複数のカメラモデルのための Structure from Motionの構築

前田尚活1 内山 英昭1 長原 一1 谷口 倫一郎1

概要:アルゴリズムとコンピュータ性能の向上により,Structure from Motion (SfM) はインターネット 上の大量の画像においても適用できるようになった.一方で近年,魚眼レンズのカメラが一般に普及して おり,透視投影カメラ以外で撮影された特殊な画像も、今後インターネット上に集まると考えられる.そ こで,本研究では複数のカメラモデルの画像を入力として3次元復元を行うためのSfMのフレームワーク を提案する.初めに,Affine-SIFTを用い複数のカメラモデルで撮影された全入力画像から特徴点を検出 し、全画像間で対応付けを行う.この特徴点を球面座標系に変換することで異なるカメラモデルを同じ枠 組みで考えることができ、その後の処理の多くを球面座標系上で行うことで、各画像の外部パラメータと 撮影シーンの3次元復元を行う.提案手法は、球面座標系を導入することでカメラモデルに依存しない3 次元復元を行うとともに内部パラメータも同時に最適化する、汎用性の高いフレームワークである.実験 では、透視投影カメラと魚眼カメラの両方を用いた画像群に対して提案手法を適用し、各カメラのみを用 いた場合と比較することで、手法の有効性を検証した.

Structure from Motion for Generic Camera Models

Maeda Naoyuki¹ Uchiyama Hideaki¹ Nagahara Hajime¹ Taniguchi Rin-ichiro¹

1. はじめに

入力した画像群から撮影シーンの3次元点群と各カメ ラの位置・姿勢を同時に推定するStructure from Motion (SfM)は、Multi-view stereoの前処理[1][2][3]や屋内外で の位置認識[4]にも利用される基盤技術である.SfMの多 くは、カメラ以外の機器を必要としないために様々な場面 で使用することができ、今日まで様々な研究が行われてき た.近年では、アルゴリズムの改善やコンピュータ性能の 向上によって、インターネット上から自動で集めた画像に 対してSfMを適用したり[5]、1つの街を復元するような 大規模なSfMも可能となってきている[3][6].

前述のようにこれまでの研究によって SfM は多彩なシー ンや多数の入力画像に対しても適用可能となった.これら の SfM のほとんどは透視投影カメラを対象として設計され たものである.一方で,RICOH THETA*1 や Gear 360*2 をはじめとする魚眼カメラが普及しており,今後透視投影 カメラ以外のカメラで撮影された画像が増加していくと予 想される.しかし,これまでに提案された SfM のほとんど は透視投影カメラのカメラ幾何を前提としているため他の カメラモデルで撮影された画像を利用可能とすることは難 しく,今後増えていくであろうこれらの画像を使用するこ とはできない.透視投影画像以外を対象とした SfM につ いてもこれまでに研究されてきたが,これらの手法は透視 投影画像に適用することができない [12].

本研究の目的は、複数のカメラモデルで撮影された画像 を同時に利用可能な、汎用性の高い SfM のフレームワーク を提案することである.今日までに撮影された透視投影画 像とこれから増加すると予想されるそれ以外の画像の双方 を利用可能とすることで、より多くの画像に対応した SfM の構築を目指す.過去に提案された SfM はの多くは1つ

 $[\]overline{*^1}$ https://theta360.com/

^{*2} http://www.samsung.com/global/galaxy/gear-360/

のカメラモデルのみを対象としており,複数のカメラモデ ルを対象とした SfM も,適応できるカメラモデルに制限が あったり [14] 内部パラメータの最適化を考慮に入れていな いといった問題点があった [15]. そこで本論文では,複数 のカメラモデルを利用可能かつ内部パラメータの最適化ま で考慮に入れた SfM を提案する.提案するフレームワー クでは,はじめに Affine-SIFT [7]を用いて全画像の特徴点 を検出し,全画像間で対応付けを行う.次に,異なるカメ ラモデルを同じ枠組みで考えるために特徴点を球面座標系 に投影し,球面座標点を用い各2画像間で3次元復元を行 う.最後に,求めた2画像間の3次元復元結果から全体の 3次元復元を行い,各カメラの内部パラメータ,外部パラ メータ,各特徴点の世界座標点群を最適化する.

なお、本論文では2枚の画像ペア間で復元したカメラの 位置姿勢を相対的な位置姿勢あるいは相対的な外部パラ メータと呼称し、世界座標系で見たカメラの位置姿勢を絶 対的な位置姿勢あるいは絶対的な外部パラメータと呼称 する.

