Teach Me To Tango - Fidelity Estimation of Sensors using CNNs

STANFORD UNIVERSITY

Motivation and Problem Statement

Robust navigation in uncertain, cluttered environments is one of the major unsolved technical challenges in robotics. The goal of this project is to predict the error rate of a a black-box estimation system: the Google Tango, without knowledge of the estimation method that it employs.

These error rate estimates can then be used in risk-sensitive planning algorithms such as the approach described in [1]. Other related work involves estimating pose [2] and optical flow [3] using CNN/RNN architectures, as well as using feature and filter based methods [4].

[1] B. Ichter, B. Landry, E. Schmerling, and M. Pavone. Robust motion planning via perception-aware multiobjective search on GPUs. Sept. 2017. Submitted.

[2] R. Clark, S. Wang, H. Wen, A. Markham, and N. Trigoni. Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem. In *Conference on Artificial Intelligence*. AAAI, 2016.

[3] P. Fischer, A. Dosovitskiy, E. Ilg, P. Ha¨usser, C. Hazırbas¸, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox. Flownet: Learning optical flow with convolutional networks. *arXiv preprint arXiv:1504.06852*, 2015.

[4] J. Gui, D. Gu, S. Wang, and H. Hu. A review of visual in- ertial odometry from filtering and optimisation perspectives. *Advanced Robotics*, 29(20):1289–1301, 2015.

Dataset and Preprocessing

Images: 16349 images taken with a Google Tango aboard a quadcopter.

IMU Data: Linear and angular rate data collected from a Pixhawk.

Objective: Predict rate of change in the error of the Tango pose estimate

Preprocessing steps:

Transform coordinates

⇩

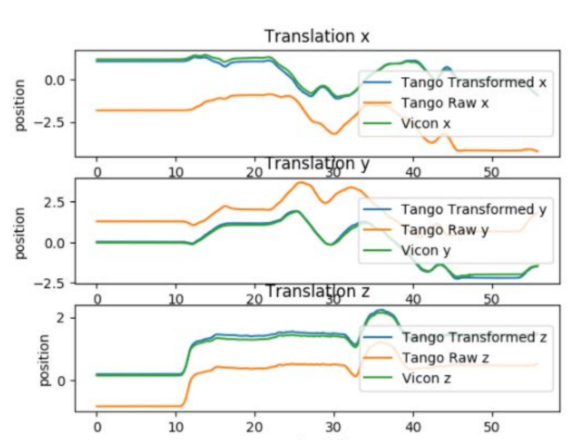
Downsample images

⇩

Compute optical flow (cv2)

⇩

Normalize inputs.



Transformed Pose vs. Raw Pose

Training: 60% training and 40% validation split

Metrics

Quantitative

Correlation:

$$(f \star g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m]\, g[m+n].$$

RMSE:

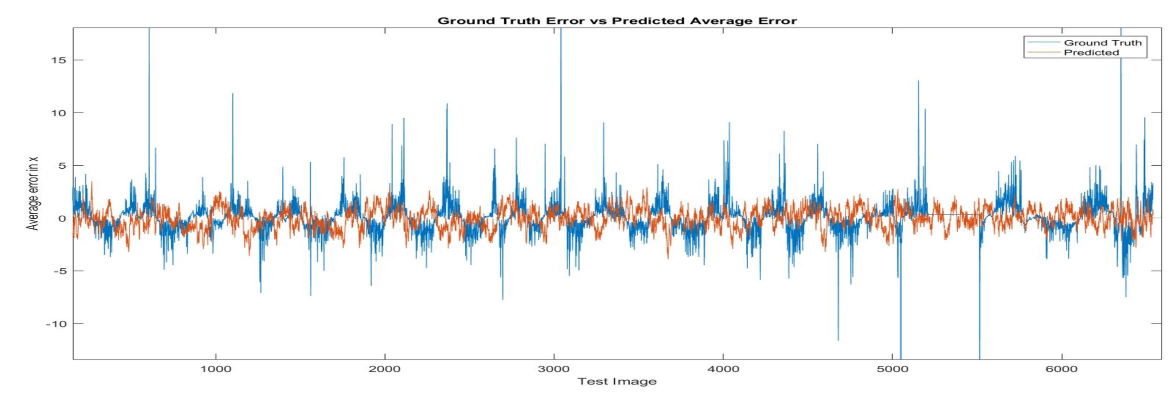
$$\sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$$

