3 次元形状特徴の位置合わせに基づくカメラの相対姿勢推定

Relative Camera Pose Estimation based on 3D Shape Feature Alignment

松﨑 康平†‡　河村 圭‡　川西 康友†‡　村瀬 洋†‡
Kohei MATSUZAKI†‡ Kei KAWAMURA‡ Yasutomo KAWANISHI†‡ and Hiroshi MURASE†‡

†名古屋大学　‡株式会社 KDDI 総合研究所
†Nagoya University　‡KDDI Research, Inc.

Abstract　In this paper, we proposed a relative camera pose estimation method based on 3D shape feature alignment. It extracts 3D shape features individually from two images and measuring their similarity. Experiments showed that the proposed method improved the relative pose estimation accuracy by 13.8% and more compared with conventional methods.

1. はじめに

カメラ間の相対姿勢は，2 つの画像間での物体の見えの重複が少ない場合には推定が困難である．この問題に対処するために，本稿では画像から物体の 3 次元形状を表す特徴を生成し，それらの位置合わせに基づいてカメラの相対姿勢を推定する手法を提案する．本稿の実験では，提案手法は従来手法と比べて相対姿勢推定精度を平均で 13.8%以上改善することを示した．

2. 関連研究

カメラ間の相対姿勢推定に対しては，2 枚の画像から抽出される局所特徴を対応付け，RANSAC アルゴリズムを用いて相対姿勢を推定する方法[1]が広く用いられている．しかし，この方法では画像間で物体の見えの重複が少ない場合には局所特徴を十分に対応付けることができず，推定が困難となる．

このような場合でも推定が可能な方法として，深層学習に基づいて画像から回帰によって相対姿勢を求める方法がある．En ら[2]は RPNet と呼ばれるモデルを用いて，2 枚の画像から個別に抽出した特徴を全結合層へ入力し，回帰によって相対姿勢を推定する手法を提案した．これにより，見えの重複が少ない場合であっても高精度に相対姿勢を推定できる．

また，画像から仮想カメラの姿勢を反映した 3 次元特徴を生成し，それらを用いて相対姿勢を推定する方法もある．Banani ら[3]は，NOVE (Novel Object Viewpoint Estimation) と呼ばれる手法を提案した．この手法は，画像から抽出した特徴を仮想カメラを用いて 3 次元格子空間へ投影することによって，3 次元特徴を生成する．そして，2 枚の画像から生成された 3 次元特徴を識別器に入力することによって，仮想カメラの相対姿勢の誤差の大きさを予測する．相対姿勢を推定する際には，様々な相対姿勢を持つ仮想カメラを用いて 3 次元特徴を生成し，識別器が予測する誤差が最小となる相対姿勢を探索する．

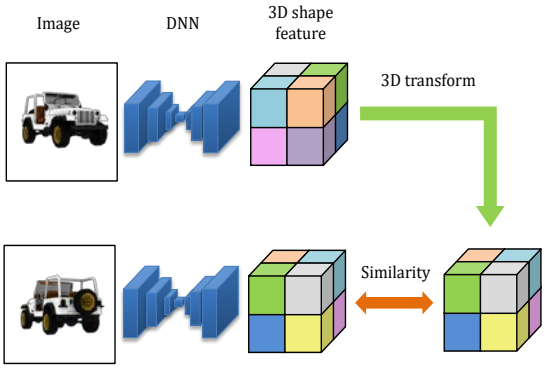


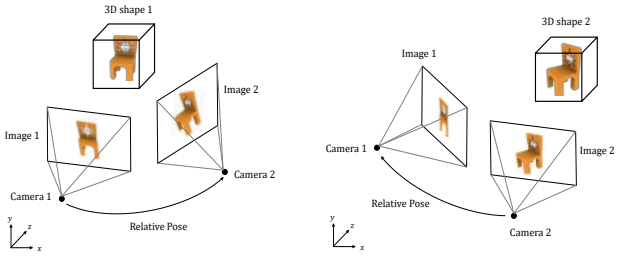
図 1　提案手法の概要

3. 提案手法

2 章で概説した NOVE は，3 次元特徴への仮想カメラの姿勢の反映と識別器による予測の両方に基づいて相対姿勢を推定する．そのため，どちらか一方に誤りが生じた場合には，相対姿勢の推定精度が低下するという問題がある．

この問題に対処するために，識別器を用いずに直接的に 3 次元特徴を比較することによって，相対姿勢を推定する手法を提案する．図 1 に，提案手法の概要を示す．本手法では，画像から DNN (Deep Neural Network) を用いて 3 次元形状を表す特徴を生成し，それらに 3 次元変換を施すことによって位置合わせを実現する．この位置合わせでは，一方の 3 次元形状特徴に様々な姿勢変換を施し，もう一方の 3 次元形状特徴との間で類似度を計算する．そして，類似度が最大となる姿勢変換を探索する．

提案手法における DNN の概要を述べる．初めに，2 枚の画像のそれぞれから畳み込みニューラルネットワークを用いて特徴を抽出する．そして，仮想カメラを用いて特徴を 3 次元格子空間へ投影し，3 次元特徴を得る．この時，各画像に対して独立した 3 次元特徴を生成するために，図 2 に示すように個別の座標系で投



