

Erratum to: Recognizing Micro-Actions and Reactions from Paired Egocentric Videos (CVPR2016)

Ryo Yonetani
The University of Tokyo
Tokyo, Japan
yonetani@iis.u-tokyo.ac.jp

Kris M. Kitani
Carnegie Mellon University
Pittsburgh, PA, USA
kkitani@cs.cmu.edu

Yoichi Sato
The University of Tokyo
Tokyo, Japan
ysato@iis.u-tokyo.ac.jp

Due to some incorrect labels annotated to the proposed Paired Egocentric Video (PEV) dataset, the dataset description (Section 3.1) and results (Table 1 and Figure 3) in the original paper were partially inaccurate. The correct results improved the detection and classification performances of the proposed method (0.89 \rightarrow 0.90 as an average AUC score in Table 1 and 0.66 \rightarrow 0.69 as an average classification accuracy in Figure 1) and did not change our conclusion that the first-person and second-person POV features of two interacting parties were complementary and essential for recognizing micro-actions and reactions. Please contact yonetani@iis.u-tokyo.ac.jp for further questions.

Dataset descriptions

1. **Pointing (182 samples)**: Pointing to a certain location, an item, or person B to initiate interaction, which is followed by B 's reactions such as orienting of attention and positive or negative responses.
2. **Attention (97 samples)**: Orienting attention with slight head motion to what is pointed to by B .
3. **Positive (159 samples)**: Responding positively by widely or subtly nodding and/or by laughing with body motion to B 's pointing or gesture.
4. **Negative (40 samples)**: Responding negatively by shaking or slightly cocking one's head and/or crossing arms to B 's pointing or gesture.
5. **Passing (119 samples)**: Initiating or finishing passing an item to B in order to exchange it.
6. **Receiving (143 samples)**: Initiating or finishing receiving what B is trying to pass.
7. **Gesture (171 samples)**: Doing head and/or hand gestures to converse with B , which can be followed by B 's gesture and positive or negative responses.

References

- [1] Y. Li, Z. Ye, and J. M. Rehg. Delving into egocentric actions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 287 – 295, 2015. 2
- [2] Y. Poley, C. Arora, and S. Peleg. Temporal segmentation of egocentric videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2537–2544, 2014. 2
- [3] M. S. Ryoo, B. Rothrock, and L. Matthies. Pooled motion features for first-person videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 896 – 904, 2015. 2
- [4] H. Wang and C. Schmid. Action recognition with improved trajectories. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3551 – 3558, 2013. 2
- [5] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, pages 568–576, 2014. 2
- [6] L. Wang, Y. Qiao, and X. Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4305 – 4314, 2015. 2

