Efficient Multi-Structure Robust Fitting with Incremental Top-$k$ Lists Comparison

Hoi Sim Wong, Tat-Jun Chin, Jin Yu and David Suter
School of Computer Science, The University of Adelaide, South Australia

Introduction

Multi-Structure Robust Fitting Issues:
1. Hypothesis sampling tends to be time-consuming for heavily contaminated data
   - previous methods depend on assumption or domain knowledge and fail to handle multi-structure data
2. Lack of a principled approach to fit multiple structures in data
   - previous methods based on sequential fitting or clustering often miss the true structures or find spurious structures

Overview

Data Similarity Matrix

Guided Sampling with Simultaneous Hypothesis Filtering

Fitting Result

Agglomerative Clustering on “Good” Hypotheses

Idea

- Multi-structure guided sampling by incrementally comparing top-k lists derived from residual sorting information
- Simultaneous hypothesis filtering such that only “good” hypotheses remains
- Agglomerative clustering on surviving hypotheses for accurate model selection

Proposed Method

Data Similarity by Comparing Top-k Lists:
- For each datum $z_i$, its top-$k$ list $\tau_i$ is obtained by sorting its residuals as measured to all hypotheses in ascending order
- To compare top-$k$ lists, we use the Spearman Footrule distance:
  $$F^{(\tau)}(\tau_i, \tau_j) = \sum_{m \in \Omega} |\tau_i^*(m) - \tau_j^*(m)|$$
  where $\tau_i^*(m) = \tau_j^*(m)$ if $m \in \Omega_i$; otherwise $\tau_i^*(m) = k + 1$.
- Data similarity matrix $K$ is obtained by comparing top-$k$ lists between all data
  $$K(i, j) = 1 - \frac{F^{(\tau)}(\tau_i, \tau_j)}{m}$$
- $K(i, j)$ is between 0 (dissimilar) and 1 (identical)
- To be efficient, $K$ is incrementally computed during sampling

Guided Sampling:
- Use data similarity matrix $K$ to sample data in a guided fashion
- To construct minimal subset, the first element is randomly selected from all data, then the remaining elements are conditionally sampled using rows of data similarity matrix $K$ as sampling weights

Simultaneous Hypothesis Filtering:
- Filter out “bad” hypotheses during sampling
- A set of “good” hypotheses is immediately available once sampling is done
- For hypothesis $m$, we construct a feature vector $f_m = [f_{\Omega m}^D, f_{\Omega m}^G]$
  - $\Omega_m$ = set of data that include the hypothesis $m$ in their top-$k$ lists ($\Omega_m$ should contain many indices from a structure if hypothesis $m$ is a good hypothesis)
  - $f_{\Omega m}^D = \text{average similarity values between all data in } \Omega_m$
  - $f_{\Omega m}^G = \text{reciprocal of average residual of data in } \Omega_m$ as measured to hypothesis $m$
- “Good” hypotheses should have high value in both $f_{\Omega m}^D$ and $f_{\Omega m}^G$
- Clustering the feature vector using k-means

Multi-Structure Fitting:
- Minimal subsets of “good” hypotheses should mainly contain inliers from different structures
- Simply use agglomerative clustering to cluster the minimal subsets
- Fitting geometric models to each cluster of data

Experiments

Proposed Method: ITKSF
1. Multiple 2D Line Fitting (time limit: 2 seconds)
2. Multiple 2D Circle Fitting (time limit: 2 seconds)
3. Homography Estimation (time limit: 15 seconds)
4. Fundamental Matrix Estimation (time limit: 30 seconds)