(a)　　(b)

図 2　個別の座標系の模式図

影を行う．図 2(a)は 1 つ目の画像から抽出された特徴を，3 次元格子の正面に設置された仮想カメラを用いて投影する様子を模式的に表している．ここでは 2 つ目の画像を投影する仮想カメラは，1 つ目の仮想カメラに対する相対的な姿勢をとる．図 2(b)では 1 つ目の画像と 2 つ目の画像の関係性が入れ替わっており，2 つ目の画像から抽出された特徴を，3 次元格子の正面に設置された仮想カメラを用いて投影する．その後，3D UNet を用いてこれらの 3 次元特徴を洗練する．最後に，洗練後の 3D 特徴を用いて，デプス画像および 3 次元形状を表す特徴を予測する．

モデルを学習する際には，Multi-view supervision の考えに基づいて複数のデプス画像を用いて損失を計算する．初めに，予測されたデプス画像とその Ground-truth との間で L1 損失を計算する．また，3 次元形状を表す特徴を用いて文献[5]に記載の Ray Consistency 損失を計算する．また，正しい相対姿勢で類似度が最大となるように，相対姿勢の Ground-truth を用いて一方の 3 次元形状を表す特徴を変換し，もう一方の特徴との間で L2 損失を計算する．そして，これらの損失の和を用いてモデルを学習する．

4. 評価実験

様々な物体カテゴリにわたる大規模な 3D モデルを含む ShapeNet データセット[4]を用いて，提案手法の有効性を評価する．実験のために，単位球面上に位置する仮想視点を用いて 3D モデルから画像をレンダリングする．ここでは，方位角を-180°から 180°の範囲で，仰角を-90°から 90°の範囲でランダムに選択することによって，1 つの 3D モデルにつき 20 通りの仮想視点を設定する．そして，レンダリングされた画像および仮想視点を組み合わせることによって，画像のペアと相対姿勢の Ground-truth を作成する．また，ここでは各画像に対してランダムな背景を追加する．本実験ではこれらのデータを用いて，モデルの学習および相対姿勢推定精度の評価を行う．相対姿勢推定精度の指標として，角度誤差が 30°以下となる割合を表す $Acc_{\pi/6}$ [5]を用いる．相対姿勢推定に関する従来手法として RPNet と NOVE を用いる．

表 1　カメラの相対姿勢推定精度[%]の比較

Category	RPNet	NOVE	提案手法
bench	34.1	29.2	**50.2**
cabinet	40.5	41.1	**51.2**
car	41.9	**52.9**	51.7
chair	44.8	47.3	**51.5**
couch	**51.2**	46.1	51.0
display	34.2	42.0	**50.0**
lamp	11.1	13.4	**51.3**
phone	39.2	31.5	**51.5**
plane	39.8	50.8	**51.4**
rifle	43.9	35.1	**52.1**
speaker	10.6	29.2	**50.9**
table	12.8	34.9	**51.4**
vessel	33.9	33.9	**52.0**
average	33.6	37.5	**51.3**

表 1 に，従来手法と提案手法によるカメラの相対姿勢推定精度を示す．この表では，13 通りの物体カテゴリにおける手法毎の $Acc_{\pi/6}$ とそれらの平均値を示している．平均値を比較すると，提案手法が従来手法と比べて推定精度を 13.8%以上改善することがわかる．これは，提案手法が 3 次元物体形状を表す特徴の類似度を直接的に比較することの効果である．カテゴリ毎の精度に着目すると，従来手法ではカテゴリによって精度の差が大きいのに対して，提案手法は全てのカテゴリで高い精度を達成できることがわかる．

5. むすび

本稿では，2 枚の画像から生成した 3 次元形状を表す特徴間の類似度を直接的に比較することによって，カメラの相対姿勢を推定する手法を提案した．公開データセットを用いた評価実験では，従来手法との比較によって提案手法の有効性が確認された．

文　献

[1] R. Hartley and A. Zisserman : “Multiple View Geometry in Computer Vision,” Cambridge University Press, 2003.

[2] S. En, A. Lechervy, and F. Jurie. : "Rpnet: An end-to-end network for relative camera pose estimation", In Proc of ECCVW, 2018.

[3] M. E. Banani, J. J. Corso, and D. F. Fouhey : "Novel Object Viewpoint Estimation through Reconstruction Alignment", In Proc of CVPR, pp. 3113-3122, 2020.

[4] A. X. Chang, et al. : "Shapenet: An information-rich 3d model repository", arXiv preprint arXiv:1512.03012, 2015.

[5] S. Tulsiani, A. A. Efros, and J. Malik : ” Multi-view Consistency as Supervisory Signal for Learning Shape and Pose Prediction” , In Proc of CVPR, pp. 2897-2905, 2018.

†名古屋大学大学院情報学研究科
〒464-8601 愛知県名古屋市千種区不老町
‡株式会社KDDI総合研究所超臨場感通信グループ
〒356-8502 埼玉県ふじみ野市大原2丁目1番15号
E-mail: ko-matsuzaki@kddi-research.jp