2. 関連研究

本章では、アルゴリズムに着目した SfM の分類と、扱う カメラに着目した分類の 2 つの観点から SfM の従来研究 をまとめる.

2.1 Incremental SfM \succeq Batch SfM

SfM の手法は大きく分けて, Incremental SfM と Batch SfM の2つに分類できる. Incremental SfM では,入力画 像から2枚の画像を選択して3次元復元し,新しい画像を 次々と追加していくことで復元結果を徐々に拡大する.こ の枠組みの復元精度は最初に選ぶ2枚の画像に大きく依存 するため,2枚の画像の選び方が重要となる. Snavely ら は撮影位置が離れており,かつ特徴点の対応付けが十分に 取れている2画像が最初の3次元復元には適していると考 え,ホモグラフィ変換を用いて2画像を選択する方法を提 案した [5].また,Beder らは2画像間の3次元復元精度の 安定性を評価することで,最初の復元に適した2枚の画像 を見つけ出す手法を提案した [8].

Batch SfM では、複数の画像ペア間で相対的な外部パラ メータを求め、すべての相対的な外部パラメータに対して 整合性が保たれる絶対的な外部パラメータを求めることで、 3 次元復元を行う.絶対的な外部パラメータの導出法が復 元精度に大きく影響するため、これまで様々な方法が提案 されてきた.Govinduは、回転の四元数とカメラ位置の外 積を使用して絶対的な外部パラメータを求める方法 [9] や、 リー代数を利用した方法 [10] を提案した.また、Crandall らは一部の入力画像のメタデータに含まれる GPS 情報を 利用して、高精度に絶対的な外部パラメータを推定する方 法を提案した [11].

2.2 SfM に用いられるカメラモデル

SfM を扱うカメラモデルに着目し分類すると,透視投影 カメラを対象とした SfM とそれ以外を対象にした SfM に 大別できる.過去に提案された SfM の多くは,透視投影 カメラを対象にしたものであった.Snavely らは,前述し た Incremental SfM の初めの 2 画像を選択する手法を用い ることによって,時系列順でない透視投影画像群を入力と して処理のすべてを自動で行う SfM のフレームワークを 提案した [5].また,Frahm らはインターネット上から集 めた大量の透視投影画像を効率よく扱うことのできる SfM を提案した [6].

一方で,魚眼カメラや Catadioptric カメラを対象とした SfM も研究されている. Scaramuzza らは,キャリブレー ションされた Catadioptric カメラで撮影された,時系列 順に並んだ画像を対象とした SfM を提案した [12].また, Mičušík らは様々な種類の全方位カメラを対象とし,同じカ メラで撮影した 2 枚の画像のための SfM を提案した [13].

いくつかの研究グループでは、異なるカメラモデルで撮影 した画像を対象とした SfM も研究されてきた. Bastanlar らは、Parabolic-Catadioptric カメラと透視投影カメラで撮 影した画像を同時に扱うことができる SfM を提案した [14]. しかし、この手法は2種類のカメラモデルにしか対応して おらず、Hyperbolic-Catadioptric カメラや魚眼カメラに対 応することはできない. また、Pagani らは球面座標系上で のエピポーラ幾何を用いることで、透視投影カメラや魚眼 カメラなどの光学中心が画像中央にあるカメラモデルに適 用可能な柔軟性の高い SfM を提案した [15]. しかし、この 手法は各画像の内部パラメータが正確に求まっていること を前提としており、内部パラメータの最適化については考 慮されていない.

3. 提案手法

本章では、複数のカメラモデルを統合した SfM の提案 フレームワークについて述べる.初めに、3.1章でフレー ムワークの概要を述べる.3.2章では異なるカメラモデル を統一的に扱うための球面座標系について述べ、3.3章で 2 画像間の3次元復元を行う方法について述べる.最後に、 3.4章で Batch SfM により全体の3次元復元結果を得る方 法について述べる.

3.1 概要

前述したように、これまでに提案されてきた SfM のほと んどは1つのカメラモデルのみを対象としたものであった. 複数のカメラモデルで撮影された画像を入力とできる SfM も、対象とするカメラモデルが制限されていたり、内部パ ラメータの最適化を考慮していないといった問題があった. そこで本論文では、これまでに提案されてきた SfM の技術 を統合することで、複数のカメラモデルで撮影された画像



図 1: 提案手法のフローチャート



図 2: 球面座標系.カメラ座標点 X_c は透視投影カメラと 魚眼カメラの場合でそれぞれ $x_{pers.}$ と $x_{fish.}$ に投影される

を同時に入力とすることができ、内部パラメータの最適化 まで考慮した新たな SfM のフレームワークを提案する.

