

Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images

Jamie Shotton Ben Glocker Christopher Zach Shahram Izadi Antonio Criminisi Andrew Fitzgibbon
Microsoft Research, Cambridge, UK

Supplementary Material
CVPR 2013

Overview

This document provides additional qualitative and quantitative evaluation of our proposed camera relocalization approach. We show several detailed illustrations of the error distributions, statistics, and inferred camera poses in comparison to the sparse RGB baseline (see main paper for description). All results shown for our approach use Scene Coordinate Regression Forests (SCoRe Forests) using DA-RGB features, without the frame-to-frame tracking or ICP refinement extensions.

Further results on sequences are shown in the accompanying video, which compares single frame pose estimation to frame-to-frame tracking and ICP refinement.

Camera Tracks

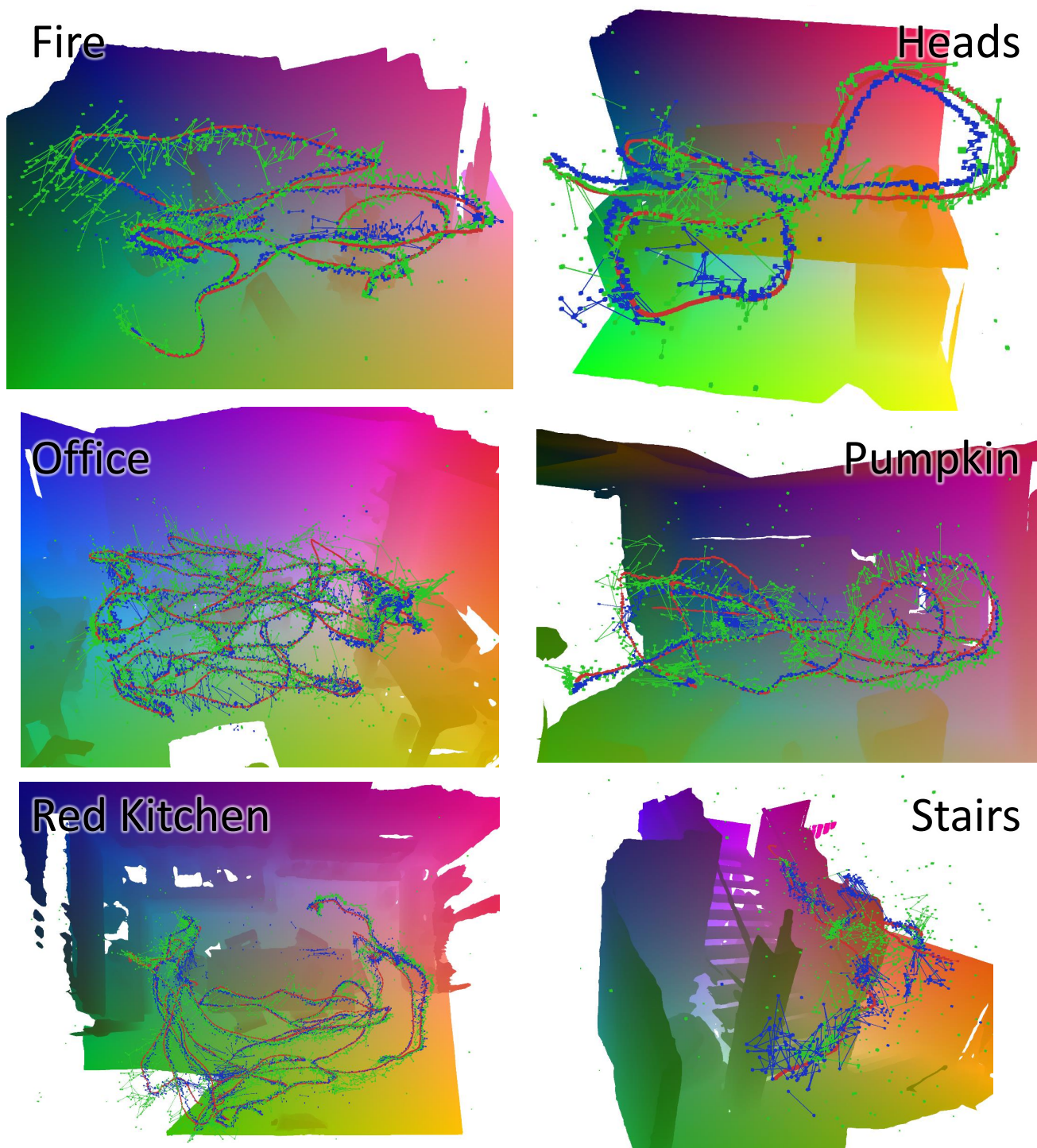


Figure 1. **Camera tracks.** Red: ground truth. Blue: SCoRe Forest result. Green: sparse baseline result. Our algorithm in many cases gives remarkably smooth camera tracks from single frames at a time. The result for Chess appears in the main paper. Best viewed digitally at high zoom. See also the quantitative comparisons below.