Qualitative

Tracking: How well the prediction seem to follow the true signal

Traditional Baseline

SVM Regression

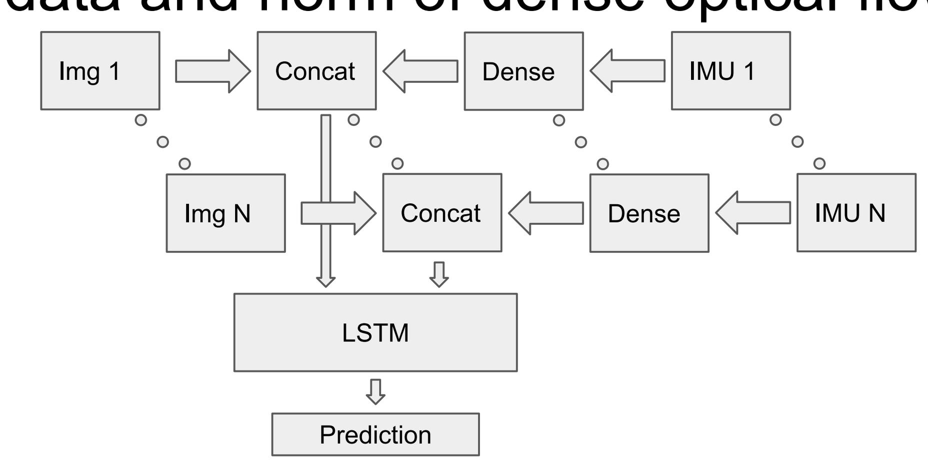


- Classical image processing features: (LBP, HOG, BRISK, SURF)
- SVM classifier for continuous predictions
- RMSE error for error rate in x - 2.1620
- Guess zero all the time - gives zero correlation

