

三次元情報を媒介とした複数画像の同時編集手法

Consistent Multiple Image Editing via Reconstructed Three Dimensional Information

本田 浩一郎 五十嵐 健夫*

Summary. We introduce a new image editing technique for an unorganized set of photographs simultaneously, by propagating edit operation on an image to others. In particular, we reconstruct 3D geometric information of a set of photographs to find corresponding pixels between images and apply the same (or similar) edit operation on them. To reconstruct 3D information, we adopt standard Computer Vision techniques, and do not rely on any hardware devices which provide geometric information (like GPS or electric compasses). In the reconstruction process, we calculate SIFT feature points and robustly estimate a fundamental matrix using eight-point algorithm on each pair in the set, then integrate each pair on an incremental Structure-from-Motion approach. In our system, however, users don't have to think about spatial relationship between photographs. For instance, when a texture-mapping effect is made onto a region in an image, the corresponding region on each other image is calculated immediately in the background, then applied the same effect to it. Apparently it is merely a simple task, but it may be far more difficult than expected to do the same thing manually, because of its visual consistency. With our system, faster and more precise editing can be created to multiple images. We demonstrate the effectiveness with several examples including a set of portrait, scene, and other natural photographs.

1 はじめに

ここ数年の間、デジタルカメラの普及は、急速に進んできた。内閣府の消費動向調査 [10] によれば、世帯主が 60 歳未満の一般家庭におけるデジタルカメラの普及率は、2009 年 3 月現在で 80 % を超えており、ほとんどパソコンの普及率と変わらなくなってきている。記憶媒体は急激に大容量化が進み、撮影写真の枚数は増える一方、デジタルカメラ本体は軽量で小型になり、持ち運びも便利になってきている。デジタル写真がより身近なものになっていると言えるだろう。

こうしたなかで、写真編集ツールが占める役割は大きい。フィルムカメラと違い、できあがった写真に対して自由な加工を加えられるデジタル写真は、より魅力的に見えるような画像補正を行ったり、デジタルコンテンツの素材として、表現の一部に使用されたりもしている。また、CEA の調査 [3] によれば、消費者の 51 % は何らかの手段を用いて写真を共有しており、共有する際に何らかの画像加工を行う機会も増えていると考えられる。

現行の画像編集ソフトは、1 枚ずつ画像を編集することが前提に設計がされており、複数枚の写真についてのバッチ処理はあまり充実していない。標準的な画像編集ソフトにはバッチ処理の機能はつけられているが、その多くは、色調補正やリサイズなど、

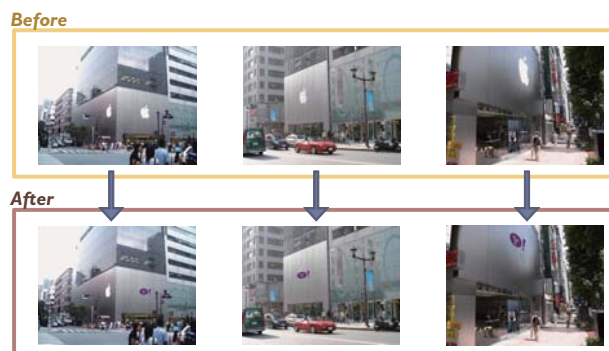


図 1. システムの概要．同じ対象を撮影した写真群に対し、一貫性のある画像編集を行う。

画像全体に処理を行うかんたんなもの限定されている。少しでも高度な処理を行わせようとすると、手動で行わなければならないのが現状である。我々が注目しているのは、同じ対象を撮影した複数の写真を同時に編集したいという場合である。それぞれの写真に写った同じ物体に対して、処理をしたい(例えば色を変える、画像を合成するなど)と考えるとき、手動で行うのは非常に手間がかかり、かつ難しい作業となる。編集したい物体は各写真内でどのように写り込んでいるか、つまり、大きさや角度、ゆがみの違いを考慮しなければならない。どの写真で見ても自然に、すなわち見た目の一貫性を保って編集を行うには、画像編集に慣れていなければなかなか難しい。2 枚ならそこまで難しくはないが、5 枚 10

Copyright is held by the author(s).