Summary of Error Statistics

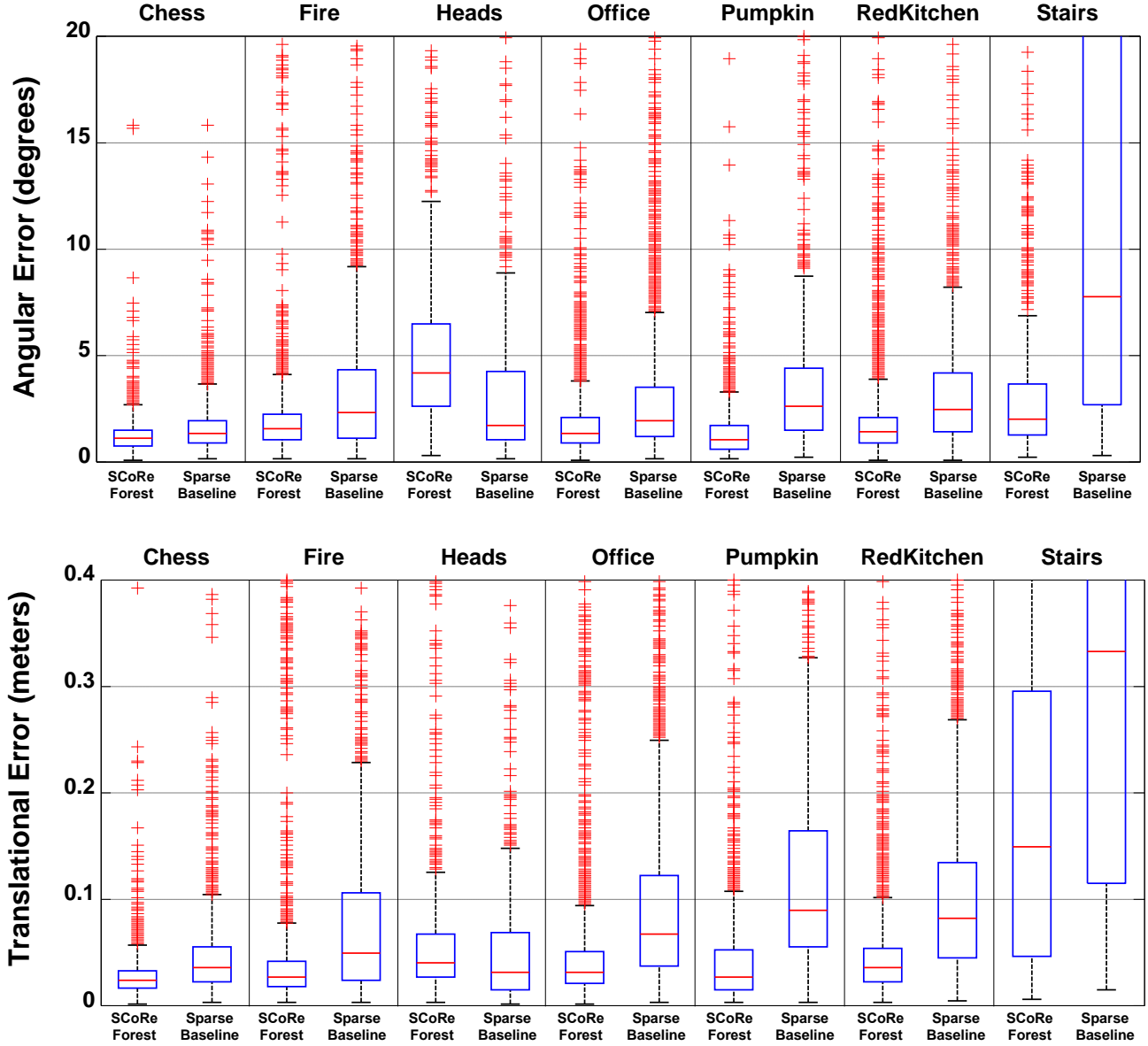


Figure 2. **Box plots summarizing angular and translational error distributions.** Red lines indicate the median. The edges of the blue boxes show the 1st and 3rd quartiles (Q_1 , Q_3). The lower and upper fences extend to the most extreme points not considered outliers (lower fence = $Q_1 - 1.5IQR$, upper fence = $Q_3 + 1.5IQR$, where $IQR = |Q_3 - Q_1|$ is the interquartile range). Outliers are plotted as red crosses. For each scene, we plot the errors for our SCoRe Forests (left box plot) in comparison to the sparse baseline (right box plot). Except for the Heads scene, our approach yields consistently lower angular and translational errors.

