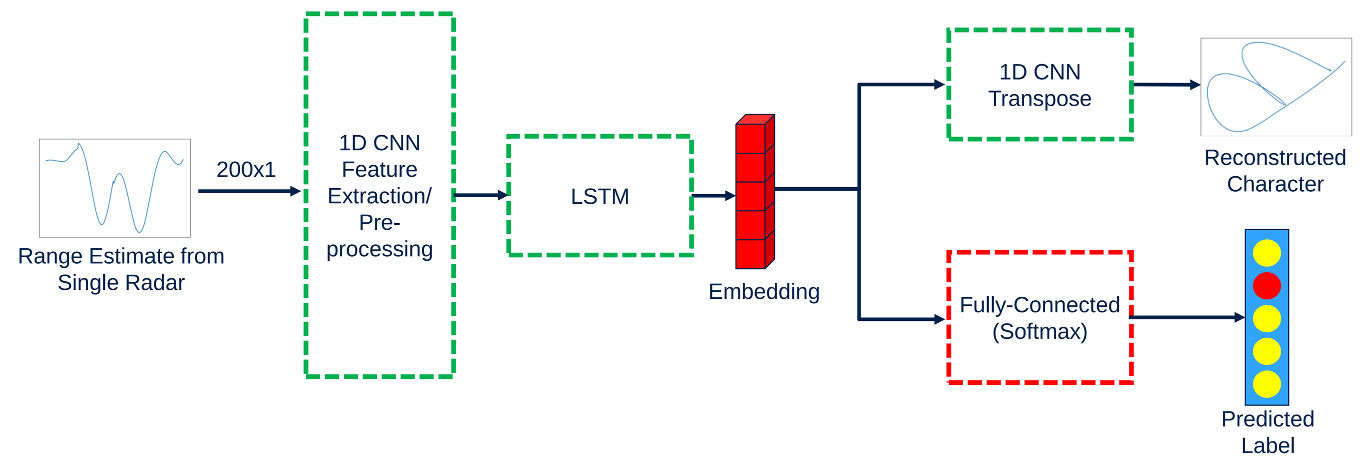
Air-Writing with Sparse Network of Radars using Spatio-Temporal Learning

Muhammad Arsalan*Ŧ, Avik Santra*, Kay Bierzynski*, Vadim IssakovŦ
* Infineon Technologies AG, Neubiberg, Germany
Ŧ Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany

Abstract

Traditionally, radar-based air-writing systems have been based on a network of radars, at least three, to localize the hand target through trilateration algorithm followed by tracking to extract the drawn trajectory, which is then followed by recognition of the drawn character by either Long-Short Term Memory (LSTM) utilizing the sensed trajectory or Deep Convolutional Neural Network (DCNN) utilizing a reconstructed 2D image from the trajectory. However, the practical deployments of such systems are limited since the detection of the finger or hand target by all three radars cannot be guaranteed leading to failure of the trilateration algorithm. Further placement of three or more radars for the air-writing solution is neither always physically plausible nor cost-effective. Furthermore, these solutions do not exploit the full potentials of deep neural networks, which are generally capable of learning features implicitly. In this paper, we propose an air-writing system based on a network of sparse radars, i.e. strictly less than three, using 1D DCNN-LSTM-1D transposed DCNN architecture to reconstruct and classify the drawn character utilizing only the range information from each radar. The paper employs real data using one and two 60 GHz milli-meter wave radar sensors to demonstrate the success of the proposed air-writing solution.

