Front. Comput. Sci., 2025, 19(11): 1911368
https://doi.org/10.1007/s11704-025-41174-x

LETTER

EMP3D: an emergency medical procedures 3D dataset with pose and shape

Hehao BAO[1], Keying DU[1], Xinyi SU[1], Jianqi FAN[1], Jing HUANG[1], Wenlong DONG[2], Lu LU (✉)[2], Kun LI (✉)[1]

1 College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
2 Institute of Disaster and Emergency Medicine, Tianjin University, Tianjin 300072, China

1 Introduction

Emergency Medical Services (EMS) play a critical role in acute emergencies, yet their effectiveness is often limited by professional complexities. For example, a European study on out-of-hospital cardiac arrest (OHCA) found survival rates below 10%, primarily due to delayed responses and insufficient bystander intervention [1].

Existing datasets for medical movement analysis have largely focused on basic patient actions like lying and standing. The NTU dataset [2] includes 2D joint data for actions such as sneezing and covering the head in everyday sickness scenarios. In daily life scenarios [3] uses motion recognition technology to monitor patients' postures, and in hospital environments [4] uses LiDAR for human posture and motion recognition. However, there is a significant gap in research on the actions of rescuers in medical emergency procedures. Creating a 3D dataset of medical emergency procedures can provide data support for the analysis or generation of emergency medical procedures, thereby facilitating the dissemination of emergency medical practices.

To address the gap in research on rescuer movements, we propose EMP3D, which offers notable contributions in terms of both innovation and dataset quality.

- **Pioneering rescuer movement dataset** EMP3D is the first dataset designed to capture the intricate movements of rescuers during emergency procedures. It includes five actions, all reconstructed in detail. This large-scale, standardized 3D dataset focuses on specialized medical actions, serving as a key resource to enhance emergency response training and improve rescue skills.
- **High-precision body reconstruction** EMP3D attains remarkable accuracy by leveraging multi-view cameras, effectively tackling challenges such as depth ambiguity and occlusion. This precision in both pose and shape modeling analysis makes EMP3D a valuable tool for advancing research in public emergency response training and enhancing real-world emergency medical interventions. Data and further details can be obtained at the website of cic.tju.edu.cn/faculty/likun/projects/EMP3D/index.html

Received November 1, 2024; accepted February 24, 2025

E-mail: lulu_998543@tju.edu.cn; lik@tju.edu.cn

Special Column—Code & Data

2 Dataset construction

Data construction The data annotation process is depicted in Fig. 1(a). Six GoPro cameras are utilized to record videos, with their placement shown in Fig. 1(e). Post acquisition, the video streams are synchronized using sound signals and other relevant information. Initially, the 2D poses of six synchronized viewpoints are estimated based on [5]. Subsequently, joint points across these viewpoints are matched through [6] to derive 3D joint coordinates. We have created the Tracking module to track the positions of the rescuers and patients in each frame. The "tracking" module primarily involves trajectory initialization, feature vector construction, cost matrix calculation, and linear matching, followed by trajectory updates based on the assignment results. Ultimately, an accurate 3D SMPL-H [7] parameterized model is obtained via a two-stage optimization method. Both the 2D and 3D results of each frame are manually reviewed, and any issues identified (e.g., poor 2D detection) are corrected to optimize the outcomes. Although the manual inspection process is time-consuming, it ensures a high-quality dataset.

We are more concerned about the posture of the rescuer's body and hands, and hence 3D SMPL-H is the best choice.

Data statistics EMP3D is a specialized dataset for emergency medicine, collected by trained professionals. There are five types of data, as shown in Fig. 1(b), listed in order: CPR, bandaging (two images), hemostasis, Heimlich maneuver, and fracture fixation. The EMP3D dataset contains 57,268 frames of data, and the resolution of the videos is 4K. It includes the original video, synchronized data from six viewpoints, camera parameters, 2D keypoints for each viewpoint, and 3D SMPL-H parameters. The data statistics are summarized in Fig. 1(d).

