

安定・高精度なオンライン SfM

鳥居秋彦[†] 奥富正敏[†]

[†] 東京工業大学

東京都目黒区大岡山 2-12-1

E-mail: †{torii,mxo}@ctrl.titech.ac.jp

あらまし 本研究では、PC に画像が入力される度に、次々とカメラポーズ・3次元空間点の復元を行うオンライン Structure from Motion(SfM) システムを提案する。提案システムでは、ワイドベースラインマッチングによるロバストな特徴点对応付けを用いることで、非連続的なカメラ運動により撮影された画像に対しても、安定な復元を可能とする。さらに、順次処理する入力画像に対してマッチングを行う画像ペアを効果的に選択することで、復元の安定性を保ったまま計算量を削減し、オンラインでの復元を実現する。画像ペアの選択基準として、時間的拘束と見かけの類似性による拘束の両方を用いることで、広範囲の画像から効果的な特徴点のリンクが行われる。この広範囲に渡る特徴点リンクを利用することで、オンラインでのループクロージングが可能となり、より高精度な復元を行える。発表では、一般的なデジタルカメラ画像、Ladybug、魚眼カメラ、様々な種類の画像からの復元をビデオデモとライブデモ両方の形式で紹介する。

キーワード Structure from Motion, Wide Baseline Matching, SURF, RANSAC, Bundle Adjustment

1. はじめに

複数画像からカメラ位置姿勢(カメラポーズ)と空間点位置の復元を行う Structure from Motion(SfM) 研究は、成熟を迎えている。特に予め用意されたデータセットに対するオフライン型 SfM に関しては、街全体という規模での復元が可能である [1], [2]。これらの研究で提案された SfM は、数千～数万枚の画像を 2-4 時間程度で処理することが可能な、非常に効率の良いシステムであるものの、あくまで閉じた入力データを想定している^(注1)。また、システムの設計上、一定の処理段階に到達するまで、3D モデルを出力することは不可能である。

一方、時系列画像に対し、オンライン型の SfM を行う様々なシステムが提案されてきた [3], [4]。実時間で復元を行うシステムも存在するが、車載カメラによる撮影や、限定された室内環境における撮影など、ある程度限定された撮影環境が想定されている。

本研究では、より柔軟な撮影条件、つまり、ユーザが復元を特に意識することなく撮影した画像やビデオ映像を入力として想定した、オンライン SfM を提案する。様々な撮影条件に適応可能とするため、ワイドベースラインマッチングを基本とした特徴点对応付け(マッチング)、複数画像間における特徴点の結びつけ(リンク)を行う。時間的拘束と見かけの類似性による拘束の両方を考慮し、効果的にマッチングを行う画像ペアの選定を行うことで、計算コストを抑えながら、安定性と精度を保った復元を実現する。

(注1)：一通りの復元処理が終了後、画像を追加して復元を拡張することは可能である。

2. オンライン SfM の概要

本研究では、オフライン SfM アルゴリズムを拡張、発展、効率化を図ることで、安定・高精度に復元を行うオンライン SfM を提案する。最先端のオフライン SfM で用いられているアルゴリズムを基盤とすることで、様々なカメラ運動、画像の見かけの変化にも適応可能な SfM を実現可能になる一方、オンライン処理を実現するためには、システム全体の効率化が重要な課題となる。

ここで、近年成熟されつつあるオフライン SfM のアルゴリズムを概略する。(1) 画像ペアマッチングにより、各画像間の特徴対応付けを行う。(2) 画像ペア間の特徴点の情報を伝搬し、複数画像間に渡る特徴点のリンクを行う。(3) 上記2ステップにより対応付け問題を解決した上で、カメラポーズ・空間点の復元を進める。オフライン型 SfM として、最も著名な復元システムである Bundler [5], [6] では、全ての画像ペアに対してマッチングを行うため、マッチングを行う画像ペアの数が指数的である。この課題を解決するために visual words and vocabulary [7], [8] を利用した類似画像検索を行い、予め類似度の高い画像ペアを選択し、それらのペアのみに対しマッチングを行うことで、計算コストを押さえた SfM が提案されている [1], [2], [9], [10]。オフライン型 SfM であっても、一度復元処理が始まれば、途中経過で生成される 3D モデルを出力することは可能である [11]。しかしながら、処理すべき画像が徐々に入力され、実用に堪えうる処理時間で順次 3D モデルを出力するような、本当の意味でのオンライン SfM とは、各ステップのアルゴリズム・システム設計の思想が異なる。

