A summary of VR quadruped embodiment using NeuroDog

Dónal Egan
Trinity College Dublin
Dublin, Ireland
doegan@tcd.ie

Darren Cosker
University of Bath
Bath, United Kingdom
D.P.Cosker@bath.ac.uk

Rachel McDonnell
Trinity College Dublin
Dublin, Ireland
ramcdonn@tcd.ie

KEYWORDS

VR embodiment, Quadruped embodiment, Motion synthesis, Deep Learning, Motion Capture, Perception

Figure 1: Our NeuroDog VR quadruped embodiment system in action. Note the synchrony between the human feet and the front paws of the dog.

1 INTRODUCTION

This work presented here is largely based off the paper *NeuroDog: Quadruped Embodiment using Neural Networks* [Egan et al. 2023]. Traditionally, users of virtual reality (VR) have been limited to embodying humanoid avatars or avatars with strong anthropomorphic qualities. When animal avatars are available, their animations are typically generated via Inverse Kinematics (IK) based on the users' motions - resulting in animations which are unrealistic for the animal type. Potentially, this could lead to a weaker virtual embodiment experience for the user. NeuroDog is the first VR qaudruped embodiment system to use deep-learning for real-time synthesis of realistic motion for the virtual qaudruped. The architecture is designed to preserve synchrony between the user's legs and the quadruped's front legs. The user is able to behave in an intuitive and unconstrained manner, performing actions such as locomotion at different speeds, sitting, or *paw*-raising, with the quadruped mimicking their actions. The NeuroDog architecture is validated via a perceptual experiment in which it is evaluated against a baseline quadruped embodiment system relying solely on IK. Users rate NeuroDog's motion as being more pleasant, more natural and more dog-like and the embodiment experience as being more fascinating. They also experience a stronger sense of animal body ownership, suggesting the importance of animation fidelity when creating VR animal embodiment experiences.

2 NEURODOG SYSTEM

Figure 2 illustrates NeuroDog's run-time pipeline, the core technical novelty of which is the two-network architecture for real-time mapping the user's motion to the virtual quadruped. The user, assuming a bipedal stance, is tracked using a HMD and three sensors located on their pelvis and feet.

If the user's and quadruped's steps are to be synchronous, the user's global velocity cannot be used as the quadruped's target velocity. This is because a human and a quadruped may travel very different distances for a given number of steps. This is the motivation for the first neural network. Its task is to predict a target velocity for the quadruped based on the user's feet velocities, which are first scaled to be in a plausible range of quadruped feet velocities.

The pose prediction network (a mixture-of-experts model [Zhang et al. 2018]) predicts the virtual quadruped's pose, including global position and orientation, at the next time-step. Its inputs include i) a target global trajectory for the quadruped to follow (constructed from the velocity predicted by the first network and the user's facing direction determined by the pelvic sensor), ii) a target action for the quadruped (determined from the user's actions using rule-based methods), and iii) target velocities and contact labels for the quadruped's front paws. The purpose of iii), obtained from the user, is to try and force the virtual quadruped's front legs to move in sync with the user's legs, both for locomotion and other actions such as paw-raises.

IK is used to animate the quadruped's neck and head so that its orientation matches that of the user's head which is determined via the HMD. The position of the camera through which the user views the virtual environment follows the position of the quadruped's head. The orientation of the camera matches that of the HMD. Full technical details of the implementation are available in the original paper [Egan et al. 2023]. Video examples of NeuroDog in action are in the supplementary material.

3 PERCEPTUAL EXPERIMENT

The experiment was a within-subjects study with two conditions, IK and NeuroDog, with the only difference being whether NeuroDog or the baseline IK method was being used to map the users' motion to the virtual quadruped. The virtual environment, illustrated in Figure 1, contained mirrors, in which participants could view their quadrupedal avatar and its movements. For each condition, the user spent approximately five minutes in the virtual environment embodying the quadruped. During this time, they were (voice) prompted by the experimenter to perform a series of tasks. Examples of NeuroDog and the IK baseline method in action are available in the supplementary material. After experiencing each condition, participants removed the HMD and completed a questionnaire.

The questionnaire was based on the Virtual Embodiment Questionnaire (VEQ) [Roth and Latoschik 2020] to investigate the three dimensions of virtual embodiment (body-ownership, agency, and change), together with additional questions to measure motion

