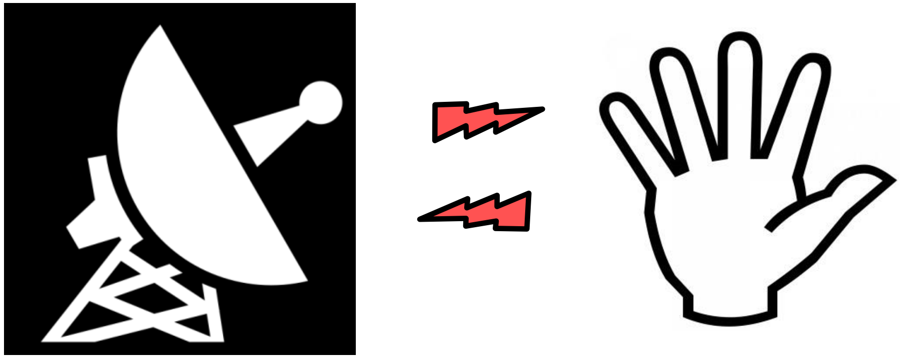
Hidden Markov Model-based Gesture Recognition with FMCW Radar

Greg Malysa, Dan Wang, Lorin Netsch, and Murtaza Ali

Texas Instruments

8 Dec 2016

Our Goal

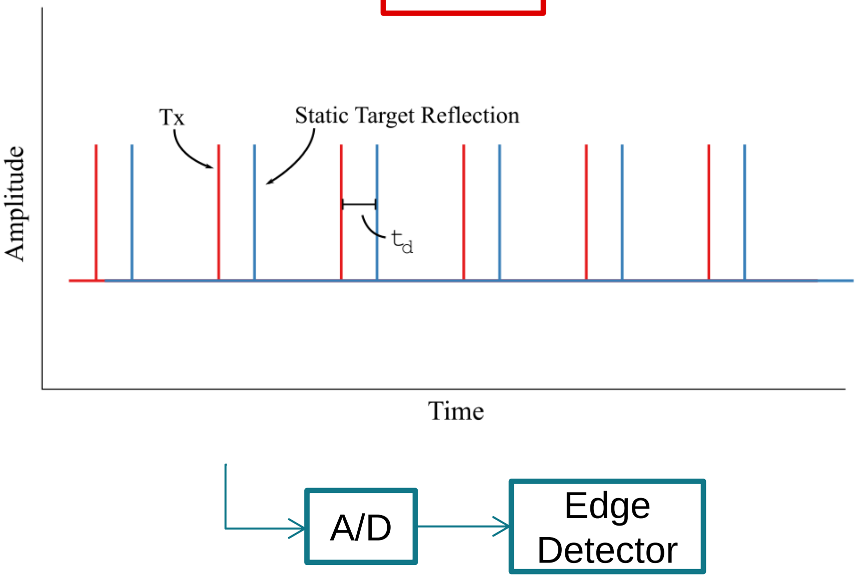


- Detect and classify coarse hand gestures using RADAR
  - Compact, low power
  - All-lighting, all-weather performance
  - NLOS capable operation
- Targeting distinct hand gestures involving motion

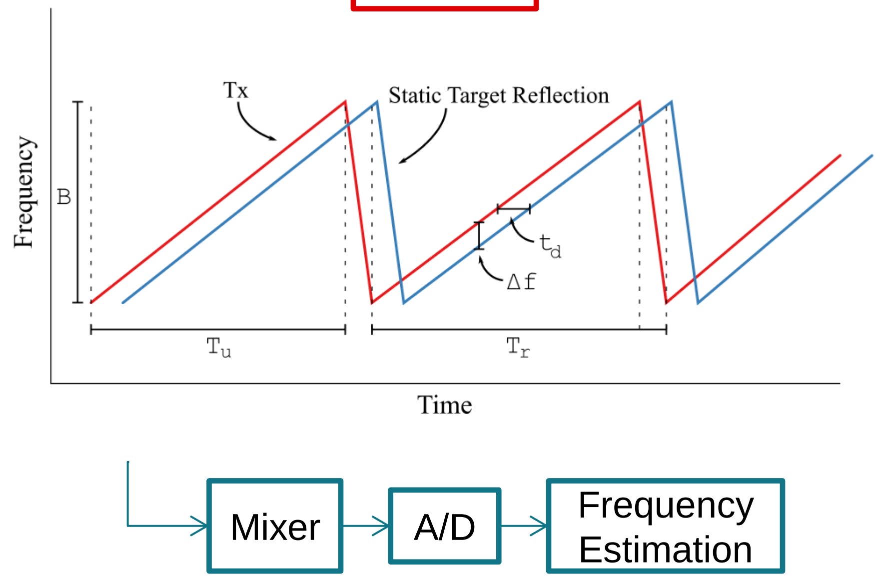
Types of Radar

- Two broad categories: pulsed and continuous wave (CW)
  - Pulsed radar directly measures time-of-flight to compute distance
  - CW radar indirectly measures distance through phase or frequency changes

Pulsed



FMCW



FMCW Radar

- Single chirp

$$s(t) = \exp(j2\pi f_c t + j\pi B T_r t^2)$$

- Down-converted response of a **point target** with moving with constant velocity v starting at some range R_0.

$$x(n,m) = \Gamma \exp\left(j\frac{4\pi}{c_0}\left(\frac{BR_0}{T_r}nT_s + vf_c mT_r\right)\right)$$