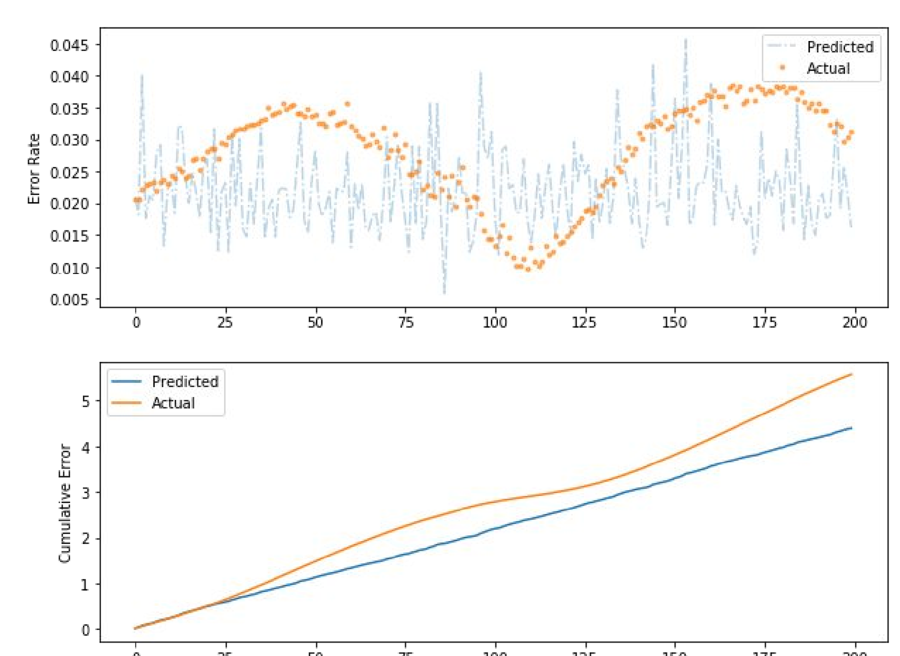
RNN Approach

Single-layer RNN with LSTM gates

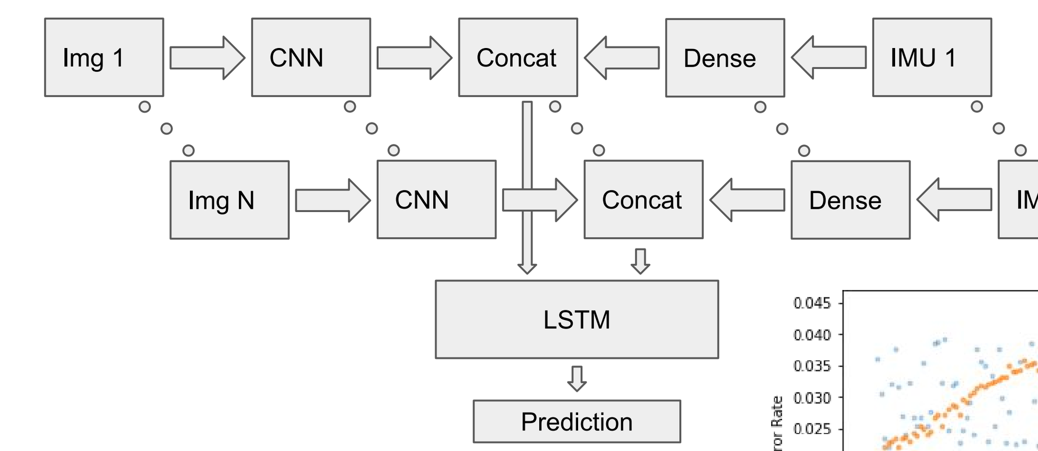
Input: IMU data and norm of dense optical flow



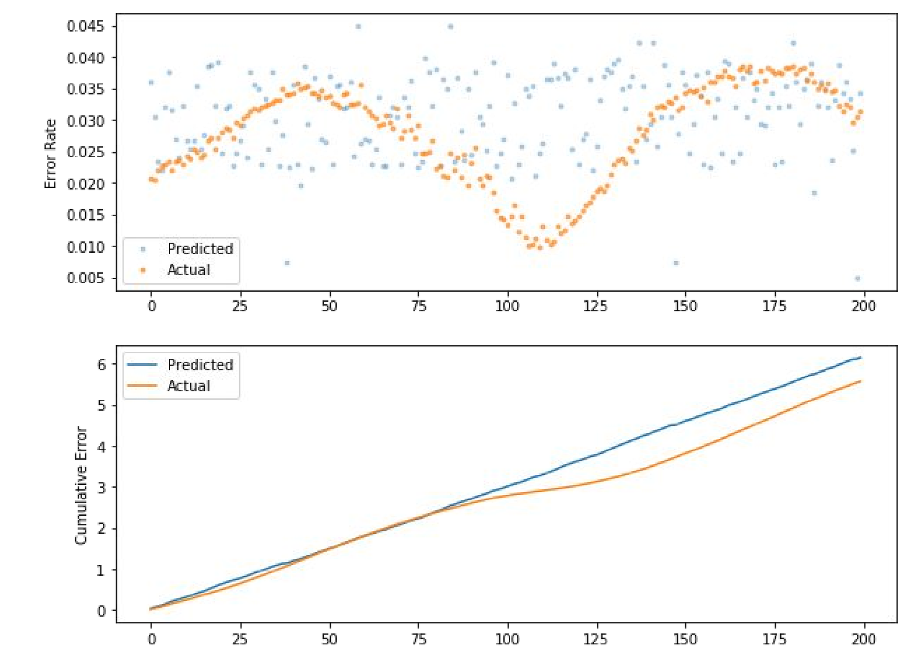
- Predict continuous value
- Correlation: **0.03**
- Tracks cumulative error poorly



