Video Processing Pipeline

Video Processing Pipeline for Cognition of Interaction

***A. Ecker, F. Wörgötter, A. Gail, P. Wieder**, Institut für Informatik / III. Institut für Physik, Universität Göttingen; Deutsches Primatenzentrum GmbH - Leibniz-Institut für Primatenforschung; Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen*

In Short

- Iterative social interactions area key feature of human and non-human primate social life
- Currently, the dynamics of social interactions and their cognitive underpinnings cannot be fully assessed
- Deep neural networks and machine learning approaches shall be advanced to enable pose estimation and motion tracking under natural conditions

Humans and their closest relatives, the nonhuman primates live in complex social systems, in which they constantly interact with others. The abilities to track and assess others, to cooperate, and to communicate in complex social groups are of utmost importance for survival. As a result, strong selection pressures shaped our socio-cognitive abilities ([1,2]). Cognitive functions such as perception, selective attention, action planning, and decision-making have been subject in cognitive sciences for decades. However, for a long time, they have predominantly been studied in simplified, often individual settings. Real-time social interactions however, unfold during direct face-to-face contact and rely on immediate time-continuous feedback about mutual behaviour ("transparency"). Many studies under natural conditions, in contrast, lack the observational and computational powers to record and process behavioural interactions at high resolution, e.g. with large crowds of people or freely living nonhuman primates. Importantly, we recently demonstrated that under high levels of "transparency" in iterative interactions, other evolutionary successful strategies emerge ([3,4]), underlining the importance of novel approaches to the study of social interactions.
To develop a better understanding of social interactions and the underlying cognitive and neurophysiological mechanisms, the Collaborative Research Center 1528 "Cognition of Interaction" conceptualized two novel experimental platforms - the Dyadic Interaction Platform and the Exploration Room Platform. Both allow to investigate real-time interactions in humans and nonhuman primates. The **Dyadic Interaction Platform** enables two human or nonhuman primate agents to observe each other in the lab and act on the same visual objects in a shared vertical workspace between them. The defining feature of this transparent configuration is that agents can monitor and react to each other's actions dynamically in real-time, emulating naturalistic interactions while still maintaining well-controlled laboratory conditions. The **Exploration Room Platform** allows cognitive testing in social pairs or groups where members are free to explore the room, might follow shared or individual action goals, and coordinate their actions. The defining feature of this innovative setup is the possibility to register behavioural and neural data from more than one individual while interacting freely, minimally constrained in their acting by experimental apparatuses. Conceptually, the Exploration Room Platform shall also be implemented at our three field sites in Madagascar, Thailand and Senegal. Here, we will focus exclusively on behavioural data: free-ranging monkeys will be recorded with multiple cameras from different angles while they interact with their group mates and potentially also with remote-controlled experimental apparatuses.
To take full advantage of these methodological and conceptual advances, we need to observe and quantify complex behaviour at high temporal and spatial resolution, including assessments of body and head orientation (posture), both in interactive lab settings and in the wild. Video recordings provide a relatively easy means of recording behaviour over extended periods of time, but analysing such recordings presents a major bottleneck. For **advanced automated analyses of massive video data**, substantial GPU-based computational power is required. We will develop and apply deep neural networks and the latest machine learning approaches to identify and track monkeys of different species from video data recorded in various environments and will estimate their poses, actions, and inter-actions based on the video data.
DeepLabCut (DLC), a deep-learning-based system for markerless pose estimation ([5,6]) and similar approaches ([7]) enable reliable behavioural tracking of animals in the lab; they can detect and track multiple keypoints at the animals' body joints with only a few hundred labelled frames. These systems work so well because training and application occur in the exact same environment with a simple/static background and constant lighting conditions. This does, however, not generalise to behavioural tracking of