  - For nth sample in mth chirp

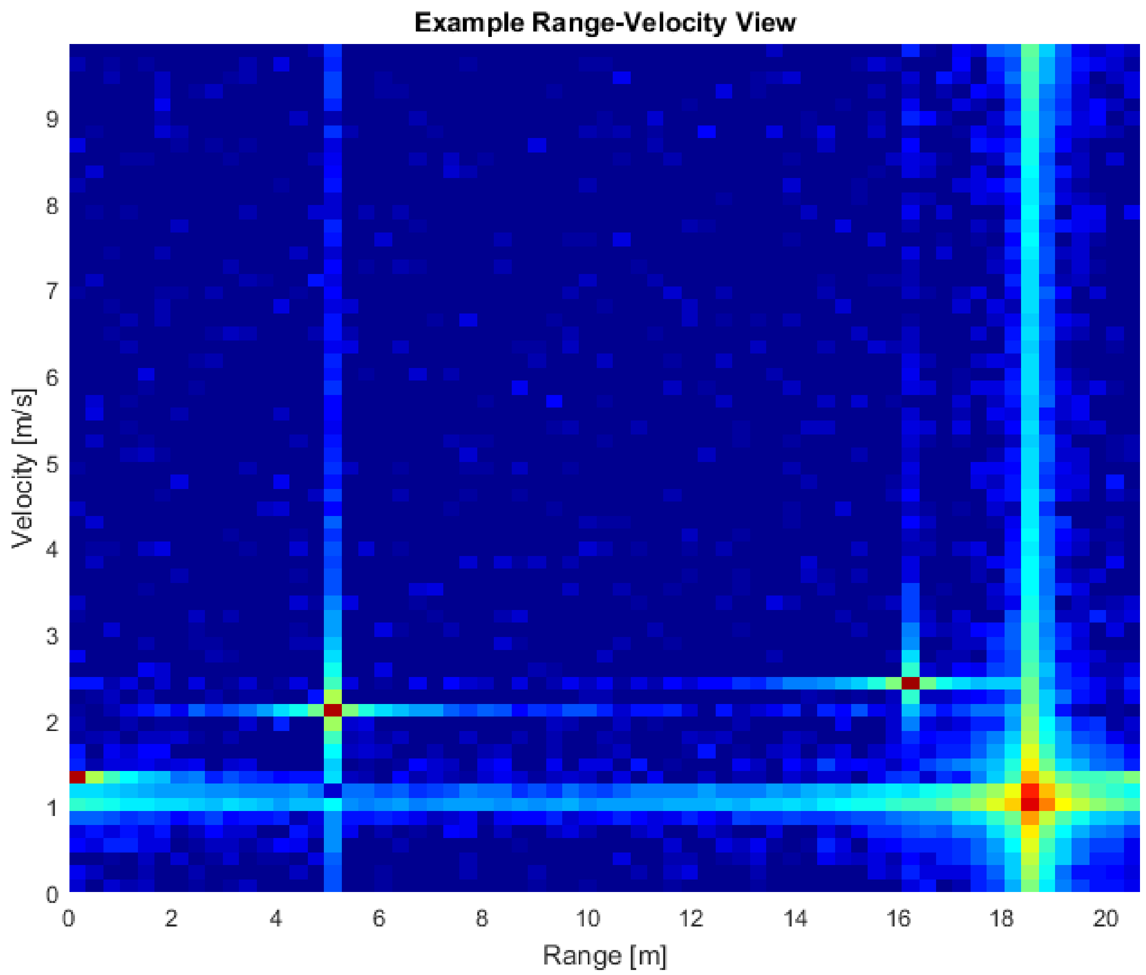
- Describe a group of point targets via 2D DFT:

$$X(k,w) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} x(n,m)\exp\left(j2\pi\frac{nk}{N}\right)\exp\left(j2\pi\frac{mw}{M}\right)$$

