータを推定する。続いて、入力映像中のあるフレーム(通常、初期フレーム)を参照フレームに選び、平面射影変換パラメータを用いて各フレーム画像を参照フレームと同一平面上に投影することで、モザイク画像を生成する。そのため、参照フレームにおいてカメラが紙面に対して傾いている場合、紙面全体を斜め方向から撮影したような、歪んだモザイク画像が生成されるという問題があった。歪みのないモザイク画像を生成するには、フレーム間の変換だけでなく、各フレームにおけるカメラと紙面との三次元的な位置関係を求める必要がある。

本稿では、カメラ外部パラメータ推定に基づくビデオモザイク手法を提案する。提案手法により、紙面に対して傾いたカメラで撮影された入力映像から、正対した高解像度のモザイク画像を生成することができる。

本手法の処理の流れは、次の通りである。はじめに、特徴点追跡結果から、各フレームのカメラ外部パラメータを推定する。次に、映像から一度消失した後、再出現した特徴点を自動検出し、消失前と再出現後の追跡結果を同一特徴点系列に統合する。続いて、映像全体でカメラパラメータの最適化を行い、推定誤差を最小化する。最後に、全フレーム画像をモザイク画像平面上に投影する。その際、超解像処理を適用することで、高解像度モザイク画像を生成する。

2. カメラパラメータ推定によるビデオモザイク

図1に、本手法の処理の流れを示す。はじめに、紙面を近接位置からなぞるように動かしたカメラの外部パラメータを推定し(A)、推定されたパラメータを用いてモザイク画像を生成する(B)。以下では、カメラ外部パラメータと誤差関数を定義した後、前記(A)、(B)の詳細について述べる。

2.1 カメラ外部パラメータと誤差関数

一般に、平面对象のモザイクにおいては、時間的に連続した2フレーム間の位置関係を、以下に示す8自由度の平面射影変換パラメータ \mathbf{H}_f で表す[1]~[3]。

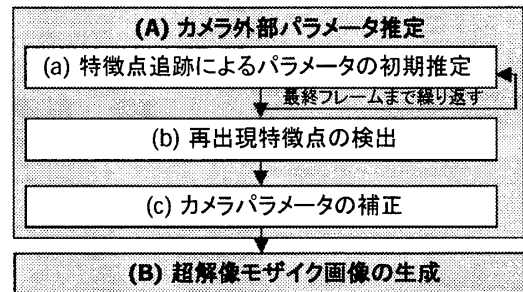


図1 提案手法の処理の流れ

$$(au_{(f+1)p}, av_{(f+1)p}, a)^T = \mathbf{H}_f(u_{fp}, v_{fp}, 1)^T \quad (1)$$

$$\mathbf{H}_f = \begin{pmatrix} m_{1f} & m_{2f} & m_{3f} \\ m_{4f} & m_{5f} & m_{6f} \\ m_{7f} & m_{8f} & 1 \end{pmatrix} \quad (2)$$

ただし、

a : パラメータ,

(u_{fp}, v_{fp}) : 第 f フレーム画像上の特徴点 p の2次元座標。

上式を用いれば、カメラ内部パラメータが未知の場合でも、2フレーム間の変換を定義できる。従来のモザイク手法では、この変換パラメータ \mathbf{H}_f を用いて、各フレームから参照フレーム(通常、初期フレーム)への変換を求め、この変換に基づいて各フレーム画像を参照フレームと同一平面上に投影し、モザイク画像を生成する。そのため、参照フレームにおいてカメラが紙面に対して傾いている場合、紙面全体を斜め方向から撮影したような、歪んだモザイク画像が生成されるという問題があった。従って、従来手法を用いる場合には、参照フレームにおいて、紙面とカメラを厳密に正対させることが重要となるが、手持ちのカメラでこれを実現するのは困難である。

