EPiC Series in Health Sciences

Volume 7, 2024, Pages 145–149

Proceedings of The 24th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery

Markerless 6D Pose Estimation of Surgical Tools and Anatomical Shapes from RGB-D images: a comparison of two approaches based on synthetic data

Basile Longo, Salaheddine Sta, Éric Stindel and Guillaume Dardenne

Laboratory of Medical Information Processing (LaTIM), UMR1101 INSERM, Brest, France

Abstract

In the realm of Computer Assisted Surgery (CAS), the accurate localization of anatomical features and surgical tools is crucial for enhancing surgical outcomes. Important efforts have been recently focused toward markerless navigation to reduce intraoperative intrusivity. Unfortunately, recent studies don't always satisfy the required accuracy and precision to be used clinically. This research investigates the application of deep learning models for pose estimation of synthetic anatomical features and surgery instruments to improve the localization accuracy. Based on the models from CosyPose and Coupled-Iterative-Refinement methods, we applied a three-step pose estimation process. Using 3D meshes and BlenderProc we generated a synthetic RGB-D dataset of scenes including tibias, femurs, shoulder's glenoids and surgical instruments (Cutting guide) to train and test our model. Moreover, we compared our results to the Point Pair Features (PPF) method, a conventional pose estimation algorithm based on point cloud data. We found a significant enhancement regarding the accuracy with our model, achieving sub-millimetric and sub-degree accuracy, surpassing the PPF algorithm. We also provided qualitative examples of the estimated poses to visualize the accuracy of our model. While the proposed method shows promising results, challenges remain in particular passing from synthetic to real-world data. Future efforts will focus on collecting annotated real-world data.

1 Introduction

Accurate localization of anatomical features and surgical tools plays a vital role in Computer Assisted Orthopaedic Surgery. By providing assistance to surgeons, CAOS contributes to improve accuracy of surgeries [1]. However, navigation systems, widely used in CAS, often rely on markers which can disturb the surgical workflow as they may requires additional bone incisions, surgical time and are bulky [1][2]. These drawbacks led to preliminary works to develop navigation methods that limit the use of these markers.

Sta et al. presented a method for the pose estimation of surgical tools from a depth camera based on the Point Pair Features (PPF) algorithm [3]. Rodrigues et al. used a RGB camera to segment the bone surface of a femur with a U-Net model and a depth camera to register the bone with the point cloud from the segmented bone surface [2]. Their registration method was

based on an improved version of the 4PCS algorithm [4] and didn't require any extra step from the surgeon.

However, both studies didn't always reach the wanted clinical accuracy. Recent advances in pose estimation, in particular through the use of deep learning models, have shown significant improved performance over more traditional approaches [5]. Our objective was to assess their ability to accurately localize surgical instruments and anatomical features such as the glenoid of the shoulder or parts of the tibia and femur.

2 Methods

2.1 Model

We used the Coupled Iterative Refinement method [6], from the BOP Challenge, and built upon another model, CosyPose [7], the most successful pose estimation algorithm in the 2020 BOP Challenge. The pose estimation process consists of three main successive stages:

1. **Detection** of the object of interest (the femur, tibia, shoulder's glenoid or cutting guide) in RGB images using a Mask R-CNN model [8] which outputs a bounding box and a mask.

2. **Coarse pose estimation** using the RGB image cropped around the detected object of interest and an initial alignment computed with its bounding box, we compute an initial pose estimation with an EfficientNet model.

3. **Pose refinement**. From the previously estimated coarse pose, we add noise to generate a 7 other poses. We render images of the object of interest with the new poses and with the previously cropped RGB, mask & depth images, we feed a RAFT [9] inspired architecture model that will output a pose update. This is repeated in several loops to converge to a final pose estimation.

Here, we used a single model to estimate the pose of each object of interest.

2.2 Dataset generation

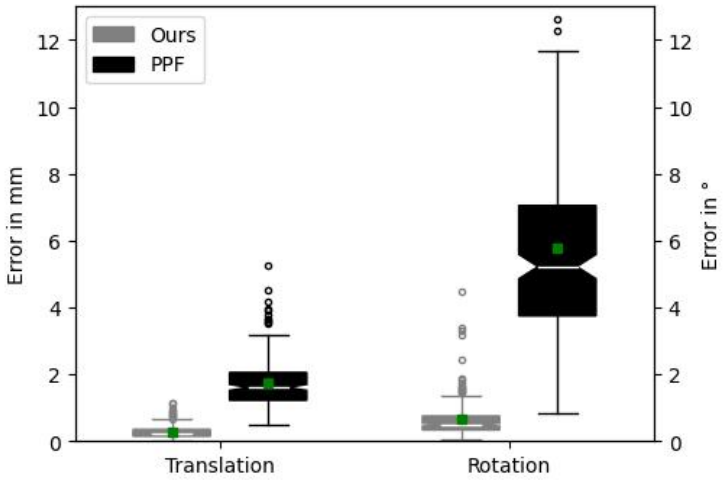
The model was trained and tested on synthetic data. 3d meshes of one tibia, one femur and one glenoid were obtained from CT scans and a 3d mesh of one cut guide was designed. Using BlenderProc [10], a photorealistic synthetic image generation tool, we created numerous pose-annotated scenes with customized parameters. Our dataset included scenes with the four objects of interest positioned at various locations and rotations, we generated 24.000 images for training the models.

