Exploiting Random RGB and Sparse Features for Camera Pose Estimation

Lili Meng, Jianhui Chen, Frederick Tung,
James J. Little, Clarence W. de Silva
lilimeng@mech.ubc.ca

University of British Columbia
Vancouver, Canada

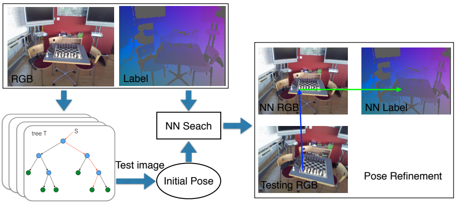

Figure 1: Pipeline. Our method first trains a regression forest using random RGB features. At test time, the camera pose is initially estimated by the random forest predictions, and then is refined by sparse feature matching.

1 Overview

We extend recent advances in scene coordinate regression forests [2] for camera relocalization in RGB-D images to use RGB features, enabling camera relocalization with only a single RGB image at test time. Furthermore, we integrate the random RGB features and sparse feature matching in an efficient and accurate way, broadening the method for fast sports camera calibration in highly dynamic scenes.

2 Methodology

Fig.1 shows the pipeline of our method. During training, the scene information is encoded in a random forest. At test time, an initial camera pose is calculated using the random forest predictions with real-time response. A nearest neighbor (NN) image is queried using the initial camera pose. The camera pose is refined by sparse feature matching between the test image and the NN image. In our method, the labels can be any information associated with pixel locations.

The random RGB features We use features based on pairwise pixel comparison:

$$f_{\phi}(\mathbf{p}) = \mathtt{I}(\mathbf{p}, c_1) - \mathtt{I}(\mathbf{p} + \delta, c_2) \qquad (1)$$

where δ is a 2D offset and $\mathtt{I}(\mathbf{p}, c)$ indicates an RGB pixel lookup in channel c. Our feature does not require depth information, and so is suitable for large scale sports camera calibration.

Methods	SCRF[2]	PoseNet[1]	Ours
Train	RGB-D	RGB	RGB-D
Test	RGB-D	RGB	RGB
Avg. Err	0.08m,1.60°	0.44m,10.4°	0.17m,5.26°

Table 1: 7 *Scenes* results.

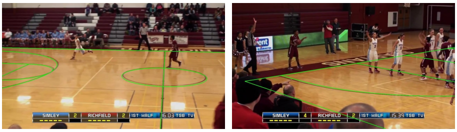

Figure 2: Sports camera calibration examples, best viewed in color. The court lines are overlaid on the images to indicate the accuracy of calibration.

Pose Refinement The 2D-3D correspondences are found by SIFT feature matching between the test image and the NN image. Then, the camera pose $\mathtt{P}$ is optimized by minimizing the reprojection error:

$$\mathtt{P}^* = \arg\min_{\mathtt{P}} \sum_k d(\mathbf{x}_k, \mathtt{P}\mathbf{X}_k)^2 \qquad (2)$$

where $\mathbf{x}$ are the feature locations in the image, and $\mathbf{X}$ are the correspondent 3D world coordinates associated with the NN image.

3 Evaluation

Our method is evaluated on the 7 *Scenes* dataset and a new basketball dataset using standard metrics. Table 1 shows quantitative results in 7 *Scenes* dataset. Fig.2 illustrates qualitative results in the basketball dataset. Experiment results demonstrate the efficacy of our approach, showing superior or on-par performance with the state of the art.

[1] Alex Kendall, Matthew Grimes, and Roberto Cipolla. PoseNet: A convolutional network for real-time 6-DOF camera relocalization. In *ICCV*, 2015.

[2] Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew Fitzgibbon. Scene coordinate regression forests for camera relocalization in RGB-D images. In *CVPR*, 2013.