Error Distributions

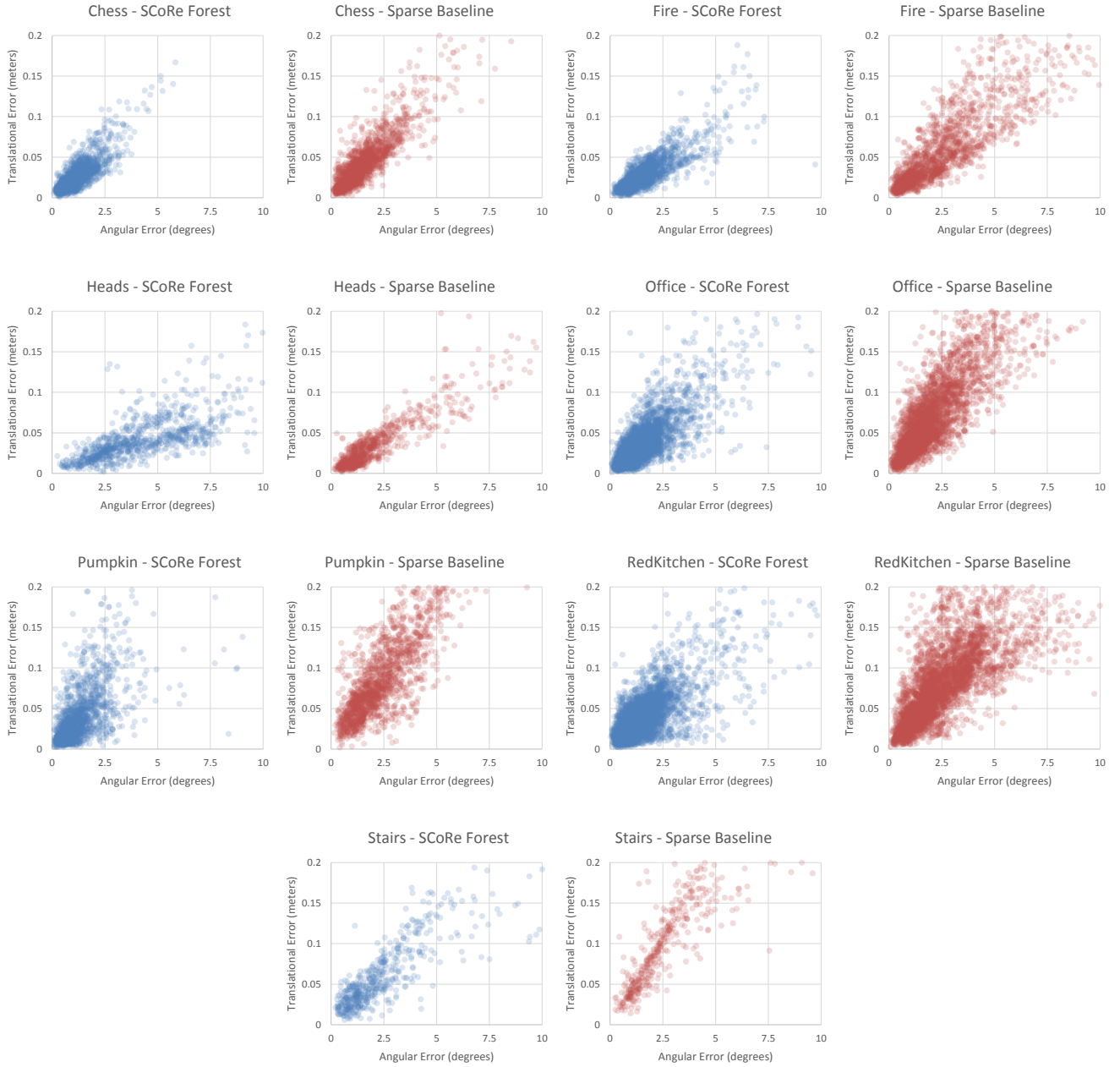


Figure 3. **Scatter plots of angular vs. translational errors.** Each dot indicates the errors for a single frame and pose. We compare the distributions of SCoRe Forests (blue plots) with the sparse baseline (red plots). Our approach yields errors which are mostly concentrated within the box of less than 2.5° and 5cm error, while the distributions for the baseline method are more spread out.

