Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The Infant-Robot Interaction (IRI) video dataset was created as part of an ongoing project to study dyadic infant-robot interaction. The dataset was used to develop a socially assistive, non-contact, infant-robot interaction system. The goal of the system is to provide contingent positive feedback to increase exploration and expand early movement practice in infants. This research has the potential to advance knowledge about which aspects of movement infants can adjust and how to most effectively guide their movement to enhance their neuromotor control, both for infants with typical development and for infants at risk for developmental disabilities. The IRI video dataset was created to evaluate the infant-robot interaction system and to study infant behavior during the interactions.

Who created this dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The IRI video dataset was jointly collected by the Infant Neuromotor Control Lab led by Dr. Beth A. Smith at Children's Hospital Los Angeles and the Interaction Lab led by Prof. Maja J. Matarić at the University of Southern California.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

Funding was provided by the National Science Foundation under grant NSF CBET-1706964.

Any other comments?

None.

Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

The instances in this dataset are videos recorded during 33 infant-robot interactions. Interactions were simultaneously documented with three cameras: one focused on the face of the infant, one focused on the full body of the infant, and one captured the infant and robot in profile-view. Face-view videos had variable frame rates while body-view videos had frame rates of 29.97 frames per second.

How many instances are there in total (of each type, if appropriate)?

There are 27 face-view videos, 32 front-view videos, and 33 side-view videos. There are 266.86 minutes (4.45 hours) of face-view video and 300.39 minutes (5.01 hours) of body-view video.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

Due to piloting and adjustments to the recording setup, certain camera views were unavailable during some of the interactions. Additionally, user testing for the infant-robot interaction system is ongoing. This dataset consists of recordings from the first 3 studies conducted as part of this ongoing project.

What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.

Each instance consists of raw RGB video data.

Is there a label or target associated with each instance? If so, please provide a description.

Behavioral annotations describe each video frame along a five-point arousal scale (alert, fussy, crying, drowsy, or sleeping). Each annotator achieved over 85% label agreement. Agreement was assessed by comparing the annotations of videos labeled by 2 separate annotators. After excluding interactions with missing video views, the proportions of affective labels are 84.3% alert, 13.4% fussy, 2.3% crying, 0.0% drowsy, and 0.0% sleeping. Participant demographic information including age, race, ethnicity, and sex is also included but is not used as a label in machine learning given the size of the dataset.

Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

Two of the interactions were not annotated for affective state at the start of the affect recognition project. Six interactions did not capture face-view video and one interaction did not capture body-view video due to changes

in the number of cameras as the infant-robot interaction setup was developed.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.

Certain infants engaged in the infant-robot interaction twice, at two different ages. This is documented in the naming scheme of participant data. The corresponding side, front, and face-view videos and the demographic information and labels are identified by the file naming schema in the dataset.

Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.

When developing models that will be used with new participants, each infant's data should belong in a single data split. Should the same infant appear in both training and validation or test sets, learning models could exploit patterns in an individual infant's behavior and adjacent video frames that are unlikely to generalize across infants.

Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.

In many cases, infants occluded parts of their own bodies or faces with their hands or looking away from the camera (e.g., toward their caregivers or at the floor). Some infants also wore an eye-gaze tracker, which sometimes occluded their faces. Though face and body occlusions are not explicit errors to the video themselves, they inhibit landmark prediction and complicate downstream modeling of the infants.

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a future user? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

The dataset is entirely self-contained and does not rely on external resources.

Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals non-public communications)? If so, please provide a description.

The dataset contains identifiable videos of infants and is therefore only available to researchers named in University of Southern California Institutional Review Board under protocol #HS-14-00911.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.

The dataset does not contain content that is commonly considered offensive, insulting, threatening, or anxiety-causing.

Does the dataset identify any subpopulations (e.g., by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

Ages of infant participants ranged between 6 to 9 months, with a mean age of 220 days (SD=23). Of the 26 infants, 19 were female and 7 were male. Ethnicity was reported by each infant's parent, with 7 identified as Hispanic or Latino. Race was also reported by parents, with 2 identified as Asian, 1 as Black or African American, 13 as White, 9 as "other", and 1 parent declined to answer.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset? If so, please describe how.

Yes: the dataset contains videos of the participating infants.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.

The demographic information includes the race and ethnicity reported by the infants' parents.