* Koichiro Honda and Takeo Igarashi, 東京大学大学院情報理工学系研究科

枚になってくると、各写真への編集を、見た目の一貫性を保ちながら行うのは殆ど不可能である。そこで我々は、複数の写真間で見た目の一貫性を保ちながら編集を行うツールを提案する。

そのために我々は、写真群を三次元的に再構成して、写真間で物体の位置の対応を取ることによって、同じ編集を他の写真にも伝搬させて加えるということ考えた。単純なバッチ処理ではなく、同じ物体に対して同時に処理を施すインテリジェントなバッチ処理とも言える。図1に簡単なイメージ図を示す。

三次元構成には、標準的なコンピュータビジョンの技法を用いており、GPS や電子コンパスなどの機器を必要としない。再構成のプロセスでは、写真群から特徴点を計算し、それを元に画像間で基礎行列を推定する。その後、再帰的な Structure-from-Motion の手続きによって写真群の位置関係を構築する。編集を伝搬する過程では、ユーザは各画像の空間的な位置関係を考える必要はなく、一枚の画像のある領域に編集を施すと、他の画像での対応する領域はバックエンドで即座に計算され、同じ処理が適用される。これは一見してシンプルな処理であるが、手動で同じことを行おうと、見た目の一貫性を保つために、非常に難しい作業となる。提案手法はこのような複数枚の画像群に対し、より速く正確に編集を加えることが出来る。我々は、肖像写真や、風景の写真など、様々な写真群の例を通じて、提案手法がより効果的であることを示す。

2 関連研究

2.1 複数写真の一括編集

複数写真に編集を加える機能をもつソフトウェアは数多く存在する。画像編集ソフトの中で最も代表的な Adobe Photoshop[2] は色調補正やフィルタなど標準で持ち合わせた編集機能からアクションと呼ばれる処理ルーチンを作成し、同社の画像管理ソフト Adobe Bridge[1] 内で複数の画像に対して自動処理を適用することが出来る。しかしこれは、純粋に全く同じ動作を行うため、画像のある部分を選択するなどの処理は、画像内の「同じ物体」ではなく「同じピクセルの位置」に対して処理が行われるため、使用する用途は限定的である。多くは画像のリサイズや画像全体のトーン補正等に使われている。代表的な画像編集・管理ソフトは Aperture など他にも多く存在するが、基本的な Photoshop/Bridge と同等な機能しか有していない。我々の目指す編集ツールは、同じピクセル位置に対して編集を行うというシンプルなバッチ処理ではなく、同じと考えられる場所に対して同じ処理を行うことを目的としている。このような写真の空間的な関係性に着目したバッチ処理をもつソフトは現状のところ存在していない。

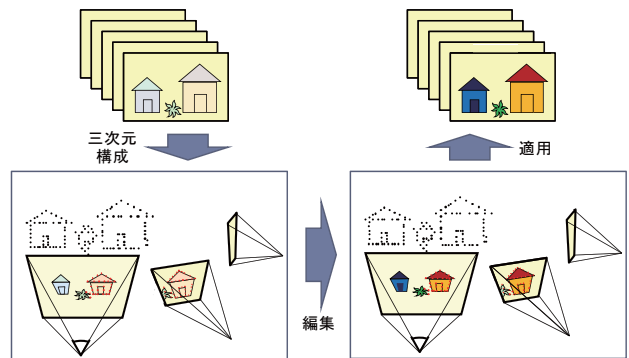


図 2. 三次元情報を用いた処理の伝搬の概要。

2.2 複数写真の三次元再構成

写真を三次元的に再構成する研究は、近年にわかに注目を浴びてきた。Devebec らは、幾何モデルを考慮したステレオマッチングを行うことで、写真群から簡単な三次元モデルを構成した [4]。Dick らは、同じ建物の写真群から三次元モデルを作成したが、ベイズ理論を応用して自動的に生成できるようにした [5]。近年では Sinha らがより直感的でインタラクティブにモデルを作成できる手法を提案し、スケッチベースで簡単にクリックを繰り返すだけで比較的正確な三次元モデルを作成できるようになった [6]。Snavely らはオンラインアルバム flickr[9] から収集した画像群に対して、特徴点のみからカメラ変数の推定を行い、空間上にマップして直感的なブラウジングができるビューアを作成した [8, 7]。我々も、カメラ変数の推定を行う際には本手法を用いている。だが、これらの研究は三次元形状を獲得したり、空間上の位置を推定したりすることが目的であり、画像自体の編集はしていない。本稿の焦点は、このようなコンピュータビジョンの手法を補助的に使い、複数の画像に対して、よりインテリジェントな編集を加えようという点にある。