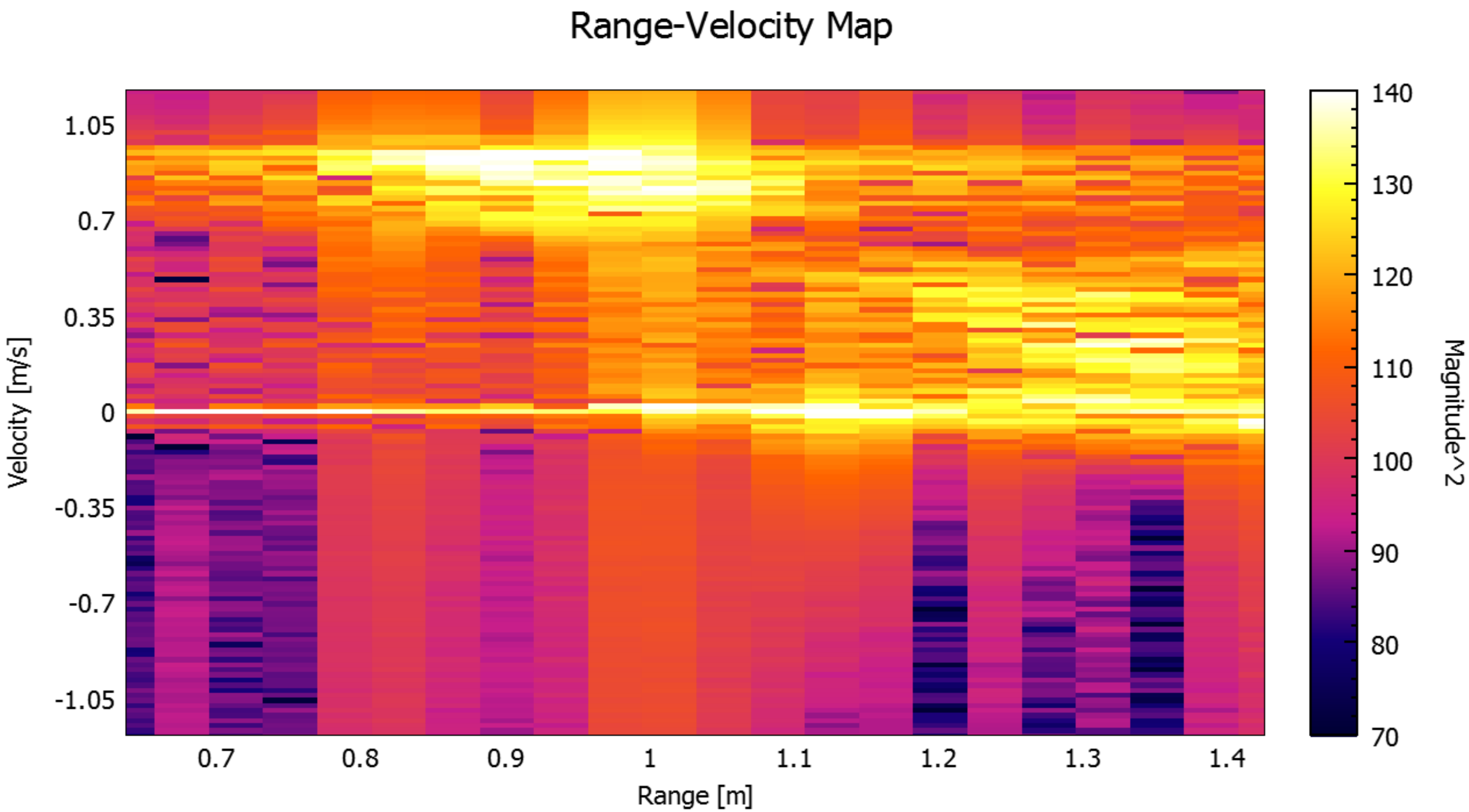
- Resulting pixel resolution

$$\Delta R = \frac{c_0}{2B} \qquad \Delta v = \frac{c_0}{2f_c M T_r}$$

Model Signature



Real Signature

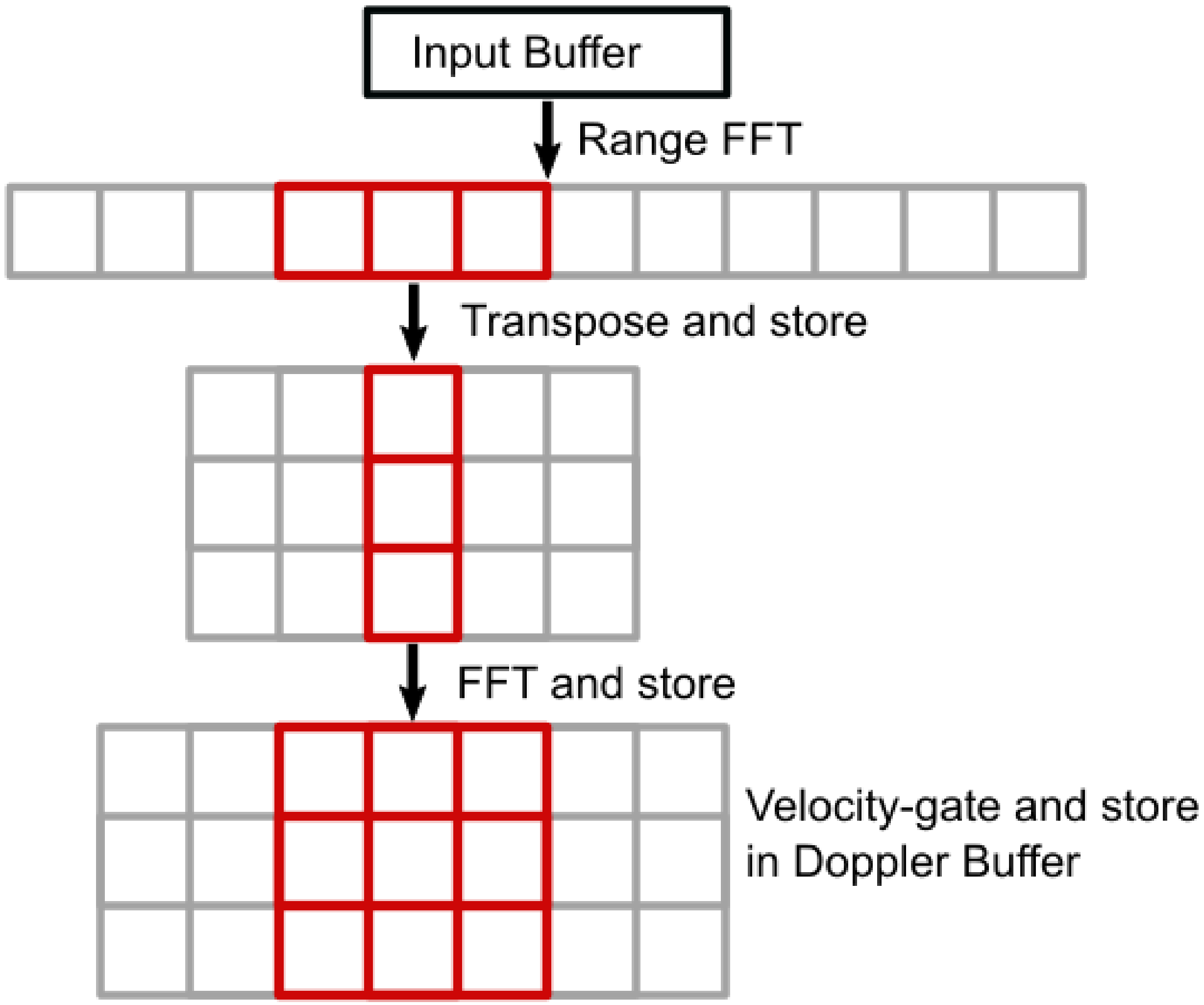


Micro-Doppler

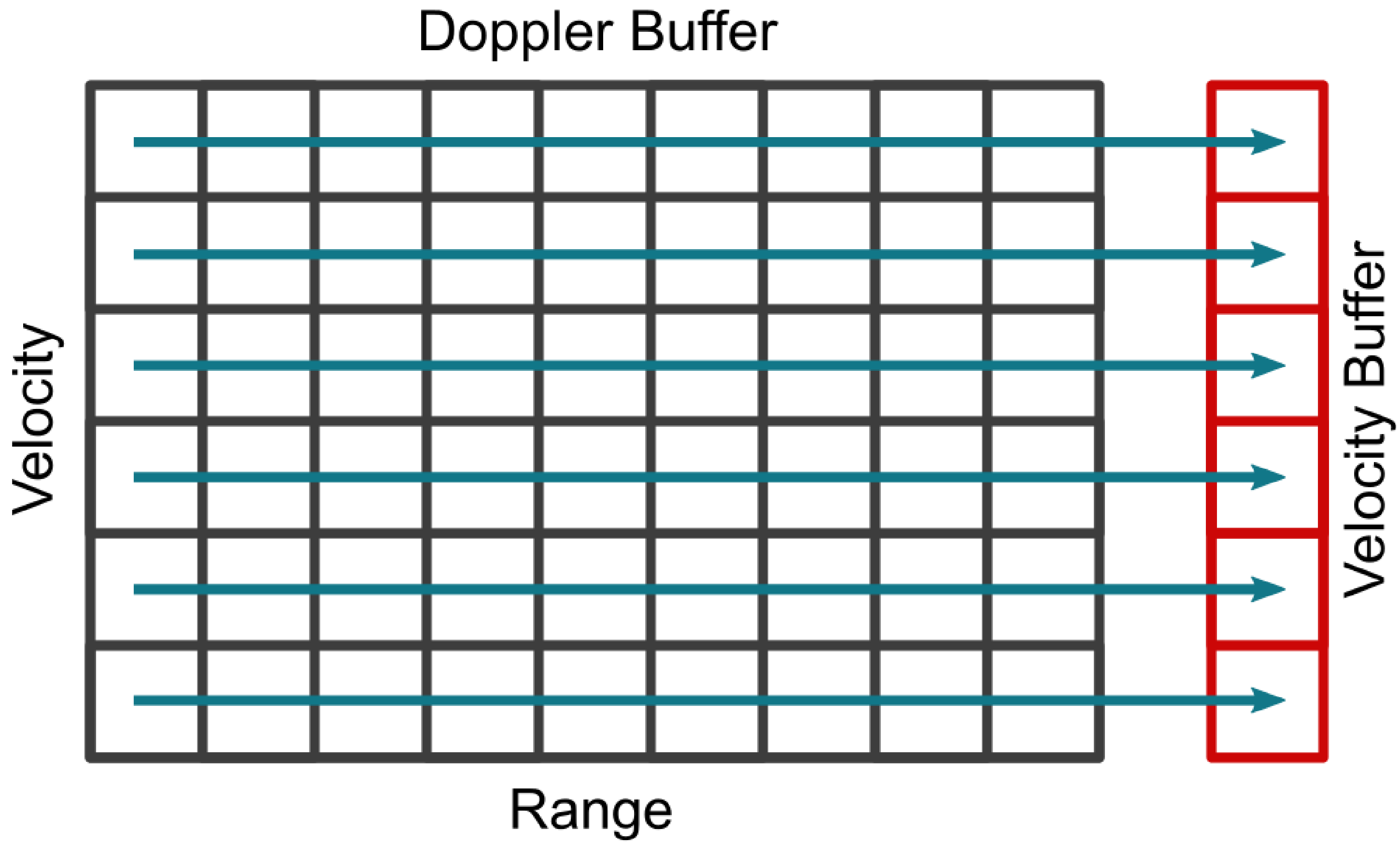
- Reality complicates matters
  - **Surface** composed of point targets
  - Multiple reflections, continuous internal scattering
  - Time-varying configuration, spatially diverse parameters

- Most relevant effect: micro-Doppler
  - Velocity signature associated with composite targets experiencing non-uniform motion
  - Difficult to characterize for all but the most trivial of configurations