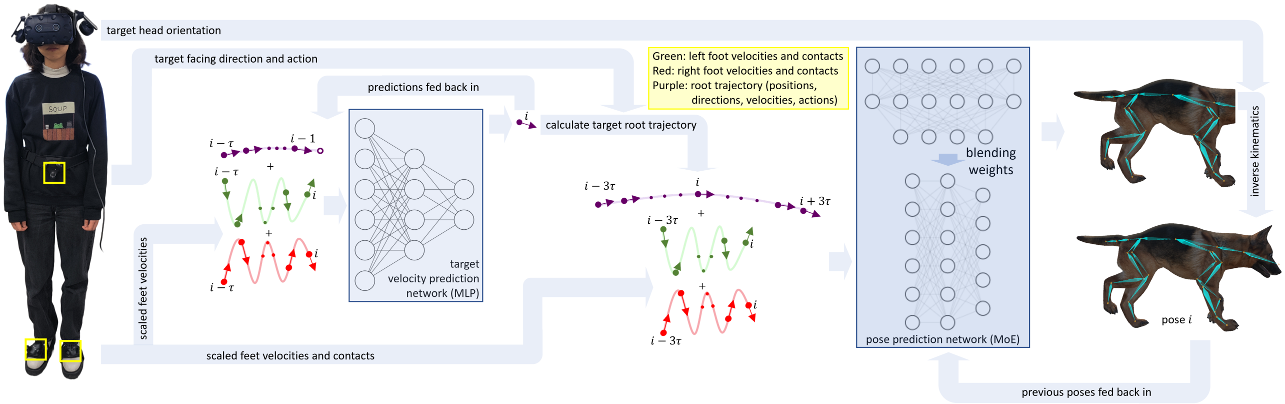

Figure 2: Run-time: Three tracking sensors are used to capture the user's feet velocities and facing direction. These are scaled to ensure they are in an appropriate range for a quadruped's feet velocities. These scaled values, along with the past root target velocity predictions, are fed into the velocity prediction network to predict the desired root velocity for the dog. This, along with the user's facing direction, feet contacts, and action are used to construct the input to the pose prediction network. The quadruped's neck and head are animated via inverse kinematics based off the user's HMD orientation.

quality (naturalness and potential uncanny effects) [Ferstl et al. 2021] and users' fascination with the experience.

The three hypotheses tested were: H1) The system that produces more dog-like natural motion will result in higher virtual body ownership of the virtual dog; H2) Agency will be higher for IK than NeuroDog due to the direct mapping of limbs; and H3) Motion quality will be rated higher for NeuroDog as the network is trained on real dog motion capture.

Results. 21 (10M, 11F) participants, ranging in age from 19 to 35 (M 24.7, SD 5.7), were recruited. A repeated measures ANalysis of VAriance (ANOVA) with within-subjects factor *System* (IK, NeuroDog) was performed for each questionnaire item that met the normality assumption, tested using Shapiro-Wilk test. When normality was not verified, a non-parametric Friedman test was conducted.

Users rated the motion of NeuroDog as significantly more natural ($p < 0.025$), more dog-like ($p < 0.008$) and more pleasant ($p < 0.008$) than for the baseline IK, validating H3. Users also found the NeuroDog experience significantly more fascinating ($p < 0.034$).

No main effect was found for *agency*, indicating that participants experienced similar levels of agency for both conditions and contradicting H2. This validates NeuroDog's architecture as having successfully produced a human-to-dog mapping with high levels of agency synchrony between the user input and quadruped animation.

A main effect of *System* was found on *Ownership* ($p = 0.004$) indicating that participants felt higher ownership in the NeuroDog condition (M4.78, SD1.25) than in the IK condition (M4.1, SD1.31). Additionally, a Spearman's correlation test showed a strong positive correlation between Ownership and Motion Naturalness ratings ($p< .001$), which implies that higher ownership occurred due to the natural animal motion achieved by NeuroDog, validating *H1*.

Further and more detailed results are available in the full paper [Egan et al. 2023].

4 DISCUSSION AND FURTHER RESEARCH

No human data is used in the training stage for NeuroDog and the system does not solve the fundamental problem of mapping or retargeting motions between characters exhibiting extreme morphological differences. Future research is to investigate whether human data can be incorporated into the training process so as to improve the robustness and expressiveness of the human to dog motion mapping. Alternative network architectures will also be explored. It is also interesting to investigate whether the preservation of features additional to, or alternative to, end-effector velocities, can be used to guide the mapping. Examples features which could be explored are the kinetic energy or the contraction index of a pose.

ACKNOWLEDGMENTS

This research was funded by Science Foundation Ireland under the ADAPT Centre for Digital Content Technology (Grant No. 13/RC/2106_P2) and RADICal (Grant No. 19/FFP/6409).

REFERENCES

Dónal Egan, Darren Cosker, and Rachel McDonnell. 2023. NeuroDog: Quadruped Embodiment Using Neural Networks. *Proc. ACM Comput. Graph. Interact. Tech.* 6, 3, Article 38 (August 2023), 19 pages. https://doi.org/10.1145/3606936

Ylva Ferstl, Sean Thomas, Cédric Guiard, Cathy Ennis, and Rachel McDonnell. 2021. Human or Robot? Investigating Voice, Appearance and Gesture Motion Realism of Conversational Social Agents. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents* (Virtual Event, Japan) *(IVA '21)*. Association for Computing Machinery, New York, NY, USA, 76–83. https://doi.org/10.1145/3472306.3478338

Daniel Roth and Marc Erich Latoschik. 2020. Construction of the Virtual Embodiment Questionnaire (VEQ). *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3546–3556. https://doi.org/10.1109/TVCG.2020.3023603

He Zhang, Sebastian Starke, Taku Komura, and Jun Saito. 2018. Mode-Adaptive Neural Networks for Quadruped Motion Control. *ACM Trans. Graph.* 37, 4, Article 145 (jul 2018), 11 pages. https://doi.org/10.1145/3197517.3201366