提案手法ではまず,複数のカメラモデルで撮影された画 像と各画像のカメラモデル,粗い内部パラメータを入力と して,各2画像間での3次元復元を行う.この復元は文 献[5]で述べられている2画像間の3次元復元手法を,文 献[15]で述べられている球面座標系で使用できるように 拡張することで,高精度に画像ペア間の3次元復元を行 う.次に,精度の高い3次元復元を目指すため文献[16]の Batch SfM を適用することで,全体の3次元復元結果を求 め,最後に Bundle Adjustment により復元結果を最適化す る.提案手法の流れを図1に示す.

3.2 球面座標系

異なるカメラモデルで撮影した画像から3次元復元を行 うことが難しい原因の1つに、レンズに入射した光が画像 上のどの画素に投影されるかは、カメラモデルによって異 なることが挙げられる.例えば、投影点と光学中心(*cx*,*cy*) の距離*r*は、図2に示すように透視投影カメラと魚眼カ メラでは透視投影カメラの方が小さくなることが知られて いる.

提案手法では、このような投影点の異なるカメラモデル

を統合的に処理するために,球面座標系を導入する [15]. 球面座標系は,図2に示すように各画像のカメラ座標系の 原点を中心とした単位球で定義され,球面座標点はこの単 位球上の座標で表現される.画像座標点 $\mathbf{x} = (x, y)^T$ から 球面座標点への投影は

$$\boldsymbol{x}_{s} = \begin{bmatrix} x_{s} \\ y_{s} \\ z_{s} \end{bmatrix} = \begin{bmatrix} \cos\varphi\sin\theta \\ \sin\varphi\sin\theta \\ \cos\theta \end{bmatrix}$$
(1)

で表現される.ここで、 $\varphi = \arccos \frac{v}{x}$ で計算され、 θ は各 カメラモデルの投影式から求める.したがって、各画像の カメラモデルと内部パラメータが既知であれば、画像座標 点を球面座標点へ変換することができる.

3.3 2 画像間の 3 次元復元

本節では,複数のカメラモデルで撮影された入力画像群 からなるすべての画像ペアについて,文献 [5] を基に高精 度に2画像間の3次元復元結果を求める方法について述べ る.はじめに,各画像の特徴点・特徴量を検出し,各画像 間で特徴点の対応付けを行う.次に,特徴点を前述した球 面座標系に投影することで異なるカメラモデルを同じ枠組 みで考え,エピポーラ幾何と特徴点追跡によって対応付け られた特徴点の中から誤対応を検出し削除する.そして最 後に,残った球面座標点の対応付けから,相対的な外部パ ラメータおよび撮影シーンの3次元点群を復元する.

すべての入力画像に対し Affine-SIFT により特徴点およ び特徴量記述を求める [7].次に,全ての画像ペアに対し て特徴量マッチングを行いて画像ペア間で特徴点を対応付 ける.対応点探索には近似最近傍探索法 [17] を用いる.こ のとき,誤対応を減らすために,求めた対応付けのうち一 方の特徴点に対し他方の特徴点以外に特徴量空間上で類似 した特徴点のあるマッチングは削除することを考える.ま ず,一方の画像の各特徴点に対して他方の画像から特徴量 空間上での距離が小さい順に特徴点を2つ求める.ある特 徴点に対し,特徴量空間上で最も距離の小さい特徴点 x_1 との距離を d_1 ,2番目に距離の小さい特徴点 x_2 との距離 を d_2 とすると, $d_1/d_2 > th$ となるマッチングは, x_1 と x_2 が類似していると判断し,削除する.なお,本研究では th = 0.6として実験を行った.

エピポーラ拘束により誤った特徴点の対応付けを削除す る.まず,3.2章で述べた方法により,各画像の各特徴点 を球面座標系へ投影する.2画像間で対応付いている球面 座標点 $x_{s}, x_{s'}$ に対し,ノイズのない理想的な対応点では エピポーラ拘束式 $x_{s'}^T E x_s = 0$ が成り立つ[15].したがっ て,この式から大きく外れる対応付けを削除することで, 誤対応を削除する.ここで,**E**は基本行列である.次に, Random Sample Consensus (RANSAC)[18]により外れ 値判定された対応付け数が最も少ない最適な基本行列を求



図 3: 曖昧性を含む追跡結果

める. 基本行列の導出には8点アルゴリズムを用い,外れ 値判定には球面座標系原点から球面座標点方向のベクトル とエピポーラ平面の角度を用いる.この角度の計算には次 式を用い,角度が閾値以上であればその対応付けは外れ値 と判定する [15].

$$\epsilon_p = \frac{|\boldsymbol{x}_s^T \boldsymbol{E} \boldsymbol{x}_s|}{||\boldsymbol{x}_{s'}||||\boldsymbol{E} \boldsymbol{x}_s||}$$

$$\theta = \arcsin \epsilon_p$$
(2)

なお、特徴点の対応付けが8点未満で基本行列を求めるこ とができない画像ペアと、最適な基礎行列計算後に正しい と判定された対応付けが8点未満の画像ペアに関しては、 2枚の画像の間には共通のシーンが写っていないと判断し、 すべての対応付けを削除する.