Processing Step 1

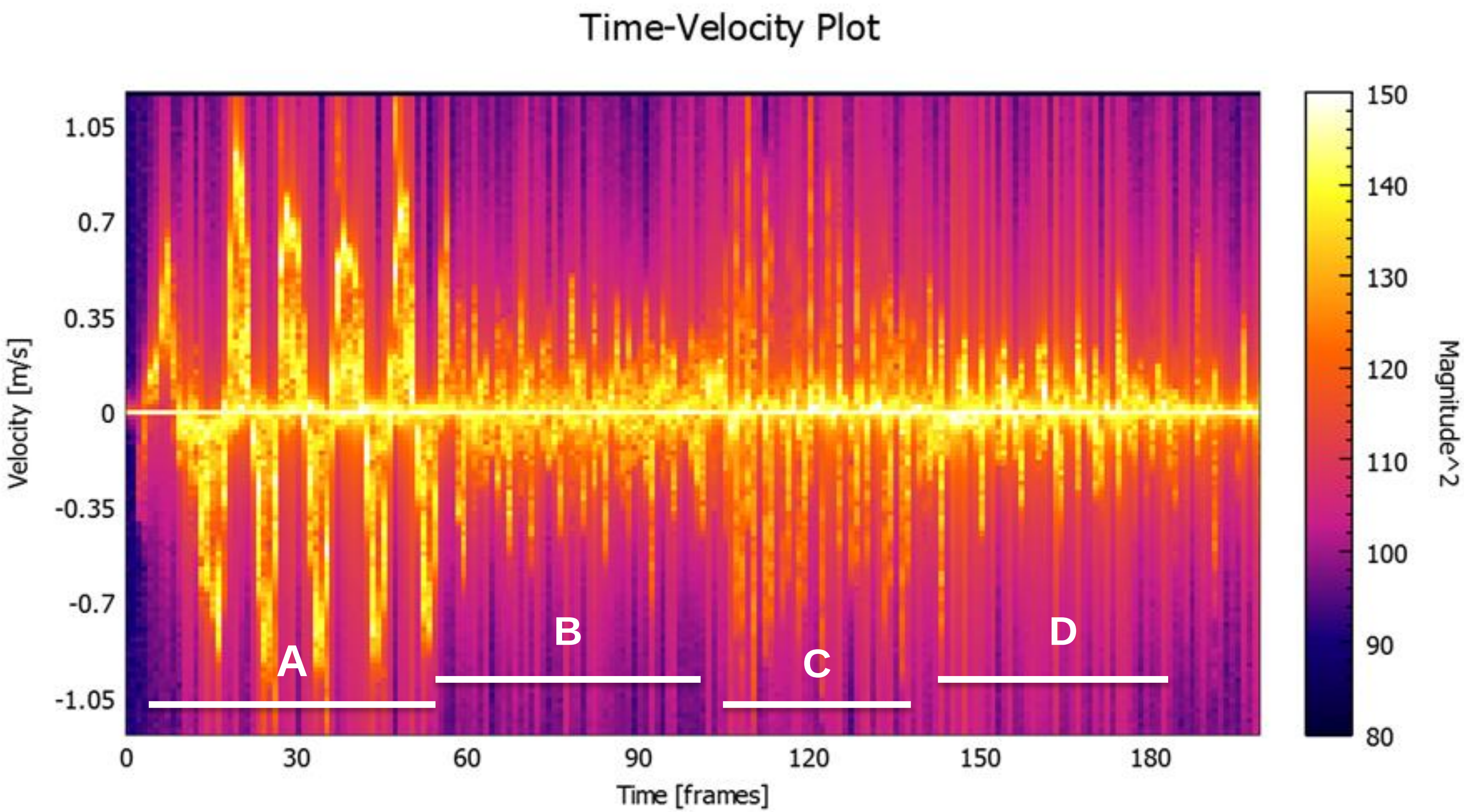


Processing Step 2

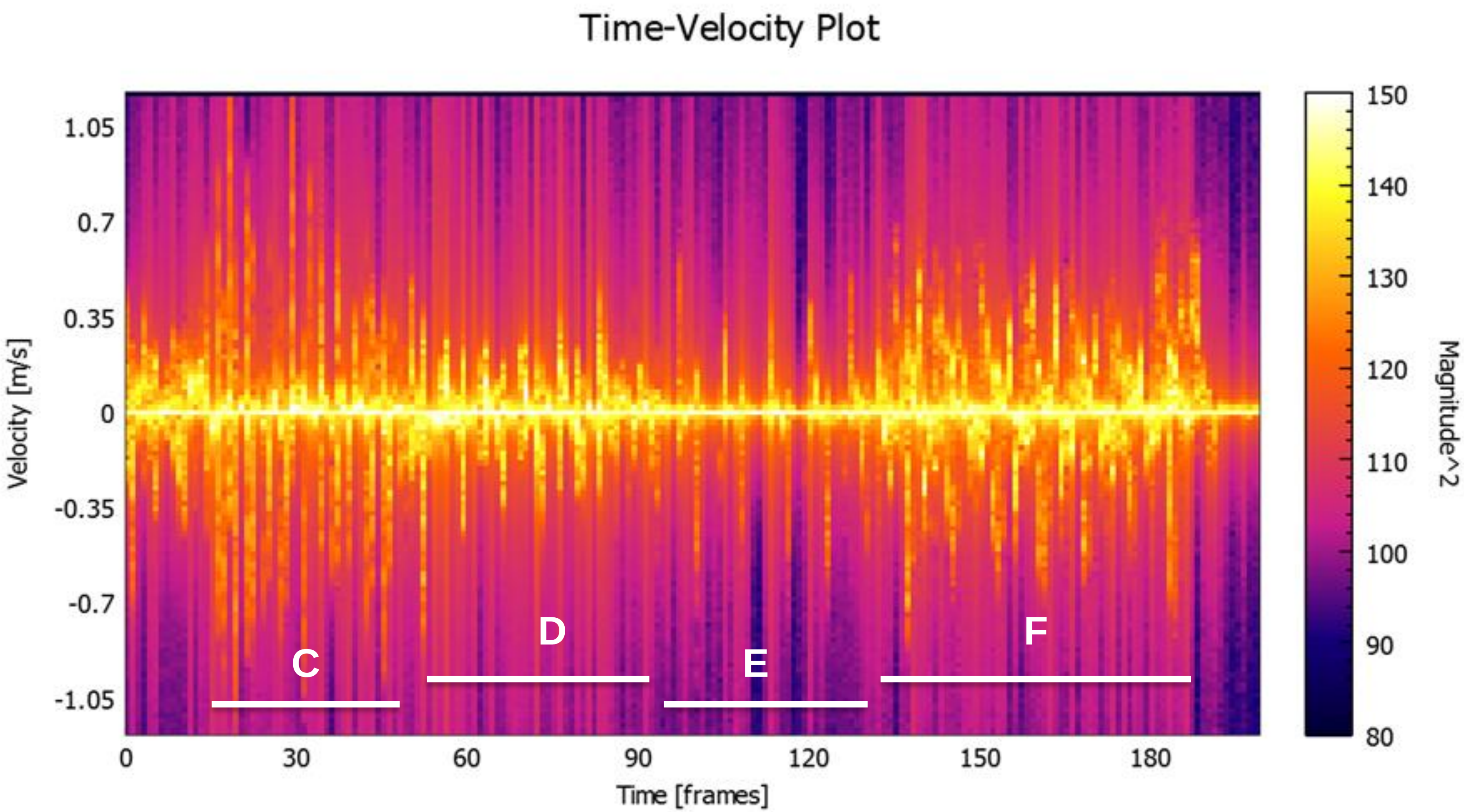


Resulting data forms “Velocity-Energy” measurement for recognition

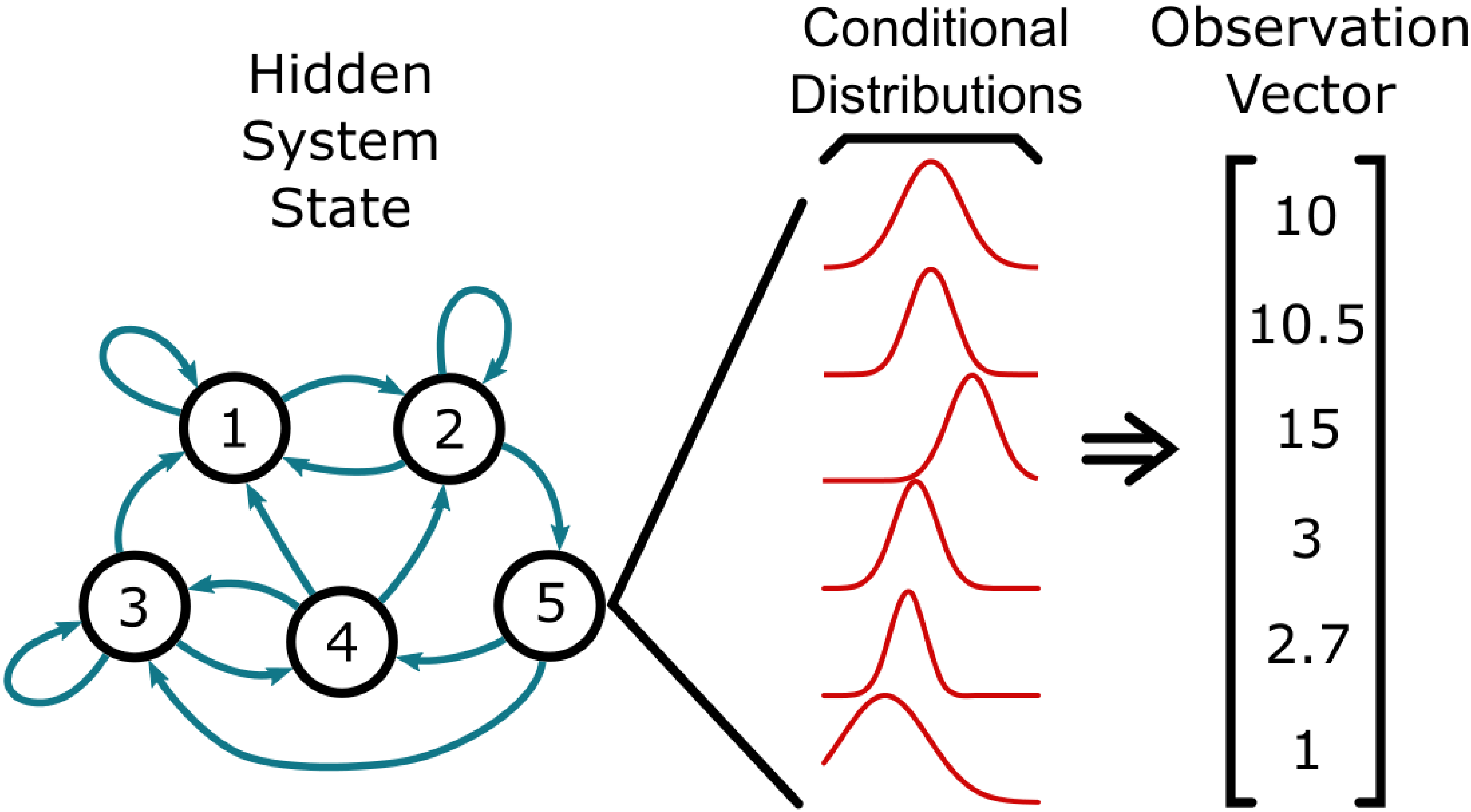
Gesture Dictionary



Gesture Dictionary (2)



Hidden Markov Models



Data Collection Parameters

- 4 GHz bandwidth starting at 76 GHz (3.75 cm range resolution)
- 1022 chirps with ~100μs chirp interval per **frame** (0.1 sec)
  - Plus ~0.2 sec dead time between frames
- 128-point velocity vector, one for each bin (~2 cm/s velocity resolution)
- 30, 15, or 10 frame “chunks”
  - Arbitrary segmentation used for both training and classification
  - Uninformed segmentation, no “markers” or silence detection
  - ~3,000 chunks per gesture collected
- Markov Model parameters
  - 5 states for each gesture model
  - 1 or 2 gaussian mixtures for each state
  - Diagonal covariance matrix for observation vector (implies independence)
- Optional feature extraction/lossy compression algorithm

Feature Extraction

- 128-point observation vectors (one velocity-energy buffer)
- Lower dimension observation vector is cheaper to classify
  - Good results if $I\left(Y; f(Y)\right)$ is suitably large