この問題を解決するには、画像間の変換だけでなく、各フレームにおける紙面とカメラの三次元的な位置関係を推定する必要がある。そこで、本研究では、各フレームにおける平面射影変換パラメータのかわりに、カメラ外部パラメータを推定する。

モザイク画像平面と各フレームにおけるカメラとの関係を図2に示す。本研究では、モザイク画像平面を $z=0$ 平面上にとり、このモザイク画像平面と第 f フレームの画像平面との変換を、カメラ外部パラメータ \mathbf{M}_f と定義する。 \mathbf{M}_f は、通常的外部パラメータ行列と基本的に同じものであるが、ここでは被写体は $z=0$ 平面上に固定されているので、 z 軸に関するパラメータを除いたものを用いる。すなわち、 \mathbf{M}_f は次のように表される。

† NEC インターネットシステム研究所

‡ 奈良先端科学技術大学院大学 情報科学研究科

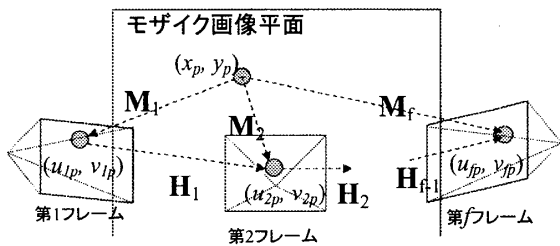


図2 モザイク画像平面とカメラの関係

$$(a\hat{u}_{fp}, a\hat{v}_{fp}, a)^T = \mathbf{M}_f(x_p, y_p, 1)^T \quad (3)$$

$$\mathbf{M}_f = \begin{pmatrix} c_1c_3 + s_1s_2s_3 & s_1c_2 & t_{1f} \\ -s_1c_3 + c_1s_2s_3 & c_1c_2 & t_{2f} \\ c_2s_3 & -s_2 & t_{3f} \end{pmatrix} \quad (4)$$

$$s_i = \sin(r_{if}), \quad c_i = \cos(r_{if}) \quad (i = 1, 2, 3) \quad (5)$$

ただし、

- (x_p, y_p) : モザイク画像平面上での特徴点 p の座標,
- $(\hat{u}_{fp}, \hat{v}_{fp})$: 第 f フレームにおける理想カメラ画像平面上での特徴点 p の座標,
- (u_{fp}, v_{fp}) : 第 f フレームにおける実カメラ画像平面上での特徴点 p の座標 ($(\hat{u}_{fp}, \hat{v}_{fp})$ から (u_{fp}, v_{fp}) への変換は、カメラの内部パラメータによって行われる),
- (t_{1f}, t_{2f}, t_{3f}) : 第 f フレームにおけるカメラの位置,
- (r_{1f}, r_{2f}, r_{3f}) : 第 f フレームにおけるカメラの姿勢.

次に、カメラ外部パラメータ \mathbf{M}_f の推定に用いる誤差関数を定義する。一般に、撮像時の量子化や、特徴点抽出時の誤差、およびカメラ外部パラメータの推定誤差により、式(3)と既知の内部パラメータによって求まる特徴点 p の座標 (u_{fp}, v_{fp}) と、実画像上での検出位置 (u'_{fp}, v'_{fp}) とは一致しない。そこで、本稿では、第 f フレームにおける特徴点 p に関する誤差 E_{fp} を次のように定義する。

$$E_{fp} = (u_{fp} - u'_{fp})^2 + (v_{fp} - v'_{fp})^2 \quad (6)$$

次節では、この E_{fp} を用いて、 \mathbf{M}_f および (x_p, y_p) を推定する方法について述べる。

2.2 カメラ外部パラメータの推定

図1に示すように、カメラパラメータ推定処理は3つの処理から構成される。以下では、各処理について述べる。

2.2.1 特徴点追跡によるカメラパラメータ推定

本処理では、特徴点を追跡することにより、各フレームにおけるカメラ外部パラメータ \mathbf{M}_f を推定する。なお、本処理は[4]の手法を拡張したものである。