Proposed Solution

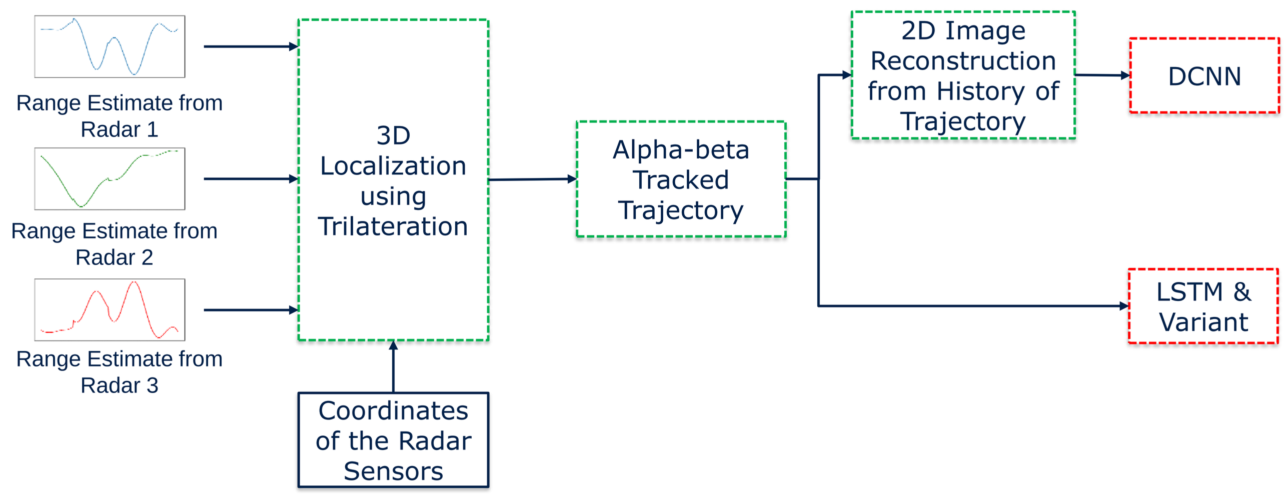


- An air-writing system using a sparse network of radars, i.e. only one or two radars
- A novel 1D DCNN-LSTM-1D transpose DCNN architecture is proposed for simultaneous recognition of the drawn character and reconstruction of the 2D trajectory image.
- The feature images are learned intrinsically by the deep neural network.
- The trilateration, $\alpha\beta$ tracking filter and tracking in the conventional pipeline are replaced by 1D-DCNN for feature extraction.
- The extracted features are fed to LSTM and 1D transposed DCNN for temporal modeling of the local and global trajectory.
- The LSTM encodes the drawn trajectory per character into an embedding feature vector, from which the 1D transposed DCNN reconstructs the global trajectory.

Conclusion

Air-writing offers a simple and intuitive alternative to touch and keyboard interfaces to machines and can be used for augmented reality-virtual reality applications among others. We propose an air-writing system based on sparse network of millimeter-wave radars, i.e. strictly less than three. The proposed system uses only the range information from these radars and is fed into 1D DCNN-LSTM-1D transposed DCNN to reconstruct the 2D drawn image with respect to a reference and classify the drawn character. We demonstrate the reconstruction and classification performance of our proposed air-writing solution using sparse network of FMCW radars.

State-of-the-Art



- Air writing system based on network of three FMCW radars.
- The drawn trajectory is sensed by individual radars, translated to global coordinates through trilateration and smoothened through $\alpha\beta$ filter
- The drawn trajectory is fed into LSTM for classification or transformed into 2D trajectory image and fed into DCNN for classification.

Drawbacks:

- Failure of trilateration due to occlusion of fingers on one or more radar and general missed detection leading to missing or unreliable trajectory coordinates.
- Placement of network of radars on a computer screen or AR-VR device to maximize the intersecting field-of-view (FoV)
- Furthermore, using the 2D trajectory image as feature images for classification using DCNN does not fully explore the capabilities of deep neural networks, which are capable

Classification & Reconstruction Results

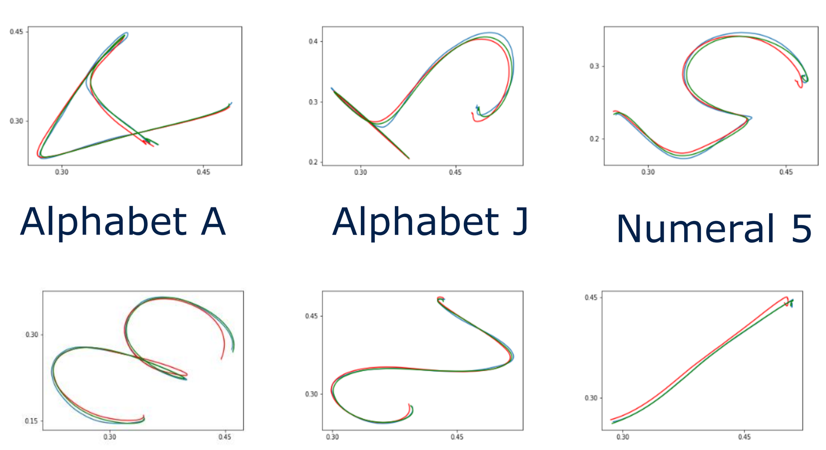
Classification accuracy of the proposed solution to other models and baseline using three radars

Proposed Method	Avg. Accuracy (%)	Baseline*	Avg. Accuracy (%)
Single radar	90.33 ± 4.44	LSTM-CTC	93.33
Two radars	97.33 ± 2.67	BLSTM-CTC	96.67
		ConvLSTM-CTC	98.33
		DCNN	98.33

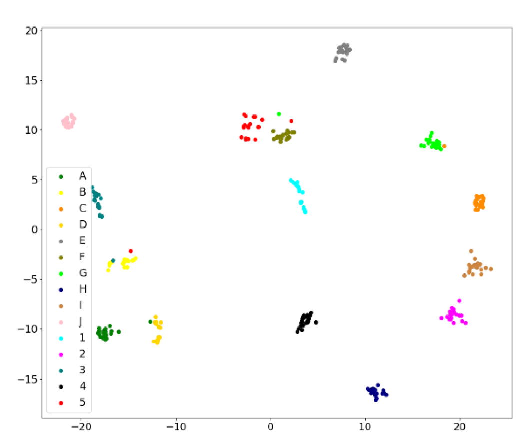
Average L_{MSE} for reconstruction

Radar configuration	Avg. L_{MSE} (e-4)
Single radar	6.8 ± 41.5
Two radars	1.4 ± 0.7

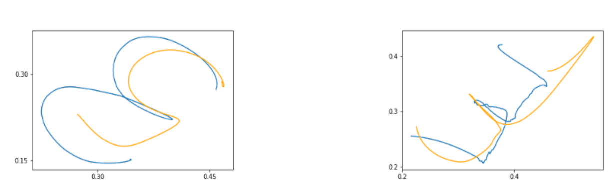
Classification & Reconstruction Results



Examples that were misclassified and wrongly reconstructed (orange). Blue represents the true label.



t-SNE of embedding learned for characters drawn in case of two radars.



Reconstruction of some of the reference trajectories (blue) of alphabets and numerals with one radar (green) and two radars (red).