これまでは、特徴点マッチングにおける誤対応を2枚の 画像ペアから検出することを考えたが、ここでは入力画像 すべてを用い誤対応を除去することを考える. すべての 画像ペアの特徴点マッチングの結果を用い、各画像の各 特徴点を複数枚の画像に渡って追跡する。追跡結果は、文 献 [5] にならい Track と呼ぶ. Track を構成する際に特徴 点の誤った対応付けが含まれる場合,図3のように Image 1と Image 2, Image 2と Image 3の特徴点の対応付けと, Image 1 と Image 3 の対応付けの結果が一致せず、1 つの 追跡結果に対して同じ画像中の2つ以上の特徴点が対応す ることがある.この場合,追跡結果を構成する特徴点の対 応付けの中には少なくとも1つの誤対応が含まれているた め、追跡結果を構築した対応付けはすべて削除する. さら に,追跡枚数が一定枚数未満の追跡結果は不安定な追跡結 果とみなし、その追跡結果を構築したマッチングはすべて 削除する.実験では、3枚未満の追跡結果は不安定な追跡 とした.

残った特徴点の対応付けすべてを用い,8点アルゴリズ ムにより基本行列を導出し,基本行列を分解することで相 対的な外部パラメータを得る.しかし,基本行列の分解で は2つの回転行列と2つの並進ベクトルの計4パターン が解の候補として得られることが知られている[19].そこ で,正しい解を見つけるため,2画像間の対応付いた球面 座標点と各解の候補を用い三角測量を行い3次元点を導出 する.一般的には,求めた3次元点のカメラ座標系での座 標の奥行が双方のカメラで正になる候補を正しい解とする が,画角が180度以上あるカメラを対象とする場合はそれ が正しい解であるとは限らない.そこで提案手法では,求



図 4: 撮影シーン全体の3次元復元のフローチャート

めた3次元点のカメラ座標と球面座標点の内積が双方のカ メラで正になる解を正しい解であるとし、4つの候補の中 で双方が正になる対応付けが最も多いものを正しい解とし て選択する[15]. なお、選択した正しい外部パラメータで 双方の内積が正にならない対応付けは誤対応として削除 する.

3.4 撮影シーン全体の3次元復元

本節では、3.3 章で求めた相対的な外部パラメータから、 文献 [16] に基づいて絶対的な外部パラメータおよび撮影 シーンの世界座標点群を導出する方法について述べる. 導 出の流れを図4に示す. はじめに,相対的な回転行列から 絶対的な回転行列を導出する方法について述べる. 次に, 絶対的なカメラ位置の導出法について述べる. 最後に絶対 的な外部パラメータから撮影シーンの世界座標点群を復元 する方法について述べ,最後に絶対的な外部パラメータお よび世界座標点群の最適化について述べる.

各画像ペアの相対的な外部パラメータより, Martinec らの手法 [20] に基づいて各画像の絶対的な回転行列を導出する. 誤差のない理想的な回転行列は,

$$\boldsymbol{R}_j - \boldsymbol{R}_{ij} \boldsymbol{R}_i = \boldsymbol{0}_{3 \times 3} \tag{3}$$

を満たす.ここで, R_i , R_j は画像 *I*,画像 *J*の絶対的な 回転行列, R_{ij} は画像 *I*,*J*間の相対的な回転行列である. $R_i = [r_i^1, r_i^2, r_i^3]$ と置くと,式 3 は

$$\boldsymbol{r}_{i}^{k} - \boldsymbol{R}_{ij}\boldsymbol{r}_{i}^{k} = \boldsymbol{0}_{3\times 1} \tag{4}$$

と書き換えることができ、1つの画像ペアについて9個の 方程式を立てることができる.したがって、式4を各画像 ペアについて考え、それらをまとめて最小二乗法で解くこ とで、絶対的な回転行列を得る.なお、未知パラメータの 数を減らすため、ある画像の絶対的な回転行列は単位行列 とした.

