

変形 AR マーカの高速かつ高精度な姿勢推定

浅野 右京 (指導教員: 山内 悠嗣)
中部大学 工学部

1. はじめに

2次元コードが急速に普及し、キャッシュレス決済やロボットの自己位置推定など様々な用途で利用されている。しかし、2次元コードが変形すると、認識や位置・姿勢推定が難化する。深層学習により変形した2次元コードの3次元位置・姿勢を推定する手法 [1] が提案されたが、姿勢推定精度と計算時間に課題があった。

そこで、本研究では変形を除去するモデルと姿勢を回帰推定するモデルを End-to-End で学習した手法を提案する。変形を除去するタスクと同時に姿勢を推定することで、姿勢推定精度の向上が期待できる。また、従来手法は最近傍法による姿勢推定であるが、提案手法は全結合層にて回帰推定を行うため計算量を削減できる。

2. 提案手法による姿勢推定

従来手法は、物体検出器により画像中の変形 AR マーカを検出し、Augmented Autoencoder(AAE)[2]により背景、変形を除去した AR マーカの潜在変数を取得する。そして、あらかじめ作成したデータベースと潜在変数を照合することで変形 AR マーカの3次元位置・姿勢を推定する。

従来手法において潜在変数を用いたデータベースと照合することで姿勢を推定していた箇所を提案手法では回帰推定に置き換えることで姿勢を推定する。提案手法の学習の流れを図1に示す。提案手法は、AAEで事前学習済みのエンコーダ、デコーダを使用し学習する。はじめに、変形 AR マーカ画像を AAE のエンコーダに入力して潜在変数を得る。潜在変数をデコーダと MLP に入力し、デコーダでは復元画像を出力、MLP では推定した姿勢を出力する。そして、式(1)より損失 L_{AAE} を計算し、AAE のエンコーダの重みを更新する。

$$L_{AAE} = L_{rc} + \lambda L_{pose} \quad (1)$$

$$L_{rc} = \frac{1}{2} \sum_{i=1}^n (\hat{x}_i - x_i)^2 \quad (2)$$

$$L_{pose} = \frac{|\hat{r}_1 - r_1| + |\hat{r}_2 - r_2|}{2} + |\hat{p} - p| + |\hat{y} - y| \quad (3)$$

ここで L_{rc} は復元画像 \hat{x} と教師画像 x の画像誤差、 L_{pose} は推定姿勢と正解姿勢の姿勢誤差、 λ は画像誤差と姿勢誤差のバランスをとる係数である。 (\hat{r}_1, \hat{r}_2) は Roll の推定姿勢、 \hat{p} は Pitch の推定姿勢、 \hat{y} は Yaw の推定姿勢、 (r_1, r_2) は Roll の正解姿勢、 p は Pitch の正解姿勢、 y は Yaw の正解姿勢である。なお、Roll の姿勢範囲は 0 から 359[deg] のため、角度の周期性を考慮する必要がある。そのため、Roll の角度 ϕ は $(\cos \phi, \sin \phi)$ の2次元で表現することで対応する。

AAE を更新した後、AAE の重みを固定して姿勢を推定する MLP を更新する。交互に最適化する処理を繰り返すことで学習を進める。

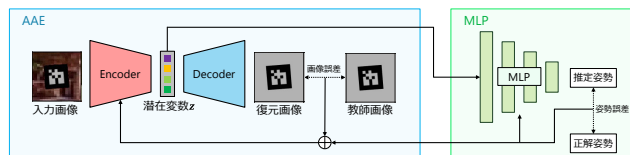


図1 AAE と回帰推定の End-to-End 学習の流れ

3. 評価実験

提案手法の有効性を確認するために、AlexNet[3]をベースとしたCNNによる姿勢推定法と従来手法 [1]、提案手法の姿勢推定精度と姿勢推定の計算時間を比較する。評価には、シミュレータ上で撮影された990枚の変形 AR マーカ画像を用いる。実験に使用したPCのスペックはCPU Intel Core i7-12700、メモリ 16GB、GPU GeForce RTX 3080 である。

表1にRoll, Pitch, Yawの姿勢推定誤差の結果を示す。提案手法は従来手法のRollの精度を同等に保ちつつ、PitchとYawにおいてAlexNetより精度が上昇した。提案手法の平均誤差は2.13[deg]であり、AlexNetの3.94[deg]からは45.9[%]、従来手法の5.28[deg]からは60.0[%]減少した。

表1 各手法の姿勢推定誤差 [deg].

手法	Roll	Pitch	Yaw	平均
AlexNet[3]	1.48	5.01	5.34	3.94
従来手法 [1]	0.69	7.84	7.32	5.28
提案手法	0.64	2.87	2.87	2.13

表2に画像1枚あたりの姿勢推定時間の結果を示す。姿勢推定の時間は提案手法で6.233[msec]となり、AlexNetの2.395[msec]と比較すると計算時間が若干増加したが、従来手法の3,270[msec]と比較すると計算時間を99.8[%]削減できた。

表2 各手法の姿勢推定時間 [msec].

手法	1枚あたりの推定時間
AlexNet[3]	2.395
従来手法 [1]	3,270
提案手法	6.233

4. まとめ

本研究では、従来手法の姿勢推定に回帰推定の導入し、AAEと回帰推定をEnd-to-Endで学習することを提案した。今後は、姿勢推定に用いる潜在変数を外観と変形に分ける手法を導入し、精度を向上させる予定である。

参考文献

- [1] 榎元等: “機械学習による変形 AR マーカの3次元位置・姿勢推定”, 動的画像処理実用化ワークショップ, 2022.
- [2] M.Sundermeyer *et al.*: “Implicit 3D Orientation Learning for 6D Object Detection from RGB Images”, ECCV, 2018.
- [3] A.Krizhevsky *et al.*: “ImageNet Classification with Deep Convolutional Neural Networks”, NeurIPS, 2012.