Errors on Camera Position

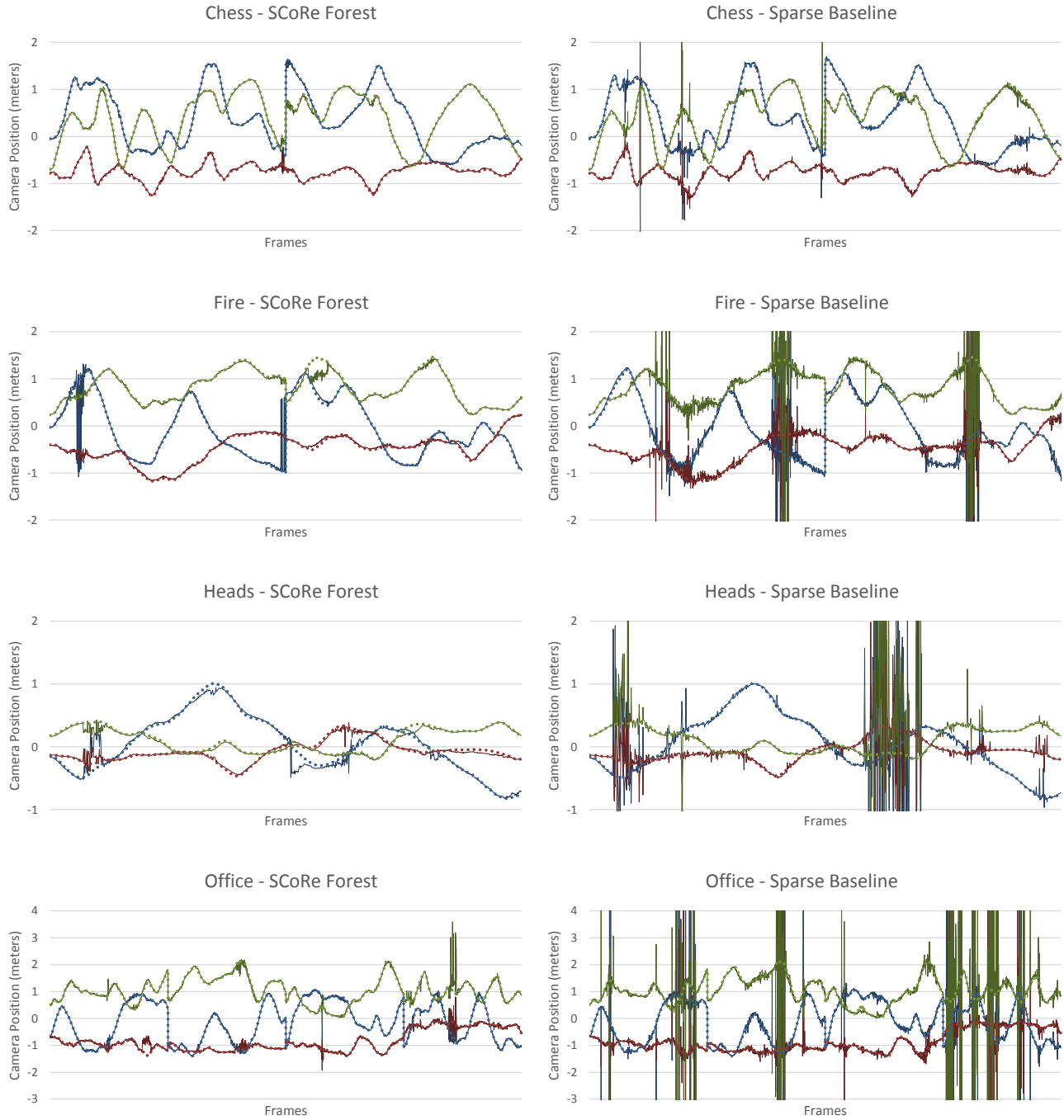


Figure 4. **Camera center coordinates for Chess, Fire, Heads, Office.** The dot curves show the ground truth camera center x-, y-, z- coordinates (blue, red, and green respectively) for all test frames. The thin, continuous curves show the inferred camera positions for SCoRe Forests (left) and the sparse baseline (right). We observe that our method produces smooth results despite the fact that no temporal information is used. Also, our method exhibits fewer outliers compared to the baseline. Best viewed digitally at high zoom.

