LivePose: Democratizing Pose Detection for Multimedia Arts and Telepresence Applications on Open Edge Devices

Christian Frisson
Gabriel N. Downs
Marie-Ève Dumas
christian@frisson.re
gndowns@gmail.com
marie-eve.dumas@mail.mcgill.ca
Society for Arts and Technology
Montréal, Québec, Canada

Farzaneh Askari
farzaneh.askari@mail.mcgill.ca
Society for Arts and Technology
McGill University
Montréal, Québec, Canada

Emmanuel Durand
edurand@sat.qc.ca
Society for Arts and Technology
Montréal, Québec, Canada

ABSTRACT

We present LivePose: an open-source (GPL license) tool that democratizes pose detection for multimedia arts and telepresence applications, optimized for and distributed on open edge devices. We designed the architecture of LivePose with a 5-stage pipeline (frame capture, pose estimation, dimension mapping, filtering, output) sharing streams of data flow, distributable on networked nodes. We distribute LivePose and dependencies packages and filesystem images optimized for edge devices (NVIDIA Jetson). We showcase multimedia arts and telepresence applications enabled by LivePose.

CCS CONCEPTS

• **Applied computing** → **Media arts**; • **Computing methodologies** → **Activity recognition and understanding**; • **Computer systems organization** → **Embedded systems**.

KEYWORDS

multimedia arts, pose detection, edge computing, telepresence

ACM Reference Format:
Christian Frisson, Gabriel N. Downs, Marie-Ève Dumas, Farzaneh Askari, and Emmanuel Durand. 2022. LivePose: Democratizing Pose Detection for Multimedia Arts and Telepresence Applications on Open Edge Devices. In *28th ACM Symposium on Virtual Reality Software and Technology (VRST '22), November 29-December 1, 2022, Tsukuba, Japan.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3562939.3565660

1 INTRODUCTION

The human pose (pose for short) is a compact feature that summarizes an image into key points representing the coordinates of human joints in the image. Several tools and libraries curate collections of pose estimation algorithms (HyperPose [5], MediaPipe [7], MMPose [1], OpenPTrack [9]); but are complex to install for optimal performance and to integrate in interactive applications.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
VRST '22, November 29-December 1, 2022, Tsukuba, Japan
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9889-3/22/11.
https://doi.org/10.1145/3562939.3565660

Immersive multimedia experiences, such as performances and installations, require scheduling and sequencing of complex modalities for human-computer interaction, with constraints in multiple dimensions: spatio-temporal aspects, amount of participants, etc. Creators of interactive multimedia artifacts rely on interoperability tools to glue all modalities together. Chataîgne [6] and OSSIA [2] are reminiscent of multimedia sequencers, while libmapper [8] / webmapper [11] focuses on mapping multimodal signals.

One challenge of telepresence applications is to transfer the realism of physical presence into 3d virtual worlds [4]. Avenues to increase realism in telepresence applications include: implementing interaction modalities such as gaze interaction and eye-tracking using computer vision [10]; distributing simulation on edge devices rather than on the cloud [3].

Livepose supports the creation and deployment of interactivity based on pose estimation for multimedia and telepresence applications, distributed on edge devices (NVIDIA Jetson), released under the GPL license[1].

2 SYSTEM

We explain the architecture of our pose estimation tool and its distribution, optimized for edge devices.

2.1 LivePose Architecture

The architecture of LivePose features a 5-stage pipeline.

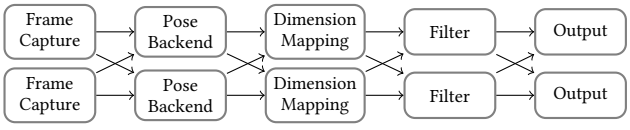

Figure 1: LivePose Data Flow

Frame capture. LivePose captures frames through OpenCV (including, but not limited to V4L2 compatible cameras, MP4 video files, and JPG image files) and pyrealsense2 (Intel Realsense depth cameras and ROSbag recorded files). Supported video frame types are currently color, single-channel, and depth. Setting intrinsic deformation parameters for each frame source is optional for cases like applying a geometrical correction to the captured frames.

[1] https://gitlab.com/sat-mtl/tools/livepose/

Pose estimation . LivePose supports multiple backends whose outputs we unified to a common representation:

- `PoseNet` as default on generic hardware;
- `MMPose` for exploring an extensive collection of pose estimation algorithms from the newest research works;
- `TRTPose` as default on NVIDIA GPUs and Jetson boards it is optimized for using TensorRT.

Dimension mapping. LivePose currently features one instance of dimension mapping: from 2D poses (in the camera frame) to floor position. Pose tracking, pose reidentification, and 2D to the 3D conversion will be implemented in the future.

Filter. LivePose features several filters for tasks, like detecting arms up, the overall orientation of a body, or whether a person is pointing at an object. The Filter stage also enables adapting data for specific uses, including controlling users' avatars in Mozilla Hubs.

Output. LivePose sends data generated by the filters through outputs, currently: Open Sound Control (OSC), libmapper, and WebSocket. These protocols enable the communication with a vast majority of software used by the creative community.

