6N-02

カメラ位置姿勢推定のためのキーポイント特徴量データベース照合の深層学習による高精度化

中島 由勝† 斎藤英雄†

慶應義塾大学†

1. はじめに

入力画像に対するカメラ位置姿勢推定は拡張現実感への応用などにおいて，仮想情報を重畳する際の幾何学的整合性を得る上で重要な技術であり，盛んに研究が行われている．

従来手法である SIFT[1]は回転及び拡大縮小に対する堅牢性は保持しているが，入力画像が対象である平面パターンに対し大きな射影的歪みを受けたような画像となると正確にカメラ位置姿勢の推定が行うことができない．その問題点を解決すべく，従来手法である Viewpoint Generative Learning[2](VGL)では対象である平面パターンに対しあらかじめ様々な視点から撮影されたかのような画像群を生成し，それらの画像群から検出された特徴点に関する特徴量を K-means 法により圧縮した後データベースすることでカメラ位置姿勢推定の精度向上を図った．しかし VGL では特徴量を圧縮するため，特徴量が大きく変化するような対象平面に対し浅い角度から撮影された入力画像には対応できないという問題点が存在する．本研究は，上記のような入力画像が浅い角度となった際に生じる問題点を解決することを目的とした．

2. 提案手法

本研究では，あらかじめ浅い角度も含む複数の視点ごとに平面パターンの特徴量データベースを非圧縮で用意し，入力画像に対してその画像がおおよそどの角度から撮影されたかを判定できるように学習させた Convolutional Neural Network(CNN)により，データベース群から入力画像に近いカメラ位置姿勢に関する特徴量データベースを一つ選択しそのデータベース内で特徴量の最近某探索によりマッチングを行う手法を提案，検証する．

2.1 データベース群の生成

はじめに，固定された平面パターンに対し複数の視点から画像を撮影し画像群を得る．次にそれらの画像に対し，正面から見た際の座標が

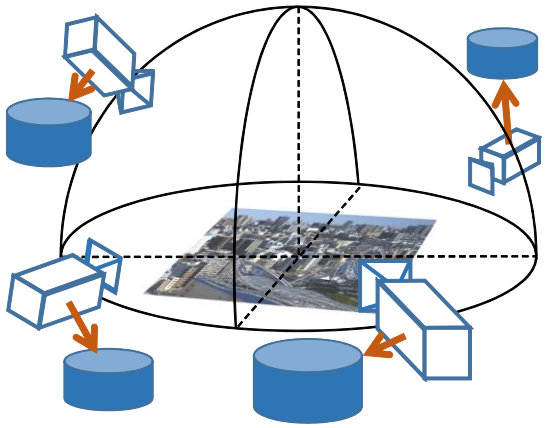

図 1 データベース群生成の概念図

明確にわかる 4 点を与え，

$$\tilde{x}'\sim H\tilde{x}$$

によりその画像中の平面パターンを正面画像のように変換する射影変換を表す行列Hをそれぞれ得る．ここで，$\tilde{x}'\sim(x',y',1)^{\mathrm{T}}$，$\tilde{x}\sim(x,y,1)^{\mathrm{T}}$であり，$\tilde{x}'$は正面画像における座標，$\tilde{x}$は各視点から撮影された画像における座標である．次に各画像について適切な局所特徴アルゴリズムにより特徴点を検出する．検出された特徴点の座標を正面画像の座標に行列Hにより変換し，記述された特徴量とひも付け，各画像ごとに非圧縮のデータベースを生成する(図 1 参照)．

2.2 視点に関する深層学習

入力画像に対しどのデータベースが最適かを推定するため，CNN を対象平面に対する大まかな視点分類に用いる．データベース群の生成時に利用した各画像について，それぞれの画像を撮影したカメラ位置から近いカメラ位置で対象平面を複数回撮影し，CNN の学習用画像群を得る．最後に各視点を教師として CNN の深層学習を行う．

2.3 カメラ位置姿勢推定

入力画像に対するカメラ位置姿勢推定ではまず，局所特徴アルゴリズムにより特徴点及び特徴量を検出する．次に CNN により入力画像が撮影された大まかな視点を推定し，その視点に関するデータベースを選択する．最後に，選択されたデータベース内で再近傍探索を行い特徴点のマッチングを行い，カメラ位置姿勢を推定する．

Accuracy Improvement of Matching with Keypoint Database by Deep Learning for Camera Pose Estimation
Yoshikatsu Nakajima† and Hideo Saito†
†Keio University