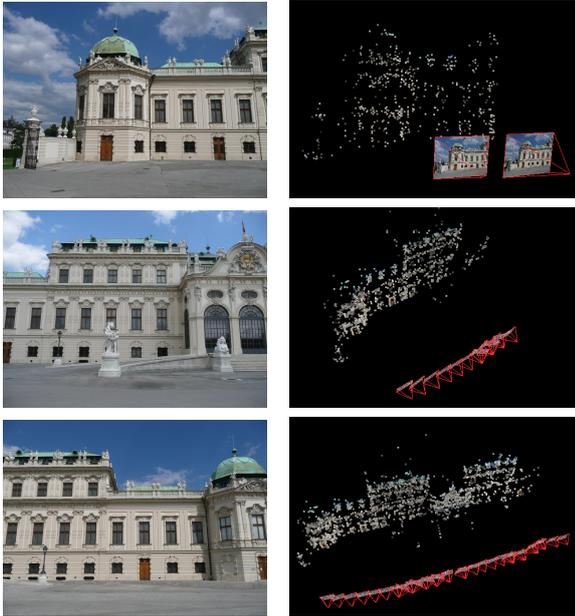


図1 Vienna Dataset: 入力画像とオンライン復元結果

オンライン SfM において、復元精度を損なうことなく、高速に復元を行うためには、マッチングを行う画像ペアの効率的・効果的な選択が鍵となる。本研究では、順次処理する入力画像に対し、マッチングを行う画像ペアの選択基準として、時間的拘束と見かけの類似性による拘束の両方を用いる。すなわち、時間的に隣接する画像と画像類似度の高い画像を数枚ずつ選択する。このバランスの良い画像ペア選択法により、一定の計算効率を保ったまま、広範囲の画像から効果的な特徴点のリンクが可能となる。広範囲に伝搬した特徴点リンクを用い、バンドル調整を行うことで、カメラ軌跡・空間点のループクロージングをオンラインで行うことが可能である。このことにより、オンライン SfM の宿命的な課題と言える誤差の累積を軽減し、高精度な復元を実現する。

その他の提案システムの特徴として、主に以下の2点が挙げられる。全自動で SfM を行う上で、安定かつ高精度に初期復元を行うことは非常に重要である。復元点から算出される視差角を用いたカメラ運動量推定アルゴリズム [12] を用いることで、ベースラインの十分に広い画像ペアを自動的に選択し、安定に初期復元を行う。さらに、システム全体の統一の規格として球面カメラを基にしたアルゴリズムを設計することで、ひとつのモジュール切り替えで、魚眼カメラ、全方位カメラ、Ladybug カメラなど様々なカメラモデルを利用可能な汎用性の高い SfM システムを実現している。

最後に実験結果例を図1, 2に示す。発表では、汎用デジタルカメラ、魚眼レンズカメラ、Ladybug など様々なカメラ画像を入力として実験を行った結果のビデオデモとライブデモの両方を行う。

文 献

- [1] S. Agarwal, N. Snavely, I. Simon, S. Seitz and



図2 Pittsburgh Dataset: 入力画像とオンライン復元結果

- R. Szeliski: “Building Rome in a day”, ICCV09, pp. 72–79 (2009).
- [2] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik and M. Pollefeys: “Building rome on a cloudless day”, ECCV10, pp. 368–381 (2010).
- [3] D. Nistr, O. Naroditsky and J. Bergen: “Visual odometry for ground vehicle applications”, Journal of Field Robotics, **23**, p. 2006 (2006).
- [4] G. Klein and D. Murray: “Parallel tracking and mapping for small AR workspaces”, ISMAR07, Nara, Japan (2007).
- [5] N. Snavely: “Bundler: Structure from motion (sfm) for unordered image collections”, <http://phototour.cs.washington.edu/bundler/> (2008).
- [6] N. Snavely, S. Seitz and R. Szeliski: “Modeling the world from internet photo collections”, IJCV, **80**, 2, pp. 189–210 (2008).
- [7] J. Sivic and A. Zisserman: “Video Google: Efficient visual search of videos”, CLOR06, pp. 127–144 (2006).
- [8] D. Nister and H. Stewenius: “Scalable recognition with a vocabulary tree”, CVPR06, pp. II: 2161–2168 (2006).
- [9] X. Li, C. Wu, C. Zach, S. Lazebnik and J. Frahm: “Modeling and recognition of landmark image collections using iconic scene graphs”, ECCV08, pp. I: 427–440 (2008).
- [10] M. Havlena, A. Torii, J. Knopp and T. Pajdla: “Randomized structure from motion based on atomic 3D models from camera triplets”, CVPR09, pp. 2874–2881 (2009).
- [11] M. Havlena, A. Torii and T. Pajdla: “Efficient structure from motion by graph optimization”, ECCV10 (2010).
- [12] A. Torii, M. Havlena and T. Pajdla: “Omnidirectional image stabilization for visual object recognition”, IJCV, **91**, 2, pp. 157–174 (2011).