はじめに、初期フレームにおける処理について述べる。画像平面は被写体と正対していると仮定し、 \mathbf{M}_f を単位行列に初期化する。撮影された特徴点のモザイク画像平面上での座標 (x_p, y_p) についても、この仮定に基づいて計算する。なお、この時点で画像平面と被写体とが正対していない場合には、誤ったカメラパラメータが推定されることになるが、これは後段の補正処理(c)によって補正されることになる。

次に、後続フレーム ($f > 1$) 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 \mathbf{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推定される。

特徴点の追跡 Harris オペレータ[5]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。さらに、RANSAC アルゴリズム[6]を用いて、アウトライヤを除去する。

カメラ外部パラメータ推定 特徴点追跡により、各特徴点のカメラ画像上での座標 (u'_{fp}, v'_{fp}) が得られる。これと、前フレームまでに推定されているモザイク画像平面上での特徴点座標 (x_p, y_p) を用いて、カメラ外部パラメータ \mathbf{M}_f を推定する。 \mathbf{M}_f は誤差関数 $\sum_p E_{fp}$ を非線形最小化することで推定される。

モザイク画像平面上の特徴点座標の推定 誤差関数 $\sum_{k=1}^f E_{kp}$ を (x_p, y_p) について最小化することにより、モザイク画像平面上での各特徴点 p の座標を更新する。

特徴点の追加と削除 正確なカメラパラメータを得るために、誤差関数の発生しにくい特徴点を選択することが重要である。ここでは、各特徴点の信頼度に基づいて、特徴点の追加、削除を行う[4]。

2.2.2 再出現特徴点の検出

撮影中はカメラが動いているため、通常、特徴点は画像上にフレームインしてから、しばらく追跡された後、フレームアウトしていく。中には、一度フレームアウトした後に、再度フレームインしてくるものが存在する。そこで、本研究ではこのような再出現特徴点を検出し、カメラパラメータの累積的な推定誤差を抑制するのに利用する。

再出現特徴点の検出は、時間的に離れたフレーム間で、類似のパターンをもつ特徴点が存在するかを検証することによって行う。その際、図3に示すように、カメラ運動の影響によって、紙面上の同じ特徴点であっても、パターンが変化してしまうことが問題となる。この問題に対し、本手法では次のようにして対処する。はじめに、すべての特徴点のテンプレートをモザイク画像平面上に投影する。次に、モザイク画像平面上での距離がしきい値より小さい特徴点のペアを選択し、両者の類似度を正規化相互相関によって評価する。その結果、相関が十分に高いものを再出現特徴点とする。この方法により、カメラ運動によって特徴点のパターンが回転したり、歪んだ場合でも、再出現特徴点の検出が可能となる。

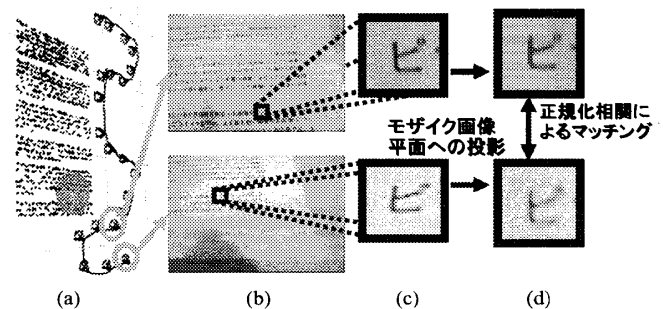


図3 再出現特徴点の検出

- (a) カメラの移動経路、モザイク画像平面上の特徴点
- (b) 入力映像中の時間的に離れた2フレーム
- (c) 各フレームにおける同一特徴点のテンプレート
- (d) モザイク画像平面上に投影された各テンプレート