3 ユーザインタフェース

本システムでは、ユーザは特に難しい計算や画像編集のテクニックを要求されないように設計している。ユーザがすべきことは、様々な角度から同じ対象を複数枚写真に納め、その画像ファイルをシステムに読み込むだけであり、三次元情報は読み込み時に自動的に復元が行われる。処理に関しても同様で、従来の画像処理ソフトのように、2次元のインタフェースで編集作業を行うだけでよい。処理の伝搬にまつわる三次元空間上の位置関係や、位置の対応を取るための計算は、あくまでもバックエンドの処理であり、ユーザには基本的には提示されない。必要に応じて三次元ビューも表示することが出来るが、三次元空間上で画像の編集を行うのはあまり直感的でないため、通常は使用しないことにした。シ

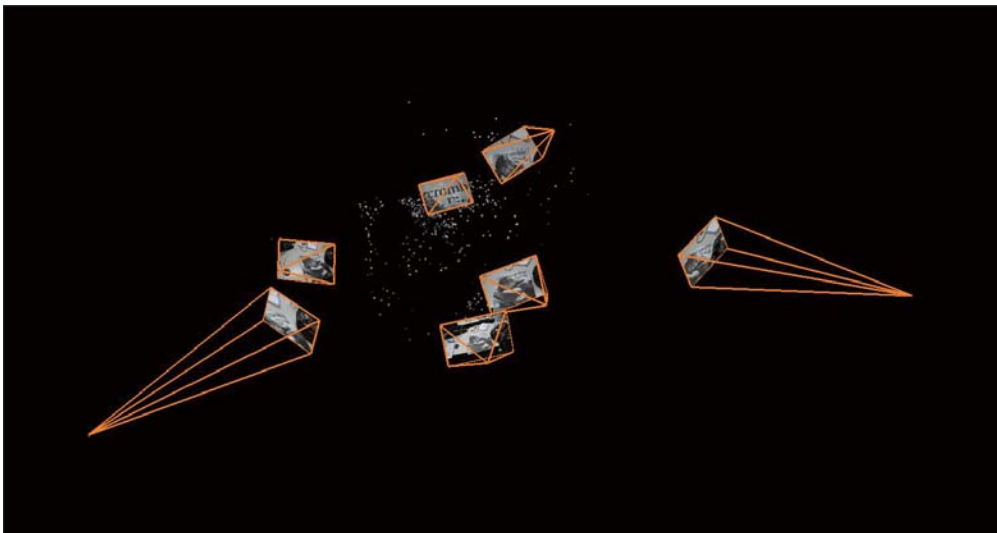
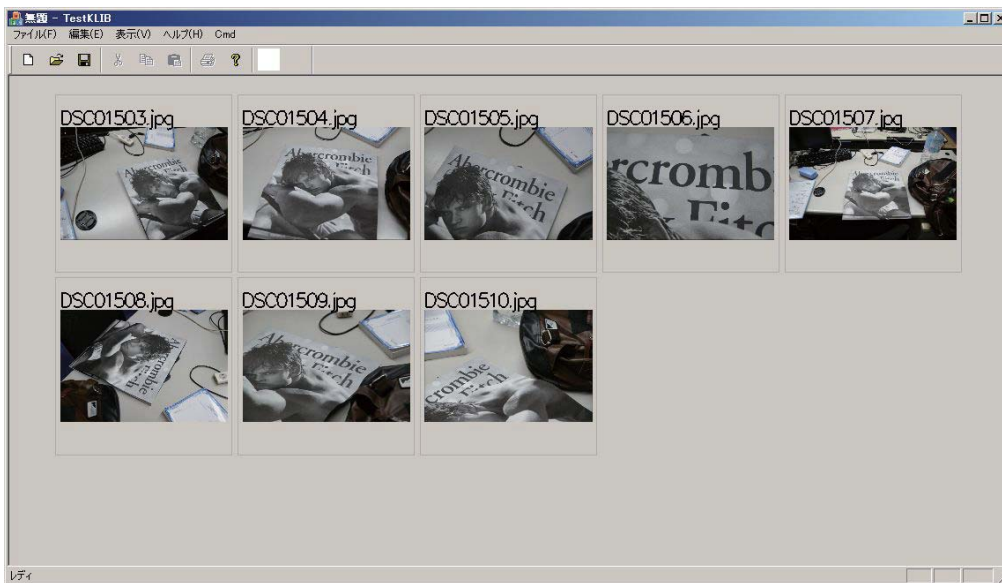