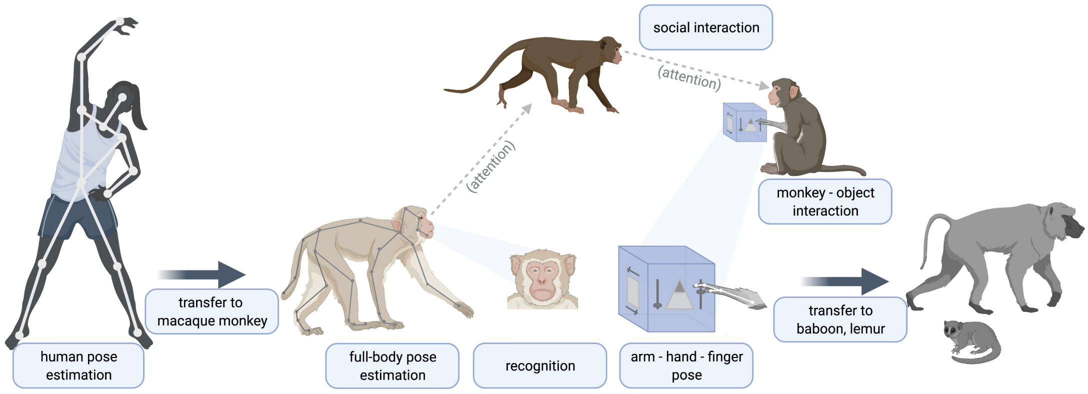

Figure 1: *Planned approach to advance currently available pose estimation and tracking approaches*

animals under field conditions. Here, environment and viewpoint change frequently, the camera is not necessarily static and a heterogeneous group of varying appearance (e.g., age, sex) has to be monitored. Current solutions are not applicable to such situations and require additional labelled data to generalise across individuals (Mathis et al. 2021).

To achieve an exhaustive understanding of social interactions and their cognitive underpinnings under naturalistic conditions and in natural environments, further advancement of fully-automated motion tracking and pose estimation systems, as well as their transfer to field conditions are essential (see 1 for a graphical illustration of the planned approach). Starting with existing a state-of-the-art system for human pose estimation, we aim to continuously expand the system to i) estimate full-body and arm-hand-finger poses of macaque monkeys, ii) be applicable under varying enclosure environments, iii) deal with larger number of individuals, iv) work in complex natural environments, and v) cope with different monkey species while minimising annotation effort. Taken together, this will significantly advance the automated analysis of complex video data, build bridges between lab and field research, and will play a key role to the understanding of cognition of interactions.

More Information

[1] J. Salle, R.B. Mars, M.P. Noonan, J.L. Andersson, J.X. O'Reilly, S. Jbabdi, P.L. Croxson, M. Jenkinson, K.L. Miller, M.F.S. Rushworth *Science* **334**, 697-700 (2011). doi:10.1126/science.1210027

[2] J.B. Smaers, A. Gómez-Robles, A.N. Parks, C.C. Sherwood *Curr. Biol.* **27(5)**, 714-720 (2017). doi:10.1016/j.cub.2017.01.020

[3] A.M. Unakafov, T. Schultze, I. Kagan, S. Moeller, A. Gail, S. Treue, S. Eule, F. Wolf *EvoApplications 2019*, 204-219 (2019). doi:10.1007/978-3-030-16692-2$_1$4

[4] A.M. Unakafov, T. Schultze, A. Gail, S. Moeller, I. Kagan, S. Eule, F. Wolf *PLoS Comput. Biol.*, **16(1)**, e1007558 (2020). doi:10.1371/journal.pcbi.1007588

[5] A. Mathis, P. Mamidanna, K.M. Cury, T. Abe, V.N. Murthy, M.W. Mathis, M. Bethge *Nat. Neurosci.*, **21(9)** 1281-1289 (2018). doi:10.1038/s41593-018-0209-y

[6] T. Nath, A. Mathis, A.C. Chen, A. Patel, M. Bethge, M.W. Mathis, *Nat. Protoc.*, **14(7)** 2152-2176 (2019). doi:10.1038/s41596-019-0176-0

[7] A. Arac, P. Zhao, B.H. Dobkin, S.T. Charmichael, P. Golshani *Front. Syst. Neurosci.*, **13** 20 (2019). doi:10.3389/fnsys.2019.00020

Project Partners

Institut für Informatik / III. Institut für Physik, Universität Göttingen; Deutsches Primatenzentrum GmbH - Leibniz-Institut für Primatenforschung; Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Funding

DFG Sonderforschungsbereich (SFB) 1528

DFG Subject Area

206-04 Cognitive, Systems and Behavioural Neurobiology