- Due to PSF, FFT windowing, and hand structure, V-E vectors have spatial correlation
  - V-E vectors might be seen as a sum of Gaussian functions,

$$f[n] = \sum_{l=1}^{L} w_l \exp\left(-\frac{(n-\mu_l)^2}{\sigma_l^2}\right)$$

$$v[n] = \eta[n] + f[n]$$

  - Then, the feature vector $\vec{z}$ is found as the parameters $\{w_l, \mu_l, \sigma_l\}$ that minimize the L2 distance between $\vec{f}(\vec{z})$ and the observation vector,

$$e^2 = \left(\vec{v} - \vec{f}\right)^T \left(\vec{v} - \vec{f}\right)$$

Classification Results

128-point vectors

- 30 frames
  - 1 GMM: 82.3%
  - 2 GMMs: 87.9%
- 15 frames:
  - 1 GMM: 81.5%
  - 2 GMMs: 84.8%
- 10 frames:
  - 1 GMM: 79.0%
  - 2 GMMs: 83.1%

10-point vectors

- 30 frames
  - 1 GMM: 83.3%
  - 2 GMMs: 88.9%
- 15 frames:
  - 1 GMM: 82.4%
  - 2 GMMS: 87.5%
- 10 frames:
  - 1 GMM: 75.85%
  - 2 GMMs: 81.7%

Conclusions

- Radar-based gesture recognition is feasible
- Learn from speech recognition
  - Previous work in activity classification focused on classification without modeling
  - Gestures can be explained using Hidden Markov Models
- Future work
  - More gesture data collection
  - Informed HMM design for gestures
  - Dealing with variation (physical size, gesture speed, viewing angle)