絶対的なカメラ位置の導出には、Jiang らの手法 [16] を 用いる. この手法では図 5 に示すように 3 つの画像ペアか らなる三角形、triplet に注目し、スケールを一意に定めた



絶対的なカメラ位置を求める.ここで、図5の c_i, c_j, c_k は 絶対的なカメラ位置である. c_i を通り c_k へ向かうベクト ル c_{ik} と c_i を通り c_k へ向かうベクトル c_{ik} は、理想的に は1点交わる.しかし、実際には誤差の影響により図 5a に 示すように、ベクトル c_{ik} と c_{jk} はほとんどの場合で1点 では交わらず、ねじれの関係になる.この手法では、図5 の3つの図それぞれでねじれの関係となっている2つのべ クトルの最接近点が絶対的なカメラ位置であると考え,3 つの最接近点を求めることで、絶対的なカメラ位置を導出 する. なお, この手法では求めた絶対的なカメラ位置は符 号についての不定性がある.この問題を解決するために本 研究では、正しい符号ではカメラへの入射光と対応する球 面座標点は理想的に同じ方向を向くことを利用する.まず Trackの追跡結果から全追跡数 10%をランダムに抽出し, 求めた外部パラメータを用いこれらの追跡点の世界座標点 の座標を三角測量により求める.そして,求めた世界座標 点をその点が見えている各画像のカメラ座標へ変換し,各 カメラ座標点と対応する球面座標点の内積がすべて正の値 になる世界座標点数を数え、過半数以上であれば絶対的な カメラ位置の符号は正しいと判断し、そうでなければ符号 を逆にする.

前述した絶対的な回転行列,カメラ位置の導出法の復元 精度は,相対的な外部パラメータの復元精度に大きく依存 する.そこで,エピポーラ幾何が正しく求まらなかったと 思われる信頼性の低い画像ペアをJiangらの手法 [16] に基 づき3つの方法により検出し削除する.なお,1つ目と2 つ目の方法は世界座標系での回転を求める前に,3つ目は 世界座標系でのカメラ位置を求める前に実行する.

(1) Triplet の成す角度による検出

Tripletを構成する3つの画像ペアのエピポーラ幾何の推定 精度が高ければ、図5に示す各角の角度 $\theta_i, \theta_j, \theta_k$ と、相対 的な外部パラメータのみから求めた各角の角度 $\theta''_i, \theta''_j, \theta''_k$ は近しい値になるはずである.したがって、各 triplet そ れぞれで前述した方法で回転行列とカメラ位置を求めて $\theta_i, \theta_j, \theta_k$ を計算し、 $\theta''_i, \theta''_j, \theta''_k$ との差の平均が閾値より大き ければ、その triplet は信頼性が低いと判断する.さらに、 図6に示すように、3枚の画像間で対応付いている球面座 標点と求めた外部パラメータで三角測量を行い、求めた3 次元点と対応する各球面座標点の成す角が閾値以下の点が



図 6:3 次元点と各画像の球面座標点の角度

1 点もない triplet も信頼性が低いと判断する. 最後に, 1 度も信頼性が高いと判断されなかった画像ペアは推定した エピポーラ幾何の信頼性が低いと判断し, 削除する. なお, 実験では2つの閾値はそれぞれ 6°と5°に設定した.

(2) 全域木による検出

最大全域木を利用し、重要度の低い画像ペアを削除する. 各画像を頂点として相対的な外部パラメータが既知の画 像間にエッジを引いたグラフを構築し、各エッジを所属す る triplet の数で重みづける.重みの小さなエッジは、その エッジに依存する triplet が少ないため全体に与える影響が 少なく、削除しても問題ないといえる.したがって、グラ フから抽出した最大全域木に triplet を構成する3つのエッ ジのうち2つが含まれていれば、その3つのエッジは重要 度が高いと判断し、一度も重要度が高いと判断されなかっ たエッジ、すなわち画像ペアは削除する.

(3) 画像ペアの成すループによる検出

画像 I, 画像 J の絶対的な回転行列 $\mathbf{R}_i, \mathbf{R}_j$ と, 画像 I, J 間の相対的な回転行列を \mathbf{R}_{ij} の間には, 誤差を含まない理 想的な環境では次の関係が成り立つ.

$$\boldsymbol{R}_{ij}^T \boldsymbol{R}_j \boldsymbol{R}_i^T - \boldsymbol{I}_{3\times 3} = \boldsymbol{0}_{3\times 3} \tag{5}$$

したがって,式5から大きく外れる画像ペアは信頼性が低いと判断し,削除する.