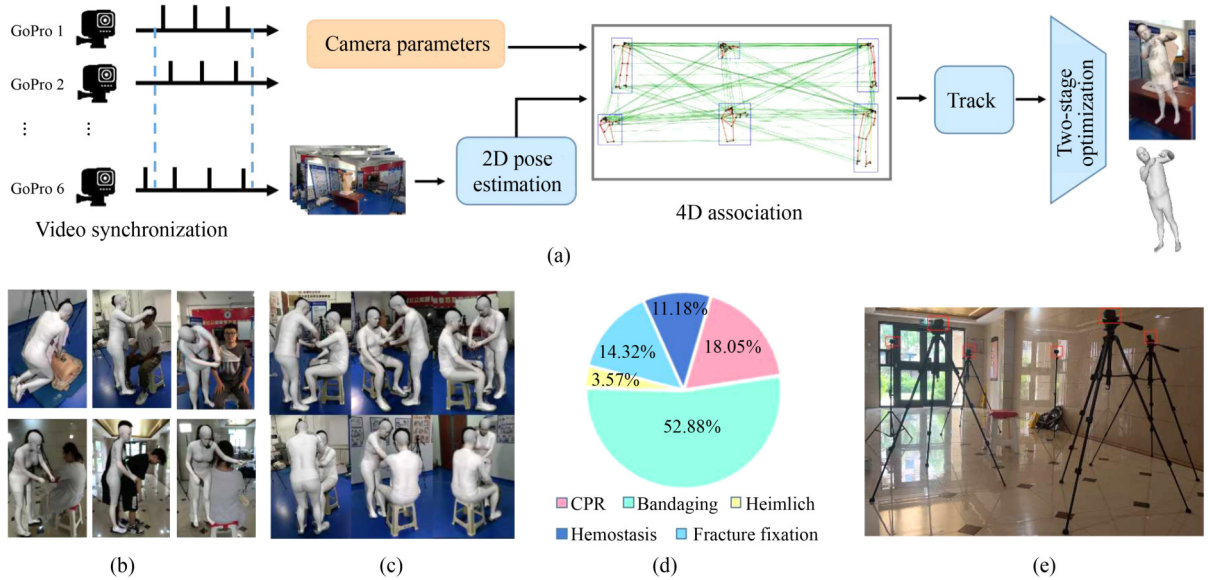

Fig. 1 A detailed explanation of the data annotation process and a visualization of data verification. (a) illustrates the data annotation pipeline; (b) showcases various types of medical emergency actions; (c) We randomly select a type of medical emergency action, and present the projection results of the 3D SMPL-H model from six different viewpoints; (d) displays statistics on data types, and (e) indicates camera positioning

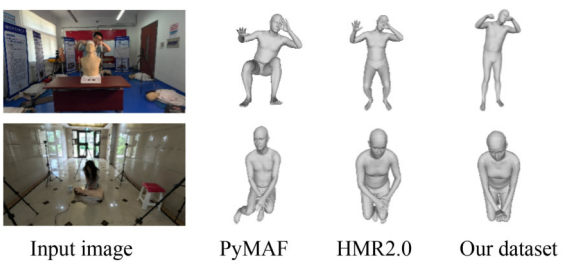

Fig. 2 Examples of benchmark results and our dataset

Data quality By using the multi-view technique shown in Fig. 1(a), our approach addresses occlusion and challenging pose issues. We choose HMR2.0 [8] and PyMAF [9] as models. As shown in Fig. 2, our dataset yields better results in the presence of occlusion or difficult poses. Notably, our dataset yields more robust and reliable results in finger pose estimation. Multi-person scenarios are common in real-world medical emergencies, and as illustrated in Fig. 1(c), our method also performs well in such scenarios.

3 Conclusion

The creation of a sophisticated and standardized 3D dataset, named EMP3D represents a significant step forward in addressing the urgent need for effective emergency response training. By employing advanced motion capture and 3D reconstruction techniques, EMP3D aims to bridge the gap in public training for emergency medical procedures, thereby alleviating the burden on medical professionals and enhancing overall emergency preparedness.

Acknowledgements This work was supported in part by the National Key R&D Program of China (2023YFC3082100), the National Natural Science Foundation of China (Grant No. 62171317), and Science Fund for Distinguished Young Scholars of Tianjin (No. 22JCJQJC00040).

Competing interests The authors declare that they have no competing interests or financial conflicts to disclose.

References

1. Gräsner J T, Herlitz J, Tjelmeland I B, Wnent J, Masterson S, Lilja G, Bein B, Böttiger B W, Rosell-Ortiz F, Nolan J P, Bossaert L, Perkins GD. European resuscitation council guidelines 2021: Epidemiology of cardiac arrest in europe. Resuscitation, 2021, 161: 61–79
2. Liu J, Shahroudy A, Perez M, Wang G, Duan L Y, Kot A C. NTU RGB+D 120: a large-scale benchmark for 3D human activity understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2684−2701
3. Huu P N, Thi N N, Ngoc T P. Proposing posture recognition system combining MobilenetV2 and LSTM for medical surveillance. IEEE Access, 2022, 10: 1839−1849
4. Wu X, Zhang H, Kong C, Wang Y, Ju Y, Zhao C. LiDAR-based 3-D human pose estimation and action recognition for medical scenes. IEEE Sensors Journal, 2024, 24(9): 15531−15539
5. Jiang T, Lu P, Zhang L, Ma N, Han R, Lyu C, Li Y, Chen K. RTMPose: Real-time multi-person pose estimation based on mmpose. 2023, arXiv preprint arXiv:2303.07399
6. Zhang Y, An L, Yu T, Li X, Li K, Liu Y. 4D association graph for realtime multi-person motion capture using multiple video cameras. In: Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 1321–1330
7. Romero J, Tzionas D, Black M J. Embodied hands: modeling and capturing hands and bodies together. ACM Transactions on Graphics (TOG), 2017, 36(6): 1–17
8. Goel S, Pavlakos G, Rajasegaran J, Kanazawa A, Malik J. Humans in 4D: reconstructing and tracking humans with transformers. In: Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. 2023, 14737–14748
9. Zhang H, Tian Y, Zhou X, Ouyang W, Liu Y, Wang L, Sun Z. PyMAF: 3D human pose and shape regression with pyramidal mesh alignment feedback loop. In: Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. 2021, 11426–11436