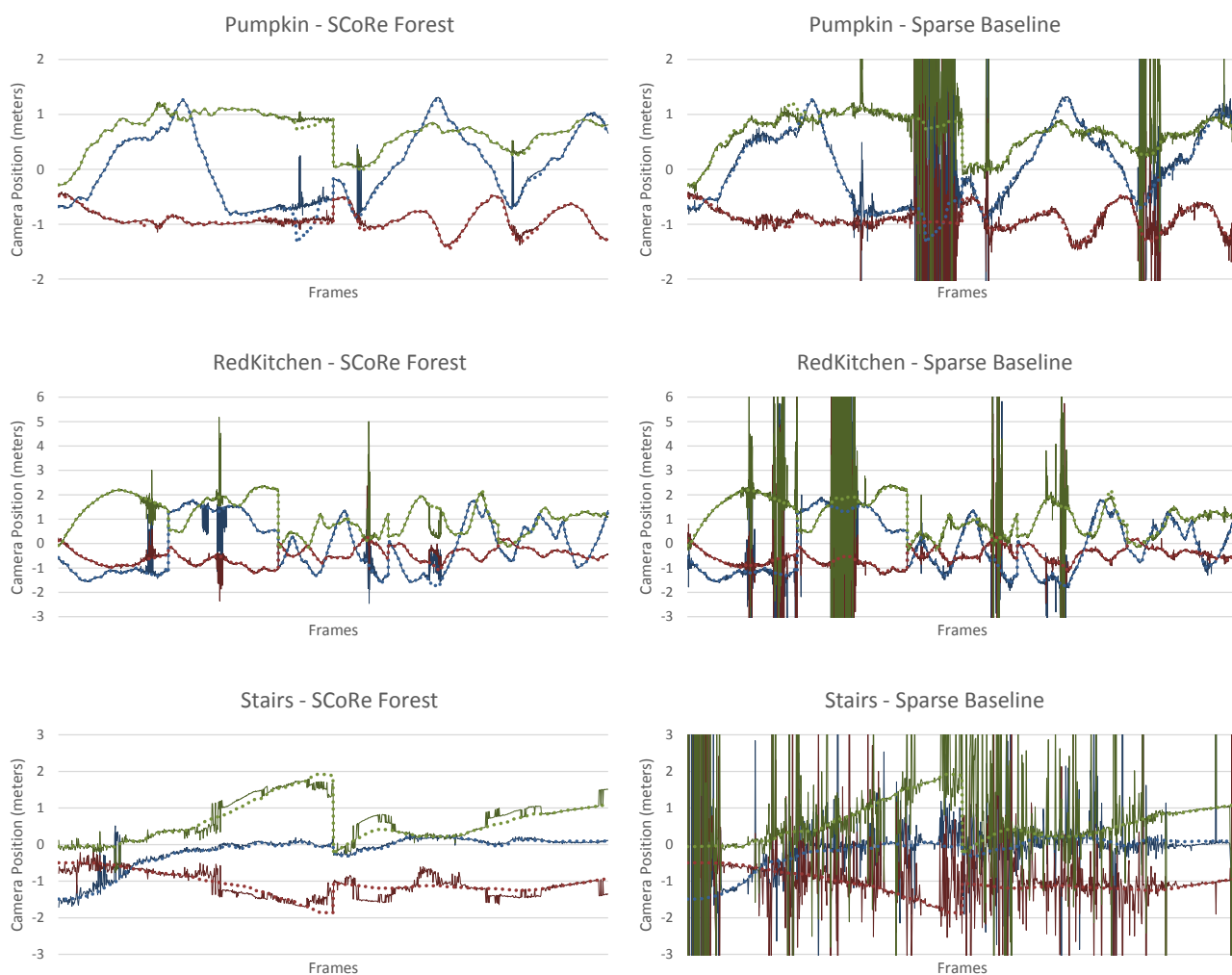


Figure 5. **Camera center coordinates for Pumpkin, RedKitchen, Stairs.** Comparison of SCoRe Forests (left) and sparse baseline (right). See Fig. 4 caption for more details. In the Stairs result, many of our errors are due to ambiguous matches to different parts of the stairs.