復元した絶対的な外部パラメータを用い,撮影シーンの 世界座標点群を復元する.*Track*により対応付いた球面座 標点と絶対的な外部パラメータを用い,三角測量により各 世界座標点を導出する.このとき,復元の精度を高めるた め,図6のように世界座標点と対応する各球面座標点の成 す角度が一定以下の点のみを保持する.また,世界座標点 から各画像の絶対的なカメラ位置方向へのベクトルを計算 し,それらのベクトルのなす角の最大が一定の角度以下な らば,その世界座標点は無限遠点とみなし削除する.実験 では,これらの角度はそれぞれ5°に設定した.

各画像の内部パラメータと,求めた絶対的な外部パラメー タおよび撮影シーンの世界座標点群を,Bundle Adjustment (BA)により最適化する.本研究では,複数のカメラモデ ルに対応するため次のような評価関数を最小化する.





図 7: データセットの撮影環境

$$Error = \sum_{i=1}^{n} \sum_{j=1}^{m} v_i w_{ij} || \mathbf{x}_{ij} - P_{pers}(f_i, cx_i, cy_i, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_{wj}) ||$$
$$+ \sum_{i=1}^{n} \sum_{j=1}^{m} v_i w_{ij} || \mathbf{x}_{ij} - P_{equi}(f_i, cx_i, cy_i, \mathbf{R}_i, \mathbf{t}_i, \mathbf{X}_{wj}) ||$$
$$+ \cdots$$

(6)

ここで、n,mはそれぞれ画像枚数,世界座標点数, x_{ij} はi番目の画像のj番目の特徴点の座標である.また、 f_i, cx_i, cy_i はi番目のカメラの内部パラメータ、 R_i, t_i は絶対的な外部パラメータである. w_{ij} は、i番目の画像にj番目の世界座標点が写っていれば1、写っていなければ0になる. P_{pers} は透視投影カメラの投影式、 P_{equi} は等立体角射影カメラの投影式で、扱いたいカメラモデルの数だけ項が存在する.また、i番目の画像のカメラモデルと対応した項の v_i は1となり、それ以外の項の v_i は0となる.

4. 評価実験

提案手法の有効性を示すため、一定間隔毎に撮影した透 視投影画像,一定間隔毎に撮影した等立体角射影画像,一 定間隔毎に交互に撮影した透視投影画像と等立体角射影画 像の3種類の入力画像に対して提案手法を適用した.デー タセットには,透視投影画像と等立体角射影画像それぞれ において,カメラの姿勢を固定し,基準となる地点から直 線上を1 cm 毎に移動させながら撮影した 81 枚の画像を 用いた. データセットを作成した環境を図7に,透視投影 画像と等立体角射影画像の一部を図8に示す.このデータ セットは、絶対的なカメラ位置が既知で、すべての画像の 絶対的な回転行列は等しい.実験では、これらの既知の値 を用い3次元復元の復元精度を評価する.なお,直線移動 にはシグマ光機株式会社の光学ステージ, SGSP46-800(X) を用いた. このステージは位置決め精度 150 µm, 繰り返 し位置決め精度 10 µm, ロストモーション 20 µm の精度 で80 cmの直線上の任意の位置にカメラを移動させるこ とができる.また、カメラにはパナソニック株式会社の DMC-GH3 を用い、レンズには透視投影レンズ LUMIX G X VARIO 12-35mm/F2.8 と等立体角射影レンズ LUMIX G FISHEYE 8mm/F3.5 を用いた.



(d) 1 フレーム目
(e) 40 フレーム目
(f) 81 フレーム目
図 8:入力画像. (a)(b)(c) は透視投影画像, (d)(e)(f) は等
立体角射影画像

4.1 撮影シーン全体の3次元復元精度

2 cm 間隔で撮影した復元結果を図 9 に示す. この結果 より,3 種類の入力画像すべてで3次元復元が成功したこ とが確認できる.また,透視投影レンズと比較して等立体 角射影レンズは広い画角を持つことから,等立体角射影画 像を含んだ画像群は透視投影画像のみの画像群と比較し広 い範囲を復元できていることも確認できる.