CNN-RNN Approach

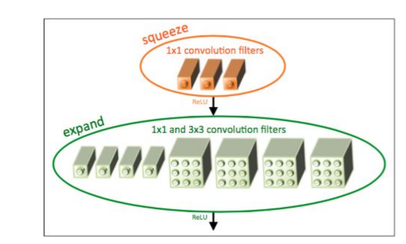


- Correlation: **0.12**
- Tracks cumulative error
- Very difficult to train

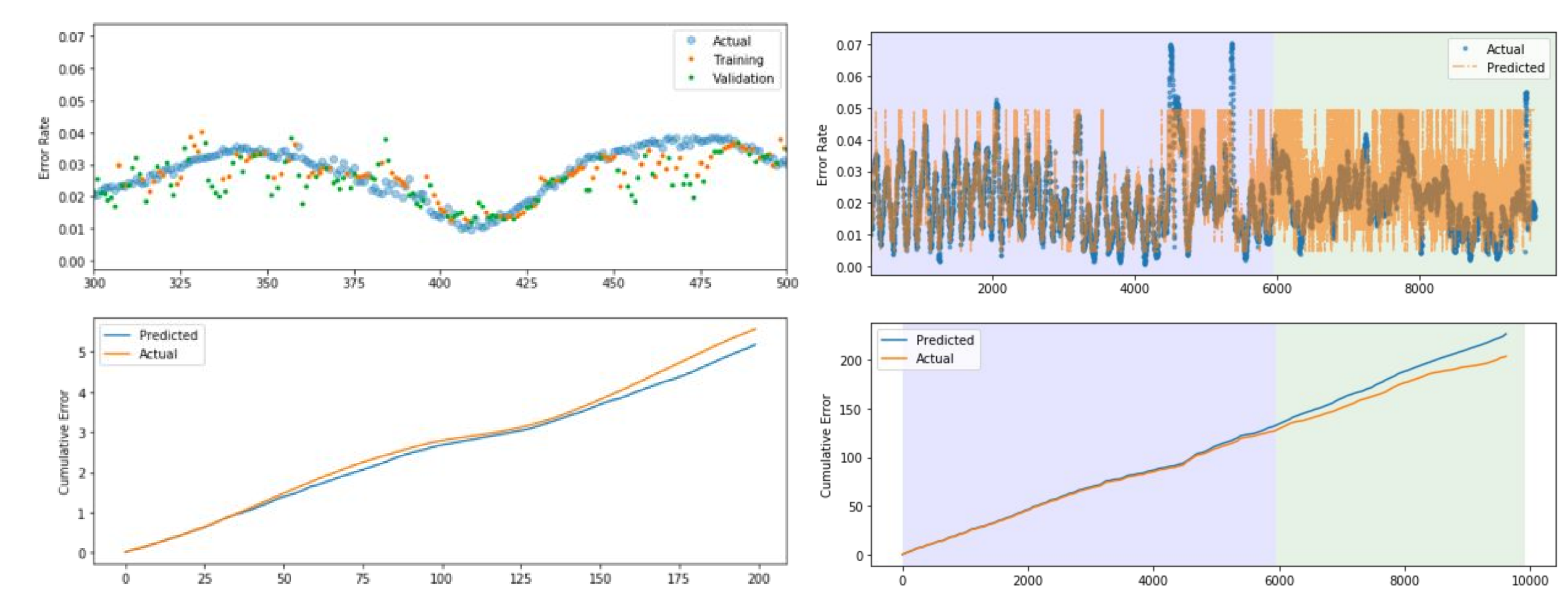


CNN Approach

- Classify discretized value
- Correlation: **0.64**
- Tracks cumulative error



Squeezenet Micro-architechture. Iandola et al., arXiv:1602.07360 [cs.CV]



Conclusions and Future Work

Conclusions:

- Baselines underperform
- Squeezenet performs best (so far)
- Recursive network difficult to train

Future Work:

- Better recursive network training
- Deeper recursive network architecture
- More data
- Direct image input