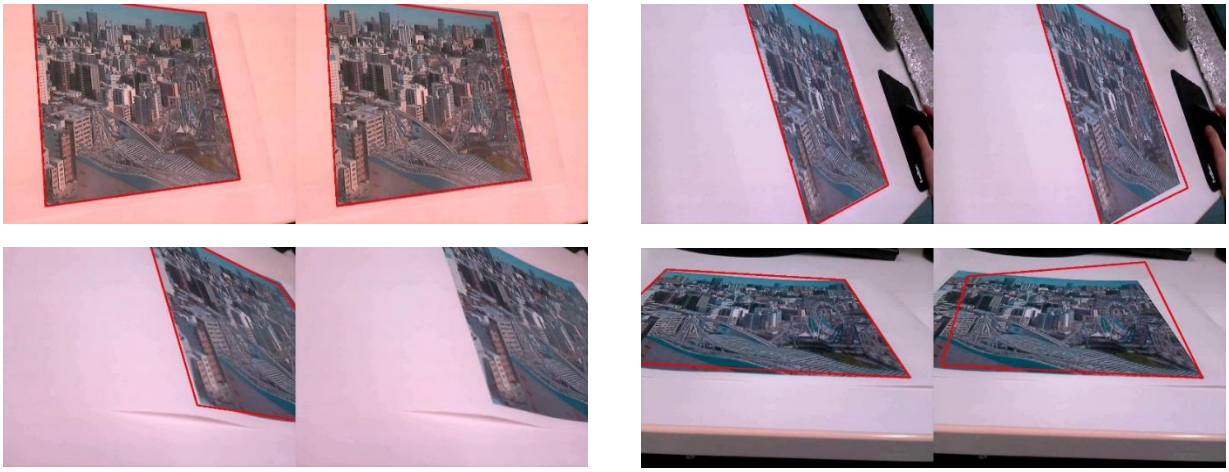

図 2 カメラ位置姿勢の推定例（左:提案手法，右:VGL）

3. 評価実験

本章では提案手法を評価するために行った実験について示す．対象である平面パターンに対しカメラ位置姿勢が浅い角度となるようなシーンを後半部に含む動画を用意し，各フレームに対するカメラ位置姿勢推定に用いたマッチング数，フレームレートを VGL と比較して評価した．本実験では平面パターンに対しまんべんなく複数の視点から撮影した 22 枚の画像群によりデータベースを生成した．また，SIFT を局所特徴アルゴリズムとして用いた．さらに，CNN の構成には Min らの Network In Network[3](NIN)を用いた．図 2 にカメラ位置姿勢を推定した結果の一部を示す．赤い枠線が示されていないものは射影変換行列の算出に足るマッチング数が得られなかったことを示す．図 2 を見ると，平面パターンに対するカメラ位置姿勢が浅い角度となった入力画像に対しても本提案手法では正確にカメラ位置姿勢を推定できたことがわかる．また，図 3 に平面射影変換行列の算出に用いたマッチング数を示す．図 2 及び図 3 を見ると，VGL は特徴量を圧縮するため各特徴点の特徴量が大きく変動した場合にマッチングを正確に行えずカメラ位置姿勢推定の精度が低下したことがわかる．一方で提案手法では適切な非圧縮のデータベースが適宜選択されたためカメラ位置姿勢推定を高精度に行うことができた．また，フレームレートの平均値は提案手法が 3.36 枚/秒，VGL が 3.58 枚/秒となり，CNN の視点推定によるオーバーヘッドが十分に小さいことを確認した．

4. 結論

本稿では，深層学習により入力画像のカメラ位置姿勢に対応するデータベースを選択し，そのデータベース内でマッチングを行うことでカメラ位置姿勢を推定する手法を提案，評価した．

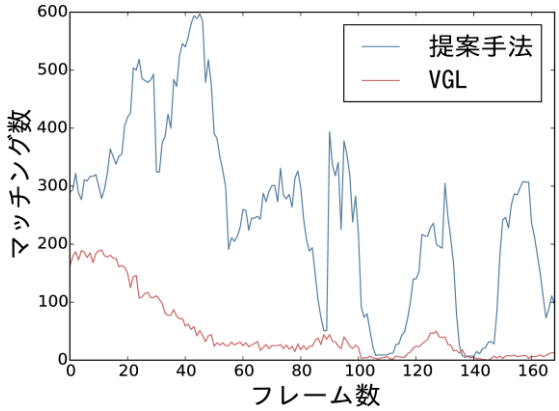

図 3 マッチング数の比較

本提案手法により，従来手法に比べマッチング数が増加することでカメラ位置姿勢推定の精度が向上した．処理時間についても即応性が十分であることを確認した．今後の課題として，3 次元物体，生成型学習への応用等が挙げられる．

謝辞

本研究の一部は，科学研究費　基盤研究(S) 24220004 の補助により行われた．

参考文献

[1] David G. Lowe. “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, 60, 2, pp 91-110, 2004.

[2] 吉田拓洋，斎藤英雄，清水雅芳，田口哲典. “視点生成型学習による頑健な平面位置姿勢推定”日本バーチャルリアリティ学会論文誌，Vol.17，No.3 pp 191-200，2012

[3] Lin Min, Chen Qiang and Yan Shuicheng. “Network In Network” in International Conference on Learning Representations, 2014