Extra Slides

Micro-Doppler in FMCW

- Bjorklund, S.; Johansson, T.; Petersson, H., **Evaluation of a micro-Doppler classification method on mm-wave data**, *2012 IEEE Radar Conference (RADAR)*, pp.0934,0939, 7-11 May 2012

- Gait identification at 77 GHz (150 MHz BW)
- Uses STFT to compute micro-Doppler signature
- Compute “Cadence Velocity Diagram (CVD)” by Fourier Transform over time for each velocity
- Construct feature vector from M peaks in CVD and corresponding velocities
- Use SVM classifier to match against training set

Micro-Doppler Signatures

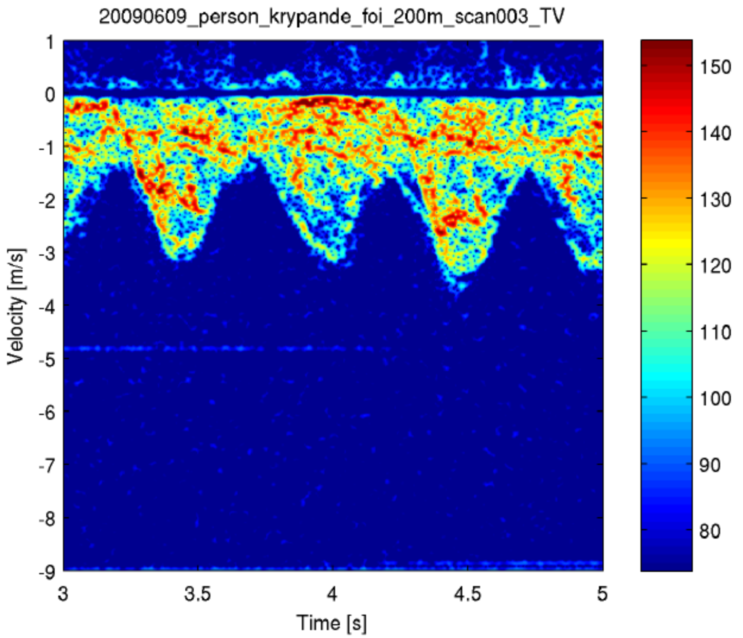


Fig. 2: A person creeping with negative Doppler. The 77 GHz SIRS 77 radar. The color shows power in dB. [runatr1019d2]

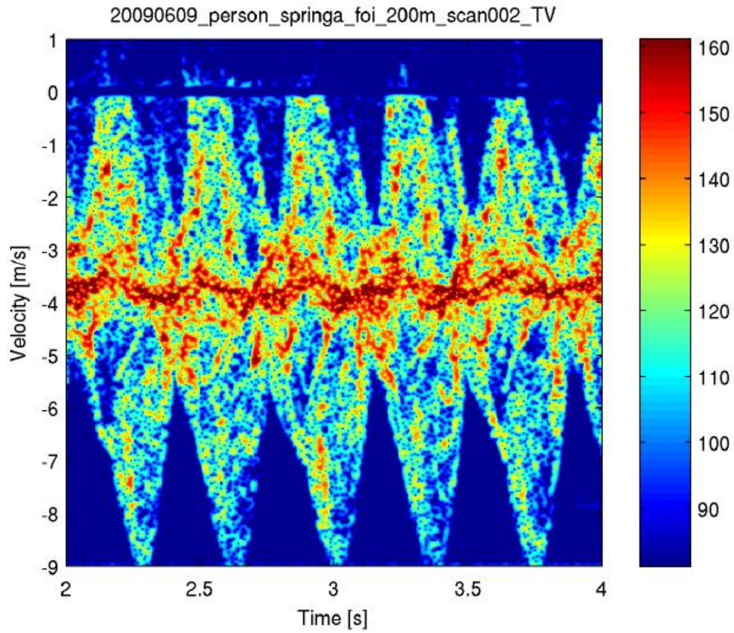


Fig. 4: A person running with negative Doppler. The 77 GHz SIRS 77 radar. The color shows power in dB. [runatr1019d2]

From "Evaluation of a micro-Doppler classification method on mm-Wave data"

Published Classification Results

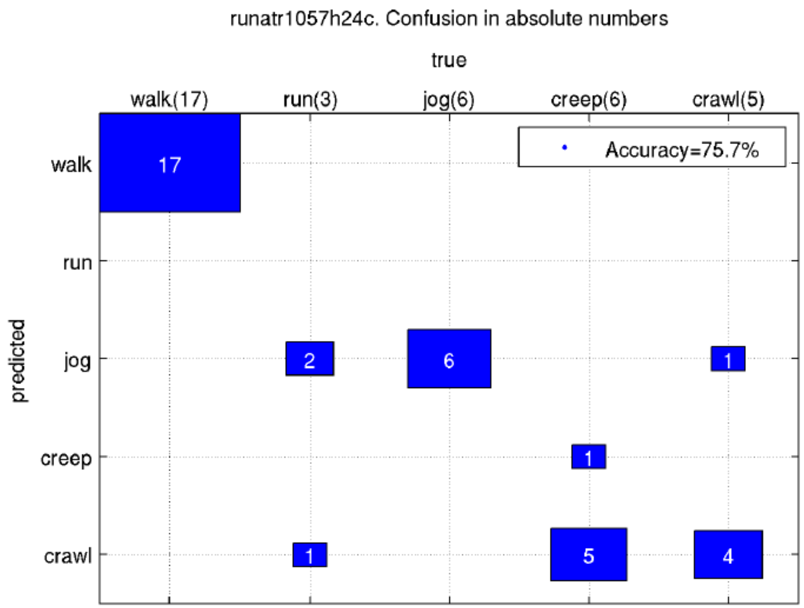


Fig. 10: Classification result (confusion matrix) with true and estimated classes. Number of training sequences: 41 (walk), 8 (run), 14 (jog), 14 (creep), 12 (crawl).