Any other comments?

None.

Collection Process

How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age

or language)? If data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

Face-view, body-view, and profile-view videos were captured simultaneously by three cameras. Details on manual annotations are discussed above.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)? How were these mechanisms or procedures validated?

The face-view camera was placed slightly in front of the infant to the infant's right. The body-view video was positioned slightly further from the front of the infant to capture the infant's whole body. The profile-view camera was placed to the side of the infant and robot. Each camera was connected to a tripod or stand.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

The dataset is not a sample from a larger data collection.

Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?

Work for this dataset was supported by the National Science Foundation under grant NSF CBET-1706964. Students and faculty were both involved in data collection. Families were compensated for participation.

Over what timeframe were the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

The video data were captured during each infant-robot interaction.

Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

This study procedure was approved by the University of Southern California Institutional Review Board under protocol #HS-14-00911.

Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?

The video data were recorded by the research team at the University of Southern California, and videos included direct footage of the infants.

Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

The details of the study for which the IRI video dataset was collected were described to each infant's parent.

Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

One parent or legal guardian signed an informed consent form before their infant participated.

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

The participants were informed that revoking their consent would prevent their data from being used or distributed for future research studies, even if all their identifiers are removed.

Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

Published papers describing the research conducted with this dataset are available here: https://uscinteractionlab.web.app/project/babies. An initial analysis of the impact of the larger project encompassing this dataset was conducted as part of the grant application for NSF CBET-1706964.

Any other comments?

None.

Preprocessing/cleaning/labeling

Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?

If so, please provide a description. If not, you may skip the remainder of the questions in this section.

Facial landmark recognition and facial action unit recognition was applied on face-view video with OpenFace 2.0. Body skeleton landmark recognition was applied with and OpenPose.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.

The raw dataset and preprocessed data are not publicly available.

Is the software used to preprocess/clean/label the instances available? If so, please provide a link or other access point.

OpenFace 2.0 source code can be found in https://github.com/TadasBaltrusaitis/OpenFace. OpenPose source code is provided in https://github.com/CMU-Perceptual-Computing-Lab/openpose.

Any other comments?

None.

Uses

Has the dataset been used for any tasks already? If so, please provide a description.

Published papers describing the research conducted with this dataset are available here: https://uscinteractionlab.web.app/project/babies.

Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

Published papers describing the research conducted with this dataset are available here: https://uscinteractionlab.web.app/project/babies.

What (other) tasks could the dataset be used for?

Future directions using this dataset include the modeling of affect in the presence of missing data.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a future user might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other undesirable harms (e.g., financial harms, legal risks) If so, please provide a description. Is there anything a future user could do to mitigate these undesirable harms?

The dataset includes raw video data to support improved landmark recognition among other preprocessing methods. A demographic breakdown should be provided in papers that use this dataset to support transparency and reproducibility. However, given the size of the dataset, claims about a specific demographic should not be made using this dataset.

Are there tasks for which the dataset should not be used? If so, please provide a description.

Due to the limited sample size, claims about a specific demographic (aside from infant age) should not be made using this dataset.

Any other comments?

None.

Distribution

Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.

The dataset will not be distributed publicly or to third party entities.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub) Does the dataset have a digital object identifier (DOI)?

N/A

When will the dataset be distributed?

N/A

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

N/A

Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

N/A

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a

link or other access point to, or otherwise reproduce, any supporting documentation.

N/A

Any other comments?

None.

Maintenance

Who will be supporting/hosting/maintaining the dataset?

The Interaction Lab at the University of Southern California and the Infant Neuromotor Control Lab at Children's Hospital Los Angeles will be maintaining the dataset.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

You may contact Lauren Klein at kleinl@usc.edu or Dr. Beth A. Smith at bsmith@chla.usc.edu.

Is there an erratum? If so, please provide a link or other access point.

No erratum exists for the dataset.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to users (e.g., mailing list, GitHub)?

No updates or corrections to the dataset are planned to be made to the best knowledge of the authors of this datasheet.

If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.

The data will be retained by the Interaction Lab at the University of Southern California and the Infant Neuromotor Control Lab at Children's Hospital Los Angeles.

Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to users.

Older versions of the dataset will not be maintained.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to other users? If so, please provide a description.

No mechanism to augment the dataset exists.

Any other comments?

None.