3種類の入力画像の復元精度を定量的に知るために、BA による最適化前,最適化後の絶対的な外部パラメータ復元 精度を,撮影間隔を 2 cm~20 cm まで 1 cm 刻みで変えな がら評価した.精度の評価には、基準となる外部パラメー タと推定した外部パラメータがどの程度近いかを評価する, Relative Pose Error (RPE) O Root Mean Squared Error (RMSE)を用いた [22]. 基準となる外部パラメータには, 推定した絶対的な回転行列を平均した回転行列と、推定し たカメラ位置を実スケールに合わせて直線フィッティング し,その直線上の一定間隔毎の点から求めたカメラ位置を 使用した. なお,回転行列の平均には Geodesic L₂-Mean を用いた [21]. 絶対的な外部パラメータの復元精度の測定 結果を図 10 に示す. なお,最適化前,最適化後の双方で 値の入っていない部分は、復元精度の低い相対的な外部パ ラメータを削除する段階で、絶対的なカメラ位置を導出に 使用する画像が3枚未満になったために3次元復元に失敗 したことを示す.

実験結果より,最適化前の3次元復元精度では透視投影 画像のみを用いた場合が最も優れているといえる.一方 で,最適化後の復元精度に着目すると,多くの場合で3種 類の入力画像すべてで同程度の精度で復元されていること を確認できる.したがって,3種類の画像群の最適化前の 復元精度の差は,BAに影響を与えない程度であると考え られる.また,撮影間隔が4,5 cmの透視投影画像を入力 としてとき,最適化によって3次元復元精度が低下する現 象が確認された.これは,3.3章で述べた手法では特徴点 の対応付けにおける誤対応を除去しきれなかったことが原

情報処理学会研究報告

IPSJ SIG Technical Report



(a) 2 cm 間隔毎に撮影した透視投影画像を用いた復元結
 果



(b) 2 cm 間隔毎に撮影した等立体角射影画像を用いた復元結果



(c) 2 cm 間隔毎に交互に撮影した透視投影画像と等立体 角射影画像を用いた復元結果



図 9: 復元結果

図 10: 絶対的な外部パラメータの復元精度

因と考えられる.この場合,*Track*に誤った特徴点追跡が 保存され,最適化のときにその追跡結果の再投影誤差まで 最小化しようとするため,全体の復元精度の低下が生じた と考えられる.

4.2 2 画像間の 3 次元復元精度

絶対的な外部パラメータの復元精度は相対的な外部パラ メータの復元精度に大きく依存する.そこで,相対的な外 部パラメータの復元精度が絶対的な外部パラメータの復元



図 11: 相対的な外部パラメータの復元精度

精度に与える影響について考えるため、カメラモデルによる相対的な外部パラメータの復元精度を評価した.一定間隔毎に撮影した2枚の画像間の相対的な外部パラメータの 復元精度を、RPEのRMSEにより評価する.評価には2 cm間隔で撮影した入力画像の3.3章で導出した相対的外 部パラメータを用い、撮影間隔が2 cm~80 cmのときの それぞれ距離に対応した画像ペアを評価した.基準となる 相対的な外部パラメータには、前述した方法で求めた基準 の絶対的な外部パラメータより計算した.なお、相対的な カメラ位置は撮影間隔に関わらず1に正規化した.相対的 な外部パラメータの3次元復元精度を図11 に示す.

2画像間の撮影間隔が小さいときは、透視投影画像の復元 結果が最も優れていることが確認できる.一方で、撮影間 隔が大きくなると2画像で共通して見える点が極端に少な くなるため、46 cm を境として復元精度が極端に低くなっ ている.等立体角射影画像は撮影間隔が小さいときの透視 投影画像の復元精度には及ばないものの、撮影間隔が大き くなっても復元精度が安定していることが確認できる.こ れは、等立体角射影レンズの画角が透視投影レンズよりも 大きく,撮影間隔が大きくなっても2画像で共通したシー ンが見えるためであると考えられる.撮影間隔が小さいと きにおいて等立体角射影画像の復元精度が透視投影画像よ りも劣っている原因としては、等立体角射影画像は画像の 中心から離れるほど空間解像度が低くなるのに対し、提案 手法では画像の中央に検出された特徴点も端に検出された 特徴点も等しく扱っているためだと考えられる. さらに, 等立体角射影画像は画角が大きな分、透視投影画像と比較 して画像全体の空間解像度が低いことも原因として考えら れる.また、カメラモデルを混在させた入力画像の相対的 な外部パラメータは、他の2つの入力画像と比べて安定し て求められる区間が短いことが確認された. 絶対的な外部 パラメータの評価時に、撮影間隔が大きくなるとカメラモ デルを混在させた入力画像の3次元復元が失敗したのは, これが原因であると考えられる.

