Paper Review : Event Driven Visual Tactile Sensing for Robots

Kartik Srinivas
IIT Hyderabad, India
Hyderabad, India
es20btech11015@iith.ac.in

I. General Notes

The paper [TSS+20] gives a full event-based perception framework(consisting of Visual and Tactile Sensors) that combines vision and touch to achieve better performance on the tasks of rotational slip detection and weight classification. They have given a comparison of the performance of DNNs in all three scenarios(Visual only, Tactile only, and Both together) and compared it to that of Spiking Neural Networks. The paper has provided extensive details of the experiments that are used for measuring performance. The paper has also provided the data collected using their sensors for public use.

II. High Level Analysis

The paper is incremental in nature because it is building upon previous applications of ANN's being applied in Visual tasks that are handled by robots and extending them to become event-driven and use SNN's. The idea to use ACES and the proposed architecture of the Spiking Neural Network is novel and remarkable. The results that are shown in the paper are very systematic and give a good picture of the performance of Spiking Neural Nets when fed with event-based data. This paper will be of good use to the reader who is trying to build systems that operate efficiently on discrete sensory data and the extensive nature of the experiments conducted along with the publically available dataset form a good starting point.

III. Technical Analysis

The paper should pay more attention to the styles in which the sensory data can be turned into input spike trains and then provide a detailed analysis of it. Instead, the paper is filled with too many details about the experiments that are carried out. These details should instead be supplied as supplementary material, and there should be more focus on explaining how the sensory data is turned into spikes. There are only around 3-4 lines of explanation for this using binary windows or **bins**. This aside, There should be an explicit table where all the tiny hyper-parameters of training are being mentioned (instead, they are spread across the paper). for example

1) Desired number of spike trains for classification
2) The parameters of the quadratic decay in the weightage of losses
3) The optimal resolution size that is passed from the Visual camera (Prophesee)
4) The supposed l_2 regularization parameters.

- No reason is given for enforcing *Early classification loss* to be lower.

1) Concerns:

- If you are using multiple instances .i.e. (15 samples) [1] of the arm raising the same particular object as **independent samples** during training, then I think this might cause a problem because you are comparing the performance of ANN's and DNN's on **dependent data**. The models are now measured on their ability to learn this supposed dependence between examples rather than the weights of the objects .
- Since you are using only 15 samples per class (which is much lesser than the usual number of examples that DNNs train on) Is it possible there is overfitting for the MLP-GRU?

2) Pitfalls: The paper has a small pitfall because the reader might think that the experiments are done to optimize the model's performance on the classification tasks. However, this is not the case [2]. The author claims that better performance can be obtained by incorporating proprioceptive data into the VT-SNN. Apparently, the author did not want to do this because he wanted to show that the multi-modal system was effective at detecting differences as compared to the single sensor-based architecture.

IV. Minor corrections

As I mentioned earlier, it would be more useful to give a full section on *How the data is passed to the VT-SNN* rather than spreading it out across different parts of the paper, thereby, mixing it with the raw details of the experiments. The diagrams are, however, spot-on and very informative. As mentioned earlier in [III] , the author should try to condense all the hyperparameters of training into one table.

V. Rating

Overall Paper rating : 8/10
Confidence: 3

References

[TSS+20] Tasbolat Taunyazov, Weicong Sng, Hian Hian See, Brian Lim, Jethro Kuan, Abdul Fatir Ansari, Benjamin CK Tee, and Harold Soh. Event-driven visual-tactile sensing and learning for robots. *arXiv preprint arXiv:2009.07083*, 2020.

[1] Mentioned in the paper at the section *Robot Motion*
[2] See the section *Container and Weight Classfication*