2.2 Distribution optimized for Edge Devices

LivePose is dependent mainly on open-source libraries, except for "freely available commercial packages" by NVIDIA: CUDA, CUDNN, TensorRT. To facilitate over-the-air updates for development and production, we packaged LivePose and all dependencies optimized with CUDA through two distributions, Metalab Package Archives (MPA), both deployed using continuous integration with Gitlab CI:

(1) mpa-jammy-amd64-nvidia: our development distribution is based on Ubuntu 22.04 LTS for amd64 CPU architectures with NVIDIA GPUs (main variant tested: Turing 75)
(2) mpa-focal-arm64-jetson: our production distribution is based on Ubuntu 20.04 LTS for arm64 CPU architectures for NVIDIA Jetson boards (main variant tested: Xavier NX)

We distribute filesystem images so that owners of NVIDIA Jetson boards only need to flash an SD card to discover LivePose. We have forked a repository by Badr Bardi to add two more contributions [2]: support for NVIDIA Jetson Xavier NX boards, and pre-installation of packages of LivePose and dependencies.

3 VALIDATION

LivePose has already been integrated into multimedia arts artifacts during artistic residencies. The *-22.7°C* [3] residency originated the creation of LivePose to design interactions based on the detection of the position of multiple participants onto the floor. We organized the Hacklab'21 hackathon together with MusicMotion [4] during which multiple teams had the possibility to experiment with LivePose to build new inclusive and accessible setups. The Percepto team employed LivePose to control musical events with hand and face movements, making musical creation accessible to people with disabilities. *ComeWithMe* by Marie LeBlanc Flanagan [5] uses pose detection to build a game-like experience based on multi-user movements in the space. *Perspective* by Charlie Leroy [6] and Sébastien Samyn is a collective experience adapting and modifying projection mapping contents depending on the participants proximity. *Podorythmie sensible* by Guillaume Coulombe [7] involves LivePose to complement a footboard equipped with sensors to detect more subtle feet movements.

We employed LivePose to increase the accessibility of telepresence applications by reusing off-the-shelf devices (webcams) to offer alternative interaction means (visual and facial interaction, in addition to gestural interaction), what may benefit people with specific interaction needs. We connected LivePose and Mozilla Hubs via WebSocket to control users' avatars via facial pose estimation. We implemented and paired two WebSocket controller components: the LivePose `Hubs` Filter, and a Mozilla Hubs user-input device[8].

4 DISCUSSION

Future work includes: developing a web-based user interface, automating camera calibration, compressing and synchronizing frames.

REFERENCES

[1] 2022. MMPose: Pose Estimation Toolbox and Benchmark. OpenMMLab. https://github.com/open-mmlab/mmpose

[2] Jean-Michaël Celerier, Pascal Baltazar, Clément Bossut, Nicolas Vuaille, Jean-Michel Couturier, and Myriam Desainte-Catherine. 2015. OSSIA: Towards a Unified Interface for Scoring Time and Interaction. In *TENOR 2015 - First International Conference on Technologies for Music Notation and Representation*. Paris, France. https://hal.archives-ouvertes.fr/hal-01245957

[3] Sahraoui Dhelim, Tahar Kechadi, Liming Chen, Nyothiri Aung, Huansheng Ning, and Luigi Atzori. 2022. Edge-Enabled Metaverse: The Convergence of Metaverse and Mobile Edge Computing. *arXiv:2205.02764 [cs]* (April 2022). arXiv:2205.02764 [cs] http://arxiv.org/abs/2205.02764

[4] John David N. Dionisio, William G. Burns III, and Richard Gilbert. 2013. 3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities. *Comput. Surveys* 45, 3 (July 2013), 34:1–34:38. https://doi.org/10.1145/2480741.2480751

[5] Yixiao Guo, Jiawei Liu, Guo Li, Luo Mai, and Hao Dong. 2021. Fast and Flexible Human Pose Estimation with HyperPose. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, New York, NY, USA, 3763–3766. http://doi.org/10.1145/3474085.3478325

[6] Benjamin Kuperberg. 2022. Chataigne. https://benjamin.kuperberg.fr/chataigne/en

[7] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. *arXiv:1906.08172 [cs]* (June 2019). arXiv:1906.08172 [cs] http://arxiv.org/abs/1906.08172

[8] Joseph Malloch, Stephen Sinclair, and Marcelo M. Wanderley. 2013. Libmapper: (A Library for Connecting Things). In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 3087–3090. https://doi.org/10.1145/2468356.2479617

[9] Matteo Munaro, Filippo Basso, and Emanuele Menegatti. 2016. OpenPTrack: Open Source Multi-Camera Calibration and People Tracking for RGB-D Camera Networks. *Robotics and Autonomous Systems* 75 (Jan. 2016), 525–538. https://doi.org/10.1016/j.robot.2015.10.004

[10] Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2022. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-worn Extended Reality. *Comput. Surveys* 55, 3 (March 2022), 53:1–53:39. https://doi.org/10.1145/3491207

[11] Johnty Wang, Joseph Malloch, Stephen Sinclair, Jonathan Wilansky, Aaron Krajeski, and Marcelo M. Wanderley. 2019. Webmapper: A Tool for Visualizing and Manipulating Mappings in Digital Musical Instruments. In *Proceedings of the 14th International Conference on Computer Music Multidisciplinary Research (CMMR)*.

[2] Jetson image generation https://gitlab.com/sat-mtl/distribution/jetson-images
[3] *-22.7°C* immersive experience: https://sat.qc.ca/fr/molecule-22-7
[4] Hacklab'21: https://musicmotion.org/hacklab21
[5] Marie LeBlanc Flanagan: https://marieflanagan.com/
[6] Charlie Leroy: https://tech-art.ca/
[7] Guillaume Coulombe: https://trad.wiki/creations/podorythmie-sensible/
[8] Mozilla Hubs: https://gitlab.com/sat-mtl/tools/forks/hubs - websocket-controller.js https://vimeo.com/604196712