図 3. 本システムのインターフェース．上が通常編集に使用する二次元のインターフェース．下が三次元インターフェース．推定されたカメラ位置にマッピングされている．

システムの外観を図 3 に示す．

なお、本システムでは、伝搬させる機能として、最も基本的な画像編集の機能である、画像合成について実装を行った．

4 一貫性を保つ編集の伝搬アルゴリズム

本システムでは、前処理として、写真群から三次元位置を推定して写真間の対応を取る．

4.1 三次元情報の再構成

本稿では Snavely らの手法を元に、三次元位置の推定を行っている．その推定は以下の様に行われる．はじめに (1) 写真群は SIFT 特徴点記述子に基づいて特徴点の抽出を行う．(2) その後、画像群の中

で 2 枚のペアについて基礎行列の推定を行う．これは抽出した 2 枚の画像の SIFT 特徴に対して、近似最近探索を用いてマッチングを取る．そして、8 点アルゴリズムを用いて基礎行列を推定する．これで、2 枚の画像間の位置関係がわかる．なお、この過程では RANSAC を用いてロバストに推定を行っている．また、どの画像とも SIFT 特徴のマッチングが取れない画像に関しては再構成のプロセスから除かれる (3) そして基礎行列の推定が終わると、全てのペアを組み合わせる束調整 (Bundle Adjustment) を行う．ここでは、組み合わせた画像群の中で対応が取れている SIFT 特徴点の空間的な位置に対して、再投影の誤差を最小化するような最適化問題を解く．用いているアルゴリズムは Levenberg-Marquardt

法であり、コンピュータビジョンの分野では比較的良好に用いられているアルゴリズムの一つである。このプロセスについての詳細は文献 [8] を参照されたい。

4.2 編集処理の伝搬

前節にて、写真群は空間的な位置関係と特徴点の対応がとれたが、編集の過程ではその位置関係の情報を元に、編集の伝搬が行われる。一つの画像に加えられた編集を、他の画像の対応する位置に適用するには、「編集した範囲だけ」伝搬すればよい。色調補正であろうが、画像の合成であろうが、どのような処理であっても、対応する位置さえ適切に選択することが出来れば、どのような処理であっても伝搬が可能である。従って、本稿では選択範囲、即ち、選択した位置の伝搬を行う。

4.2.1 位置の伝搬

選択された位置を伝搬させるには、近傍の特徴点から重心座標系 (Barycentric Coordinates) を定義し、そのパラメータの情報の伝搬を行う。画像 S で点 P_s が選択されたときに、画像 T で対応する点 P_t を見つけることを考える。ここで、点 P_s の座標を \mathbf{x}_s とし、点 P_t の座標を \mathbf{x}_t とする。はじめに、画像 S と画像 T で共通して持っている特徴点群の中から、重心座標系 (Barycentric Coordinates) を定義できる近傍の特徴点 3 点 R_1, R_2, R_3 を見つける。そしてその後、下記のように重心座標系のパラメータ $\lambda_1, \lambda_2, \lambda_3$ を計算する。

$$\mathbf{x}_s = \lambda_1 \mathbf{r}_1 + \lambda_2 \mathbf{r}_2 + \lambda_3 \mathbf{r}_3 \quad (1)$$

$$\mathbf{x}_t = \lambda_1 \mathbf{r}'_1 + \lambda_2 \mathbf{r}'_2 + \lambda_3 \mathbf{r}'_3 \quad (2)$$

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \quad (3)$$

そして、画像 T 上で、 R_1, R_2, R_3 に対応する 3 点 R'_1, R'_2, R'_3 から、同じパラメータ $\lambda_1, \lambda_2, \lambda_3$ を使って、 P_t を計算している。

この作業を、元の画像 S から、画像群の中の他の画像にそれぞれ伝搬を行っている。ここで、特徴点は、画像群の各画像すべてが同じ特徴点を持っているとは限らないため、伝搬に使われる特徴点は、伝搬先の画像によって異なる。本システムではその都度、伝搬元と伝搬先の画像で共有している特徴点の中から 3 点近傍の特徴点を探してきて使用している。

5 実験

本稿では、いくつかの画像群に対して、実験を行った。ここでは、ドアに貼られたポスターと、机に置かれた紙袋の 2 つの画像群の例を示す。伝搬させる機能としては、最も基本的な画像編集の機能である、画像合成について実装を行った。画像合成は、前述の伝搬のアルゴリズムを用いて、貼り付ける画像の四隅の点を伝搬させている。ユーザが画像群の

中の一枚に、画像合成処理を行うと、自動的に他の画像にも、処理が行われる。以下、2 つの画像群について実験した結果を示す。2 つの画像群はどちらも Sony 社製デジタル一眼レフカメラ DSLR-A300 にて撮影を行ったものであり、計算の際には 640×480 px 程度の大きさにダウンサンプリングして用いている。また、計算機的环境は CPU が Core 2 Duo T9800 2.93GHz、メモリが 4GB、GPU が GeForce 9300M GS、OS が Windows Vista Business 32bit である。計算時間は、編集の伝搬についてはリアルタイムに行っており、三次元構成については、以下 2 つの写真群ともに 10 秒以下である。