We compared our model with the PPF algorithm's pose estimation on depth images cropped from the object of interest detected by the Mask R-CNN.

3 Results

As shown in Figure 1, our approach achieved submillimetric and sub-degree accuracy. The test set contained 209 images of the different objects of interest, comprising 53 images of a femur, 51 images of a tibia, 55 images of a cutting guide and 50 images of a glenoid. Estimated poses with an error $\geq$ 40 mm or $^\circ$ were considered as False. Seven images where considered as False

for our model, all were considering a symmetrical wrong pose of femur and glenoid, and four images for the PPF.



(a) Overall errors in translation (mm) and rotation (°) for our and PPF method

	Femur		Tibia		Cutting guide		Glenoid		Overall	
	Te	Re	Te	Re	Te	Re	Te	Re	Te	Re
Ours	**0.25**	**0.44**	**0.34**	**0.58**	**0.35**	**0.89**	**0.17**	**0.65**	**0.28**	**0.65**
PPF [11]	2.34	5.90	1.83	7.63	1.41	3.90	1.43	5.77	1.75	5.76

(b) For each object of interest : mean translation error Te (mm) and mean rotational error Re (°)

Figure 1: Quantative results of our pose estimation study

Figure 2 displays qualitative examples of our 6D pose estimation results upon the different objects of interest.

4 Discussion

We presented a comprehensive pipeline integrating synthetic data generation and markerless pose estimation of anatomical shapes and instruments. Moreover, we compared the results obtained with a more traditional 6D pose estimation method used by Sta et al. [3] and noted promising improvements with mean errors less than 1° for the rotation and 1 mm for the translation. The advantage of this method is that it doesn't require markers to work, is not expensive, and only requires a depth camera and an RGB camera. But such methods face challenges related to domain adaptation, as models trained on synthetic data may exhibit diminished performance when applied to real-world data. Further work should be done toward adaptability to real-world by collecting real annotated data.

Figure 2: RGB images cropped around object of interest and respective estimated pose overlaid.

5 Acknowledgements

This work benefited from a government grant managed by the Agence Nationale de la Recherche under the Programme d'Investissement pour le Futur (Investment Programme for the Future) with the reference ANR-17-RHUS-0005 (FollowKnee Project). Authors thank PLaTIMed for techincal contribution to the project.

References

[1] Werner Siebert et al. "Technique and first clinical results of robot-assisted total knee replacement". In: *The Knee* 9.3 (2002), pp. 173–180. ISSN: 0968-0160. DOI: https://doi.org/10.1016/S0968-0160(02)00015-7. URL: https://www.sciencedirect.com/science/article/pii/S0968016002000157.

[2] Pedro Rodrigues et al. "Deep segmentation leverages geometric pose estimation in computer-aided total knee arthroplasty". en. In: *Healthc. Technol. Lett.* 6.6 (Dec. 2019), pp. 226–230.

[3] Salaheddine Sta et al. *Towards markerless computer assisted surgery: Application to TKA*. June 2021. DOI: 10.1002/rcs.2296.

[4] D. Aiger, N. J. Mitra, and D. Cohen-Or. "4-points Congruent Sets for Robust Surface Registration". In: *ACM Transactions on Graphics* 27.3 (2008), #85, 1–10.

[5] Tomas Hodan et al. *BOP Challenge 2020 on 6D Object Localization*. 2020. arXiv: 2009.07378 [cs.CV].

[6] Lahav Lipson et al. *Coupled Iterative Refinement for 6D Multi-Object Pose Estimation*. 2022. arXiv: 2204.12516 [cs.CV].

[7] Yann Labbé et al. *CosyPose: Consistent multi-view multi-object 6D pose estimation*. 2020. arXiv: 2008.08465 [cs.CV].

[8] Kaiming He et al. *Mask R-CNN*. 2018. arXiv: 1703.06870 [cs.CV].

[9] Zachary Teed and Jia Deng. *RAFT: Recurrent All-Pairs Field Transforms for Optical Flow*. 2020. arXiv: 2003.12039 [cs.CV].

[10] Maximilian Denninger et al. "BlenderProc2: A Procedural Pipeline for Photorealistic Rendering". In: *Journal of Open Source Software* 8.82 (2023), p. 4901. DOI: 10.21105/joss.04901. URL: https://doi.org/10.21105/joss.04901.

[11] Joel Vidal, Chyi-Yeu Lin, and Robert Martí. *6D Pose Estimation using an Improved Method based on Point Pair Features*. 2018. arXiv: 1802.08516 [cs.CV].