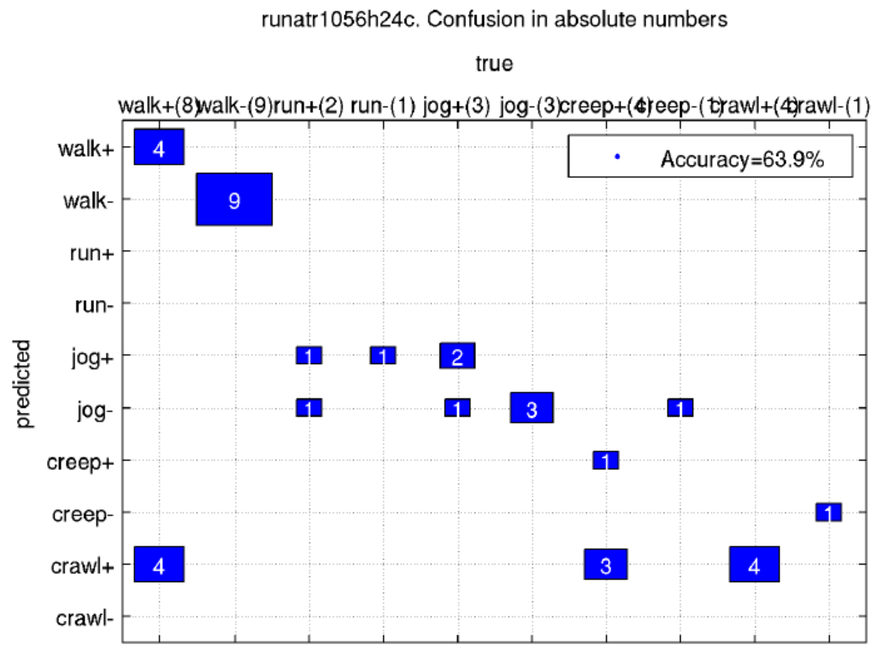


Fig. 11: Classification result (confusion matrix) with true and estimated classes. The labels above the plot should be "walk+(9) walk-(9) run+(2) run-(2) jog+(3) jog-(3) creep+(4) creep-(2) crawl+(4) crawl-(1)". Number of training sequences: 20, 21, 4, 4, 7, 7, 11, 4, 10, 2.

From "Evaluation of a micro-Doppler classification method on mm-Wave data"