5.1 ドアに貼られたポスター

はじめの例は、ドアに貼られたポスターの画像群である。枚数は 6 枚である。図 5 左に示す骨と書かれたポスターに対して、違うポスターを貼り付けている。貼り付けるポスターは図 4 右である。編集は最下段右側の画像に対して図 4 右の画像の合成を行った。他の画像への合成は位置の伝搬によって自動的に合成されたものである。この画像群は左右から撮った画像に写っている物体がかなり異なっており、特徴点の抽出やマッチング等に影響があるとも考えられるが、比較的良好に処理の伝搬が来ている。



図 4. 張り替えるポスター。左が編集前に映っていたポスター、右が張り替えたいポスター。

5.2 机の上に置かれた紙袋

もう一つの例は、机の上においた紙袋を撮った写真群である。図 6 に元の写真群と編集後の結果を示す。枚数は 8 枚であるが、紙面の関係上 6 枚を示す。適用した編集は、最上段左側の紙袋に寄った写真に対して、文字の一部を黒く塗りつぶす動作を行ったものである。他の画像は、伝搬によって同様に塗りつぶされたものである。これだけ小さい処理となると、遠方から撮られた写真では画質が荒くなってしまい、うまく処理が行えない可能性も考えられるが、比較的良好に伝搬が来ていると言えるだろう。

6 議論

本稿で提案した画像編集の伝搬手法は、複数画像の同時編集を目的としたものであり、既存の画像編集手法を置き換えるものではない。むしろ、編集処理の伝搬を「処理範囲の伝搬」と考えると、プロトタイプとして実装した画像合成だけでなく、色調補正など既存の編集手法の多くを、複数画像に適用することができるようになるため、従来の画像編集と本システムは補完的な関係にあるといえる。

また、本手法にはいくつか解決すべき課題も存在する。ひとつには、本手法は位置伝搬のために、近傍の特徴点によって定義された重心座標系を用いているが、特徴点が疎な対応しか取れていない（すなわち、画像間でピクセル単位の密な対応が取れているわけではない）ため、画像のどの位置についても正確な伝搬が行えるとは限らない。また、本手法では平面上で画像合成をする例を示したが、複雑な形状をした写真群についてはうまく伝搬が行うことができない可能性が高いことが挙げられる。

今後の発展としては、プロトタイプとして実装した画像合成だけでなく、様々な機能の追加を行い、実用性を高めるとともに、対応の精度を上げるべく、より適切な特徴点の選択や、伝搬手法の選択を検討していく必要があると考えられる。

7 まとめ

本稿では、複数枚の画像セットに対して、再構成された三次元情報を元に、一貫性を保った同時編集をする手法を提案した。具体的には、一枚の画像に加えられた編集作業を他の画像にも伝搬することによって、編集箇所を考慮した高度なバッチ処理を実現した。また、その編集作業のプロトタイプとして画像合成の処理を実装し、実際の画像群に対して実験を行った。

参考文献

- [1] Adobe. Bridge. <http://www.adobe.com/>.
- [2] Adobe. Photoshop. <http://www.adobe.com/>.
- [3] CEA. Digital Imaging: A Focus on Sharing, 2009.
- [4] P. Debevec, C. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *Proceedings of ACM SIGGRAPH 1996*, 1996.
- [5] A. Dick, P. Torr, and R. Cipolla. Modelling and Interpretation of Architecture from Several Images. *International Journal of Computer Vision*, 60, 2004.
- [6] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, and M. Pollefeys. Interactive 3D architectural modeling from unordered photo collections. *Proceedings of ACM SIGGRAPH Asia 2008*, 2008.
- [7] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections. *International Journal of Computer Vision*, 2007.
- [8] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. In *Proceedings of ACM SIGGRAPH 2006*. ACM New York, NY, USA, 2006.
- [9] Yahoo! flickr. <http://www.flickr.com>.
- [10] 内閣府. 消費動向調査 (平成 21 年 3 月調査), 2009.



図 5. ドアに貼られたポスターの例 . 左側が編集前 , 右側が編集後 . ポスターの位置に新しいポスターを貼り付けている .



図 6. 机に置かれた紙袋の例 . 左側が編集前 , 右側が編集後 . 各写真に小さく写っている文字の一部を編集し , 伝搬させた .