2.2.3 カメラパラメータの補正

ここでは、全フレームでカメラ外部パラメータの最適化を行う。これにより、初期フレームにおいてカメラが紙面に対して正対していないことなどによる推定誤差が最小化され、カメラパラメータが補正されることになる。

累積推定誤差 E は、次式に示すように、特徴点座標をカメラ画像平面上に投影したものと、実際のカメラ画像上での特徴点座標との2乗誤差の総和によって与えられる。

$$E = \sum_f \sum_p E_{fp} \tag{7}$$

これをカメラパラメータ \mathbf{M}_f とモザイク画像平面上の特徴点座標 (x_p, y_p) について最小化する。なお、再出現特徴点については、最初の追跡結果と再出現後の追跡結果とを同一特徴点系列とみなし、誤差 E_{fp} を計算する。

2.3 超解像モザイク画像の生成

最後に、カメラ外部パラメータを用いて全フレーム画像をモザイク画像平面上に投影し、重複部分においてブレンディング処理を行う。これによって得られた画像を初期モザイク画像 $S^{(0)}$ とし、Iterative Back Projection アルゴリズムによる超解像処理[7]を適用することで、高解像度のモザイク画像 S を生成する。処理の流れは、次の通りである。

いま、 m 枚の入力画像 $\{I_f\} = I_1, I_2, \dots, I_m$ が存在するとする。はじめに、初期モザイク画像 $S^{(0)}$ に対して、カメラ外部パラメータ \mathbf{M}_f で定義される幾何変換、ぼかしから成る低解像度画像生成過程を適用し、各入力画像 $\{I_f\}$ に対応する低解像度画像 $\{I_f^{(0)}\}$ を生成する。ここで、 $S^{(0)}$ の推定誤差が大きければ大きいほど、 $\{I_f^{(0)}\}$ と $\{I_f\}$ のずれは大きくなる。そこで、入力画像と推定画像間の誤差画像 $\{I_f - I_f^{(0)}\}$ を計算し、誤差画像の画素値を $S^{(0)}$ 中の対応する箇所へ逆投影 (Back Projection) し、加算することで、更新された高解像度画像 $S^{(1)}$ を得る。この処理を、次式で表される誤差関数が十分に小さくなるまで繰り返す。

$$e^{(n)} = \sum_f \sum_{u,v} (I_f(u,v) - I_f^{(n)}(u,v))^2 \tag{8}$$

ここで、 n は反復計算の回数を表す。以上の処理により、高解像度のモザイク画像を得る。

3. 実験

提案手法に基づくビデオモザイクシステムを試作した。本システムは、デスクトップ PC (CPU: Pentium-4 3.2GHz、メモリ:2GB) と、事前のキャリブレーションによって内部パラメータが既知の IEEE1394 CCD カメラ (Aplux C104T) とから構成される。この試作システムを使い、論文 (実験 1) と写真 (実験 2) の二つについて実験を行った。なお、いずれの対象も A4 サイズであり、定量評価 (後述) のために、紙面上に + 印のグリッドを 40mm 間隔で配置したものを用いた。

各実験における入力映像を図 4 に示す。いずれも手持ちのカメラによって 15fps で撮影された 150 フレームから成る映像であり、画像サイズは 640×480 ピクセルである。なお、実験 1 では、映像全体を通して、カメラを紙面に対して傾けた状態で撮影を行った。

本手法によって得られたカメラ外部パラメータ、および各特徴点のモザイク画像平面上での位置を図 5 に示す。図中の曲線は、推定されたカメラの移動経路を表し、錐体は 10 フレームおきのカメラの姿勢を表す。提案手法により生成されたモザイク画像を図 6 に示す。画像サイズは、それぞれ 1600×1916 ピクセル (実験 1) と 1600×1997 ピクセル (実験 2) であった。図 6(a) より、入力映像中にカメラと紙面が正対しているフレームが存在しなくても、正対したモザイク画像を生成できることが確認できる。