Feature Extraction Algorithm

Find Largest Peak

↓

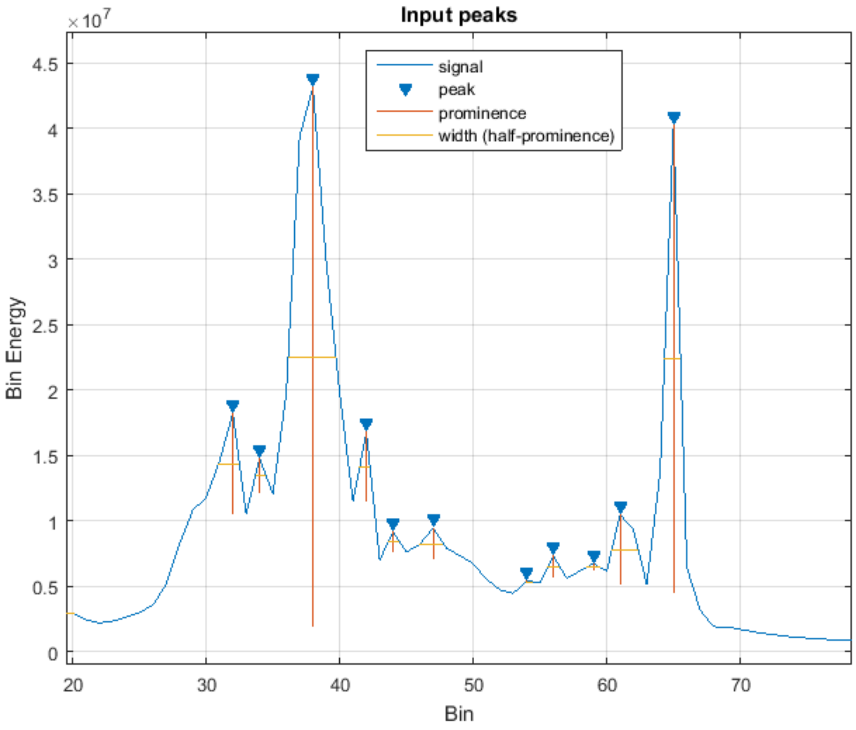
Newton Refinement for Matching

↓

Subtract Model Signal

↓

Repeat



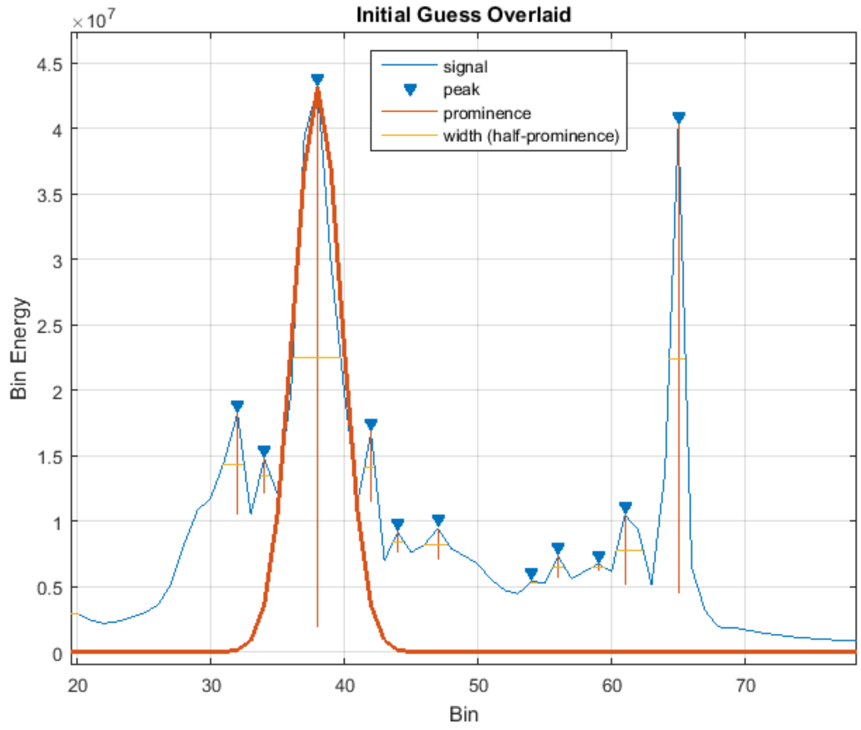
Feature Extraction Algorithm

Find Largest Peak

Newton Refinement for Matching

Subtract Model Signal

Repeat



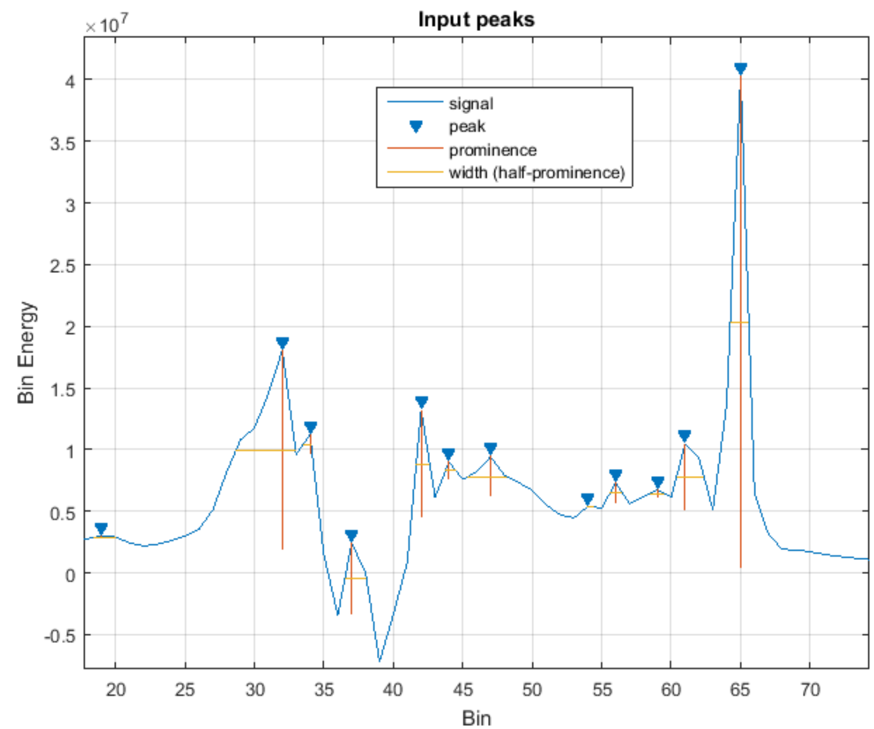
Feature Extraction Algorithm

Find Largest Peak

Newton Refinement for Matching

Subtract Model Signal

Repeat



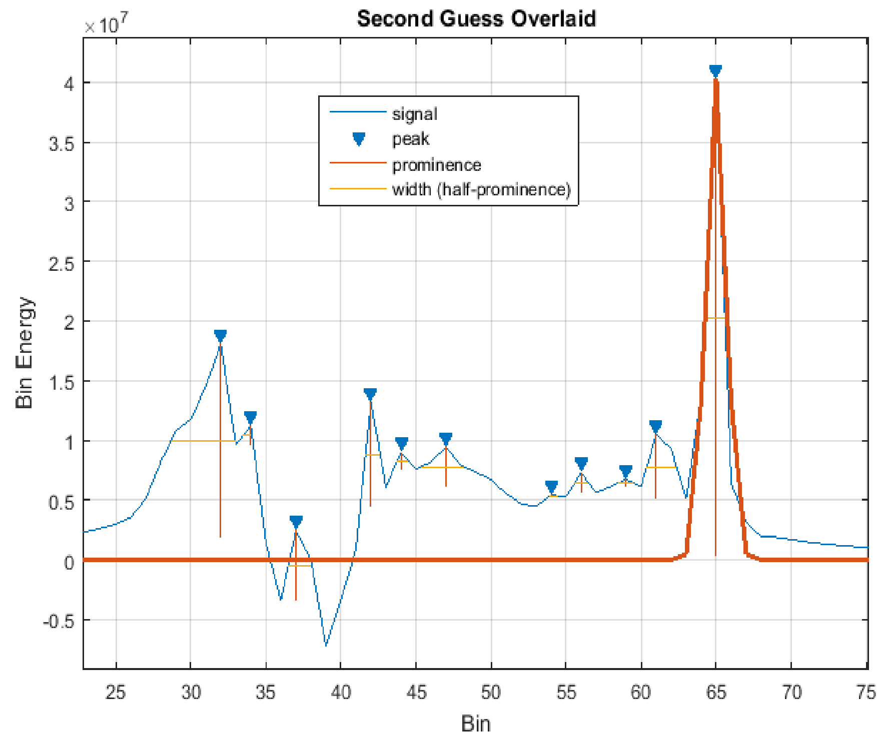
Feature Extraction Algorithm

Find Largest Peak

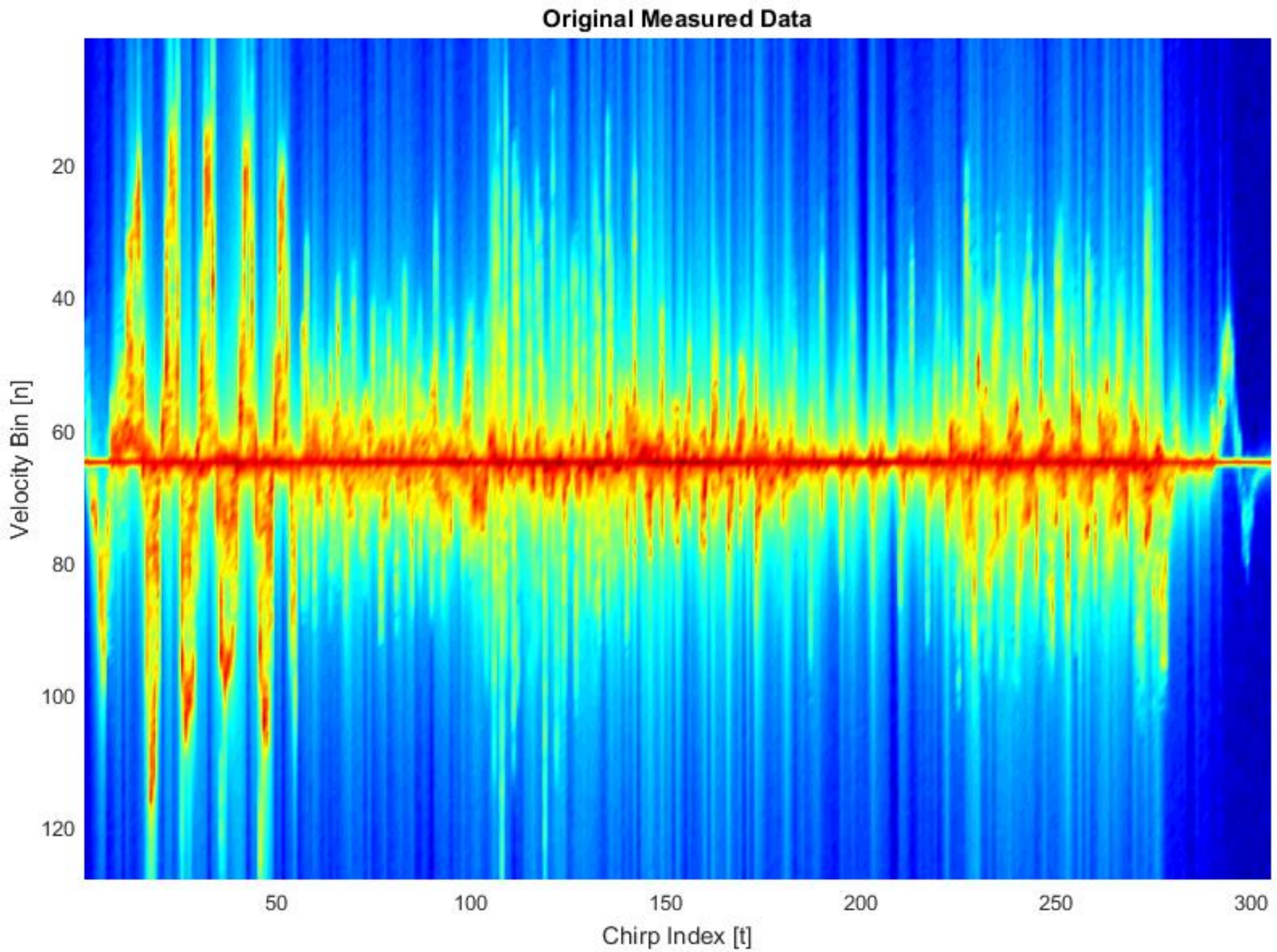
Newton Refinement for Matching

Subtract Model Signal

Repeat



Original V-E Plot



Reconstructed V-E Plot