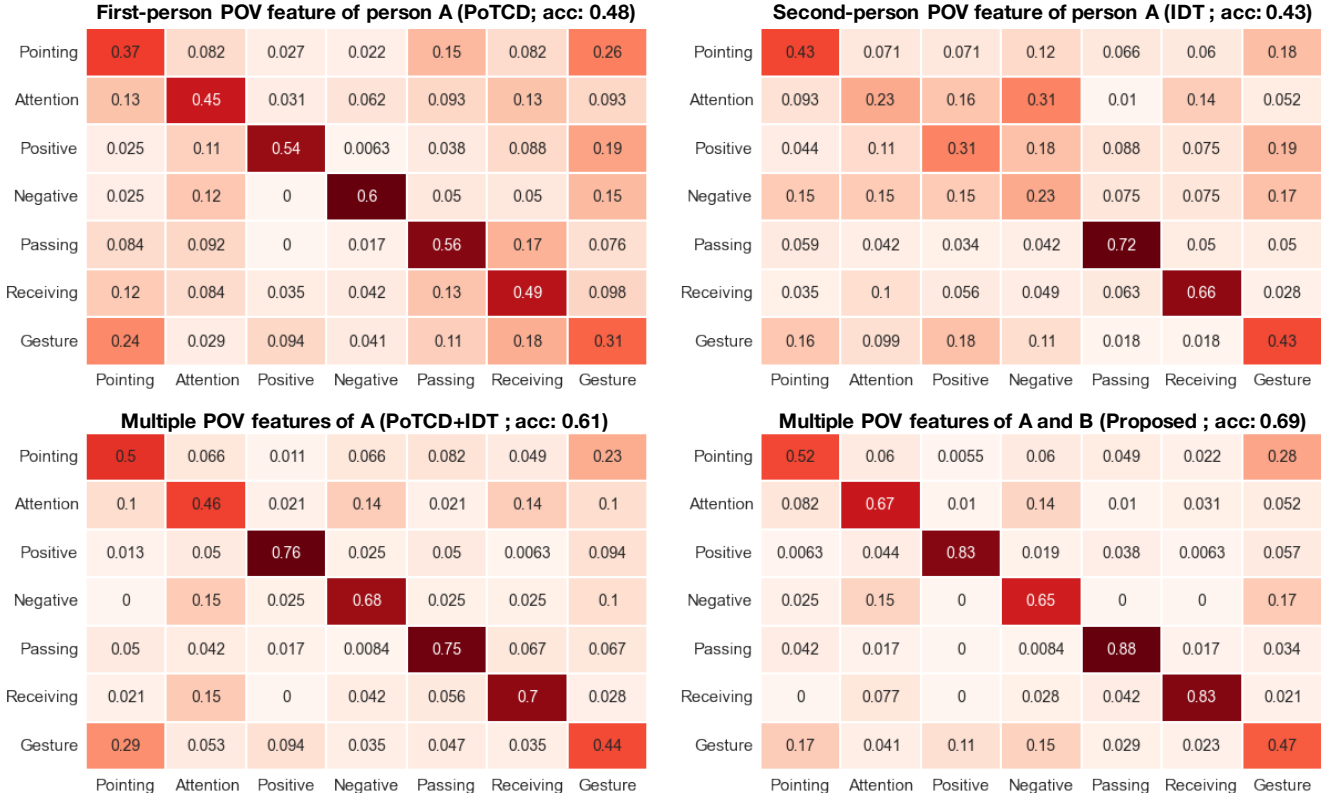


Figure 1. Confusion matrices and average accuracies of the classification task on the PEV dataset.

Table 1. AUC scores on the detection task. (1) First-person POV features of target person *A*. (2) Second-person POV features of *A*. (3) Combinations of first-person and second-person POV features of *A*. (4) Combinations of multiple POV features of persons *A* and *B*.

		Pointing	Attention	Positive	Negative	Passing	Receiving	Gesture	Average
(1) First-person POV features of <i>A</i>	E [1]	0.65	0.76	0.92	0.92	0.62	0.75	0.68	0.76
	E+O [1]	0.70	0.74	0.94	0.69	0.79	0.85	0.68	0.76
	CD [2]	0.61	0.59	0.55	0.66	0.76	0.66	0.60	0.63
	PoTCD [2, 3]	0.71	0.70	0.91	0.92	0.81	0.77	0.67	0.78
(2) Second-person POV features of <i>A</i>	IDT [4]	0.76	0.70	0.69	0.58	0.92	0.91	0.80	0.76
	TCNN [5]	0.63	0.56	0.56	0.55	0.60	0.67	0.61	0.60
	TDD [6]	0.64	0.71	0.61	0.50	0.71	0.79	0.56	0.64
(3) Multiple POV features of <i>A</i>	E+IDT	0.78	0.73	0.86	0.78	0.93	0.92	0.82	0.83
	E+O+IDT	0.82	0.77	0.95	0.85	0.94	0.95	0.77	0.86
	PoTCD+IDT	0.82	0.78	0.97	0.91	0.93	0.92	0.77	0.87
(4) Multiple POV features of <i>A</i> and <i>B</i>	Degraded-A	0.82	0.79	0.95	0.95	0.95	0.96	0.68	0.87
	Degraded-B	0.78	0.69	0.69	0.60	0.93	0.91	0.80	0.77
	Proposed	0.87	0.83	0.97	0.92	0.97	0.97	0.76	0.90