また、各実験における入力画像の一部と、モザイク画像中の対応箇所をそれぞれ図 7、図 8 に示す。いずれの実験においても、入力画像より解像度が向上しているのが確認できる。

最後に、モザイク画像に生じた歪みの定量評価を行った。ここでは、モザイク画像上でのグリッド間距離によって歪み量を評価した。各実験におけるグリッド間距離 (ピクセル) の平均、最大値、最小値および標準偏差を表 1 に示す。標準偏差は歪み量の平均を表すと考えられ、表 1 より、いずれの実験においても、歪み量が 1.7 ピクセル程度にまで抑えられているのが確認できる。

表 1 モザイク画像上でのグリッド間距離 (単位: pixel)

対象	平均	最大値	最小値	標準偏差
論文	229.0	233.5	226.0	1.65
写真	245.6	249.0	243.0	1.27

なお、両実験において、処理時間は概ね同じであった。各処理に要した時間は、以下の通りである。

- 画像入力およびカメラ外部パラメータの初期推定: 1 フレームあたり 67 ミリ秒 (15fps の実時間処理)
- 再出現特徴点の検出: 1 秒 (150 フレーム全体で)
- カメラパラメータの補正: 25 秒
- モザイク画像の生成 (ブレンディング): 16 秒
- モザイク画像の生成 (超解像 反復 3 回): 110 秒

4. まとめ

カメラパラメータ推定による超解像ビデオモザイク手法を提案した。本手法により、紙面に対して傾いたカメラで撮影された入力映像から、正対した高解像度のモザイク画像を生成することができる。実験により、その有効性を確認した。今後は、モザイク画像のさらなる高解像度化、および処理の高速化に取り組む。

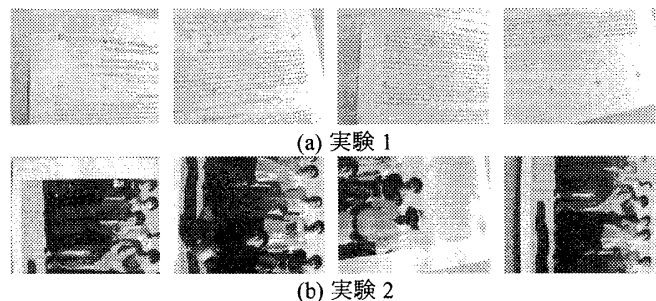
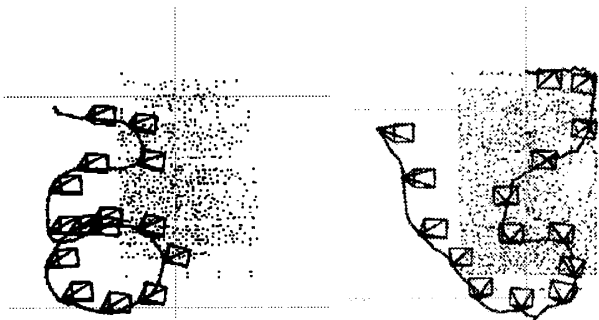


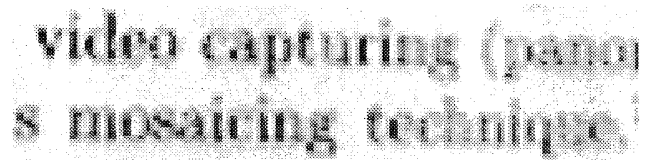
図 4 入力映像 (一部フレームのみ)



