Biologically-Inspired Spatial Filtering for Photon-Event LiDAR with Neural Networks

Matthias Aquilina[1], Pietro Bartoli [2] and Ashley Lyons[1]

[1] *Extreme Light Group, Advanced Research Centre, University of Glasgow, Glasgow, United Kingdom*
[2] *Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Milan, Italy*

The time domain is an incredibly useful resource for computational imaging techniques. Previous studies, such as Single Photon Event Driven 3D Imaging (SPEDI), have shown that by using a Single Photon Avalanche diode (SPAD) and time-correlated single-photon counting to obtain the time-of-flight of photons from an illuminated scene, you have sufficient information for a 3D scene to be reconstructed [1]. SPEDI processes temporal data directly from each photon's timestamp, reducing latency by updating the 3D scene estimate photon-by-photon. This is an extreme example where no spatial information is measured directly by the system; however, there may be an advantageous middle-ground where a small amount of spatial information is directly encoded in the measurement [2]. We explore this concept by using a biologically-inspired preprocessing filter at the SPAD photon-processing phase to adjust which photons are accepted or ignored, which streamlines the incoming data spikes into one information-rich stream. We also divide the photons into multiple data spike streams, training our autoencoder on two and three separate streams of incoming data, each one highlighting a different feature in the scene.

Akin to SPEDI, we use a LiDAR-like setup to flood-illuminate the scene with a pulsed laser and the back-scattered photons are collected and processed by the SPAD. Timestamps are obtained for each detected photon over consecutive pulses which are continuously streamed from the array to form a temporal histogram. Objects placed at different positions within the scene provide different distributions of arrival times at the sensor [2]. Cells in a creatures vision system put together the spike signals from photo-receptor cells and filter out critical information from the scene to send to the brain. We generate different training histograms based on these cells' spike-filtering method, which the network is trained on either one or multiple of, however blind to what the different histograms represent. We use an ANN autoencoder that converts histogram inputs of shape $(B, C, 1249)$ into depth maps of shape $(B, 1, 60, 80)$. The model combines a 1D convolutional encoder for temporal histogram compression with a 2D transposed-convolution decoder for spatial reconstruction. In our setup, the encoder output is projected to a $(512, 5, 5)$ seed tensor, which is progressively upsampled to the final depth map and passed through a sigmoid.

The comparative analysis presented in Figure 1 demonstrates that multi-channel histogram configurations enhance the visual quality of depth map reconstructions over single-channel baselines. By partitioning the photon stream into structured channels, this biologically-inspired preprocessing method effectively integrates additional spatial cues into the temporal representation, resulting in reconstructions that more accurately reflect ground-truth object shapes and scene structures. Experimental results on scenes with moving subjects and static backgrounds indicate that two- and three-channel inputs substantially reduce common reconstruction ambiguities while maintaining geometric stability. Future research will explore the scalability of this channelised representation in more complex scenarios involving dynamic backgrounds.

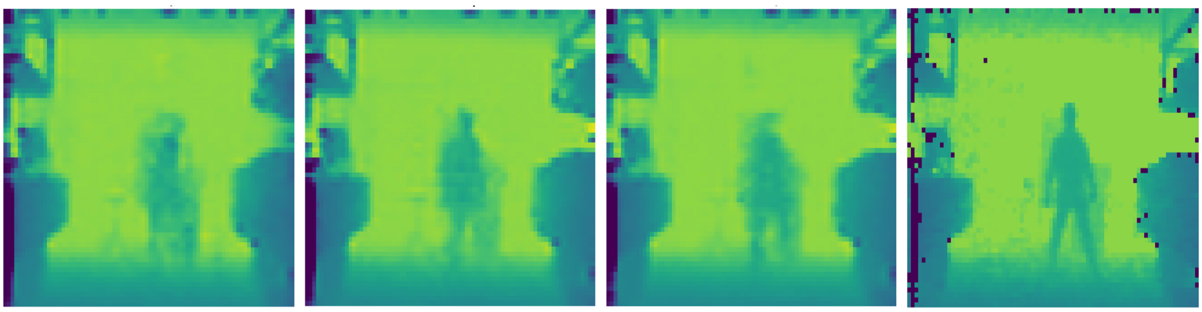

Fig. 1: *Three images showing the generated scene by the autoencoder when (from left to right) 1, 2 and 3 histograms are used to train the network after 58+ epochs of training. The* 4^{th} *image (far right) is the ground truth.*

References

[1] Alex Turpin, Gabriella Musarra, Valentin Kapitany, Francesco Tonolini, Ashley Lyons, Ilya Starshynov, Federica Villa, Enrico Conca, Francesco Fioranelli, Roderick Murray-Smith, and Daniele Faccio. Spatial images from temporal data. *Optica*, 7(8):900–905, Aug 2020.

[2] Alex Vicente Sola, Matthias Aquilina, Paul Kirkland, and Ashley Lyons. Single photon event-driven 3d imaging. *Communications Engineering*, 5, 11 2025.