5. まとめ

本論文では、カメラモデルの異なる複数のカメラで撮影

した画像を同時に扱い3次元復元を行う,SfMのフレー ムワークを提案した.提案手法は,球面座標系を導入する ことでカメラモデルに依存しない3次元復元を行うととも に内部パラメータの最適化も同時に行う,汎用性の高いフ レームワークである.提案手法では,画像と各画像のカメ ラモデル,内部パラメータを入力として,特徴点マッチン グ,2画像間の3次元復元,全体の3次元復元のすべてを 自動で行う.実験では,透視投影画像と等立体角射影画像 と双方の画像を混ぜた3種類のデータに対して提案手法を 適用し,すべてのデータにおいて3次元復元が成功するこ とを確認した.

今後の課題と展開としては、各カメラモデルの特徴を、 提案フレームワークに反映させるように拡張することが考 えられる.提案手法では、画像から検出したすべての特徴 点を等しく扱うが、実際には等立体角射影画像のように、 画像の場所によって空間解像度が変わるカメラモデルも存 在する.このようなカメラモデル毎の特徴を考慮して提案 フレームワークに反映させることで、復元精度の向上が期 待できる.

参考文献

- Michael Goesele, Noah Snavely, Brian Curless, Hugues Hoppe, and Steven M Seitz. Multi-view stereo for community photo collections. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8. IEEE, 2007.
- [2] Yasutaka Furukawa, Brian Curless, Steven M Seitz, and Richard Szeliski. Towards internet-scale multi-view stereo. In *Computer Vision and Pattern Recognition* (CVPR), 2010 IEEE Conference on, pp. 1434–1441. IEEE, 2010.
- [3] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building rome in a day. *Communications of* the ACM, Vol. 54, No. 10, pp. 105–112, 2011.
- [4] Arnold Irschara, Christopher Zach, Jan-Michael Frahm, and Horst Bischof. From structure-from-motion point clouds to fast location recognition. In *Computer Vi*sion and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp. 2599–2606. IEEE, 2009.
- [5] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, Vol. 80, No. 2, pp. 189–210, 2008.
- [6] Jan-Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, et al. Building rome on a cloudless day. In European Conference on Computer Vision, pp. 368–381. Springer, 2010.
- [7] Jean-Michel Morel and Guoshen Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, Vol. 2, No. 2, pp. 438–469, 2009.
- [8] Christian Beder and Richard Steffen. Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence. In *Joint Pattern Recognition*

Symposium, pp. 657–666. Springer, 2006.

- [9] Venu Madhav Govindu. Combining two-view constraints for motion estimation. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, Vol. 2, pp. II– 218. IEEE, 2001.
- [10] Venu Madhav Govindu. Lie-algebraic averaging for globally consistent motion estimation. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, Vol. 1, pp. I–I. IEEE, 2004.
- [11] David Crandall, Andrew Owens, Noah Snavely, and Dan Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *Computer Vision* and Pattern Recognition (CVPR), 2011 IEEE Conference on, pp. 3001–3008. IEEE, 2011.
- [12] Davide Scaramuzza, Friedrich Fraundorfer, Marc Pollefeys, and Roland Siegwart. Closing the loop in appearance-guided structure-from-motion for omnidirectional cameras. In *The 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras-OMNIVIS*, 2008.
- [13] Branislav Micusik and Tomas Pajdla. Structure from motion with wide circular field of view cameras. *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, Vol. 28, No. 7, pp. 1135–1149, 2006.
- [14] Yalin Bastanlar, Alptekin Temizel, Yasemin Yardimci, and Peter Sturm. Effective structure-from-motion for hybrid camera systems. In *Pattern Recognition (ICPR)*, 2010 20th International Conference on, pp. 1654–1657. IEEE, 2010.
- [15] Alain Pagani and Didier Stricker. Structure from motion using full spherical panoramic cameras. In Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on, pp. 375–382. IEEE, 2011.
- [16] Nianjuan Jiang, Zhaopeng Cui, and Ping Tan. A global linear method for camera pose registration. In Proceedings of the IEEE International Conference on Computer Vision, pp. 481–488, 2013.
- [17] Marius Muja and David G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. VISAPP (1), Vol. 2, No. 331-340, p. 2, 2009.
- [18] Martin A Fischler and Robert 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.
- [19] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2000.
- [20] Daniel Martinec and Tomas Pajdla. Robust rotation and translation estimation in multiview reconstruction. In 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, 2007.
- [21] Richard Hartley, Jochen Trumpf, Yuchao Dai, and Hongdong Li. Rotation averaging. *International journal of computer vision*, Vol. 103, No. 3, pp. 267–305, 2013.
- [22] Jürgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard, and Daniel Cremers. A benchmark for the evaluation of rgb-d slam systems. In *Intelligent Robots and* Systems (IROS), 2012 IEEE/RSJ International Conference on, pp. 573–580. IEEE, 2012.