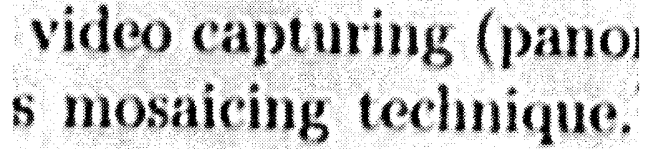
(a) 実験 1

(b) 実験 2

図 5 カメラパラメータと特徴点座標

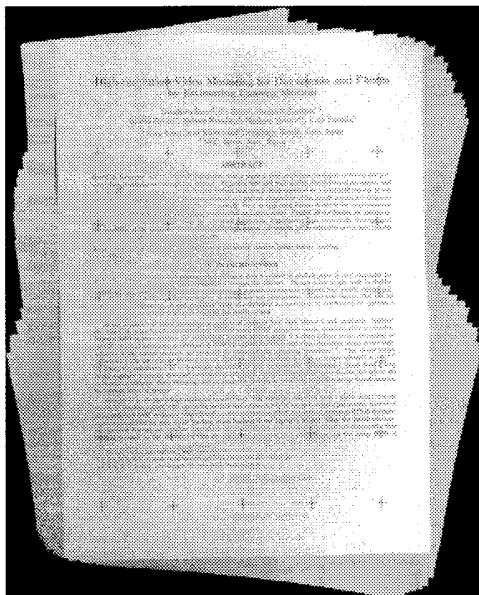


(a) 入力画像の一部

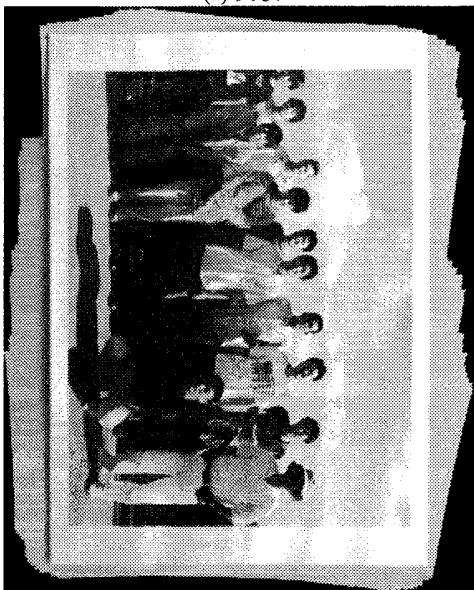


(b) モザイク画像中の対応箇所

図 7 実験 1 の入力、モザイク画像の拡大図



(a) 実験 1

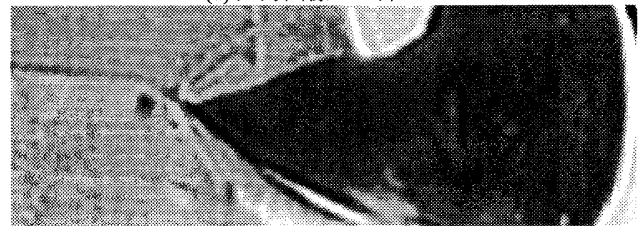


(b) 実験 2

図 6 生成されたモザイク画像



(a) 入力画像の一部



(b) モザイク画像中の対応箇所

図 8 実験 2 の入力、モザイク画像の拡大図

文 献

- [1] R. Szeliski: "Image Mosaicing for Tele-Reality Applications," Proc. IEEE Workshop on Applications of Computer Vision, pp. 230-236, 1994.
- [2] 千葉, 蚊野, 美濃, 安田: "画像特徴に基づくイメージモザイク", 電子情報通信学会論文誌 (D-II), Vol. J82-D-II, No. 10, pp. 1581-1589, 1999.
- [3] D. Capel and A. Zisserman: "Automated Mosaicing with Super-resolution Zoom," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 885-891, 1998.
- [4] 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. Jour. of Computer Vision, Vol. 47, No. 1-3, pp. 119-129, 2002.
- [5] C. Harris and M. Stephens: "A Combined Corner and Edge Detector," Proc. Alvey Vision Conf., pp. 147-151, 1988.
- [6] 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.
- [7] M. Irani and S. Peleg: "Improving Resolution by Image Registration", CVGIP: Graphical Models and Image Processing, Vol. 53, No. 3, pp. 231-239, 1991.