Image Analysis and Computer Vision

AdelaideRMF analysis and correction

Carlo Fabrizio Gabriele Giusti

Politecnico di Milano

Project presentation
July 22, 2024

Presentation Overview

1. Problem Formulation
   - Multi Model Fitting in general
   - AdelaideRMF dataset
   - Task Analysis - problems of the specific scenario
2. Outline and description of our solution approach
   - Change Point Detection
   - Performance indexes - Silhouette and Influence Function
   - Ensembling technique
3. Experimental results
   - Homography estimation
   - Fundamental Matrix estimation
4. Open problem - Soft Clustering internal validation measure

POLITECNICO
MILANO 1863

Problem Formulation

Multi Model Fitting problem formulation

Given a set of data $X = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$ possibly corrupted by noise and outliers, and given a family of parametric models Θ, the goal of multi model fitting is to automatically estimate:

- The models $\{\theta_1, \theta_2, \dots, \theta_K\} \subset \Theta$ that best explain the data
- The structures $\{U_1, U_2, \dots, U_K\} \subset \mathcal{P}(X)$ hidden in the data

POLITECNICO
MILANO 1863

AdelaideRMF dataset overview

- The AdelaideRMF dataset consists of a set of pairs of images, each pair contains a set of points correspondences P such that

$$P = (\bigcup_{i=1}^{M} U_i) \cup O \text{ where } U_i \cap U_j = \emptyset; \forall i, j \in \{1, ..., M\}, i \neq j$$

  Note that the dataset is not correct for instances of Multi Model Fitting, since the structures are disjoint sets, that is not the case in Multi Model Fitting scenarios.
- The dataset presents two different scenarios: Homography estimation & Fundamental matrix estimation.

POLITECNICO
MILANO 1863

Task Analysis

- The goals of the project are:
  - Find outliers among a set of point correspondences that were previously labeled as inliers.
  - Re-design the dataset so that it is in the form of soft clustering.
- The difference in the complexities of the two scenarios (Homography vs Fundamental matrix) forced us to develop two different specific approaches.

POLITECNICO
MILANO 1863

Task Analysis - Problems of the specific scenarios

- Homography case:
  1. Well masked outliers $\rightarrow$ Necessity of a robust approach.
  2. Absence of external validation measures $\rightarrow$ Necessity of a measure to validate our results.
- Fundamental Matrix case:
  1. All the problems of Homography affect also the Fundamental Matrix case.
  2. Adoption of the correct distance function $\rightarrow$ Sampson vs Symmetric Epipolar Distance
  3. Change Point Detection sensitivity $\rightarrow$ Robust estimators vs Shallow estimators.
  4. Fundamental matrix estimator $\rightarrow$ GC-RANSAC vs Ensemble.
  5. Insignificant outliers

POLITECNICO
MILANO 1863

Outline and description of our solution approach

General procedure - Homography

1. Apply LMEDS $\rightarrow$ LMEDS residuals.
2. Change Point Detection on LMEDS residuals $\rightarrow$ Inlier Thresholds.
3. GC-RANSAC with Inlier Thresholds $\rightarrow$ GC-RANSAC residuals & soft clustering & outliers.
4. Validate the outliers using the Influence Function.
5. New dataset creation.
    - re-fit GC-RANSAC on inliers from soft clustering and correct the outliers.

POLITECNICO
MILANO 1863

General procedure - Homography

1. Apply LMEDS $\rightarrow$ LMEDS residuals Sampson & LMEDS residuals Sed.
2. Change Point Detection
   - LMEDS residuals Sampson $\rightarrow$ Inlier Thresholds Sampson.
   - LMEDS residuals Sed $\rightarrow$ Inlier Thresholds Sed.
3. Ensemble models
   - Ensemble with Inlier Thresholds Samspon $\rightarrow$ Labels Sampson.
   - Ensemble with Inlier Thresholds Sed $\rightarrow$ Labels Sed.
   - Final Labels = Labels Sed $\cup$ Labels Sampson.
4. Validate the outliers using the Influence Function.
5. New dataset creation.
   - Eliminate the outliers only for the images for which it is worth it (Influence Function Increase Ratio above 30).

POLITECNICO
MILANO 1863

Change Point Detection

- General task: compute the data value corresponding to the change in the distribution of residuals.

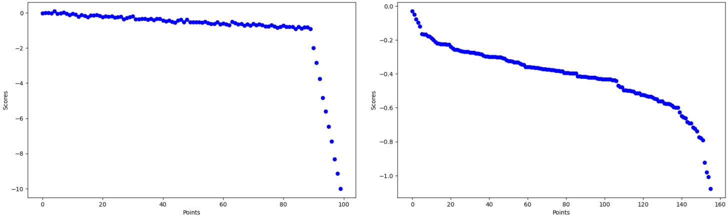


Figure: Ideal residuals on the left; Real residuals on the right

POLITECNICO MILANO 1863

Change Point Detection - Methods

- General Statistical Change Point detection method $\rightarrow$

$$\textit{thresh} = \hat{\mu} + \alpha \cdot \hat{\sigma}$$

	Location	Spread	Threshold
Inter-Quantile-Range	Q_2 (median)	$\sigma = Q_3 - Q_1$	$Q_3 + 1.5 \cdot \sigma$
Median Absolute Value	Median	σ=Median Absolute deviation	Median + 2.9 $\cdot\, \sigma$
Variance Based	Median	σ=Variance	Median + 1.5 $\cdot\, \sigma$
Estimator S_n	Median	$\sigma = S_n$	Median + 3 $\cdot\, \sigma$
Estimator Q_n	Median	$\sigma = Q_n$	Median + 3 $\cdot\, \sigma$
Forward Search	Median	σ=Median Absolute deviation	Adaptive Percentile

POLITECNICO
MILANO 1863

Performance indexes

- Performance of the Change Point Detection method:

$$\rightarrow \text{Silhouette Score} = \frac{b_i - a_i}{max\{b_i, a_i\}}$$

- Performance of the Multi Model fitting solution:

$$\rightarrow \text{Influence Function} = ||Model\_baseline - Model_{-i}||_2$$

POLITECNICO
MILANO 1863

Performance indexes - Silhouette Score (Homographies)

- For each model of each image, selection of the best Change Point Detection method based on the Silhouette Score.

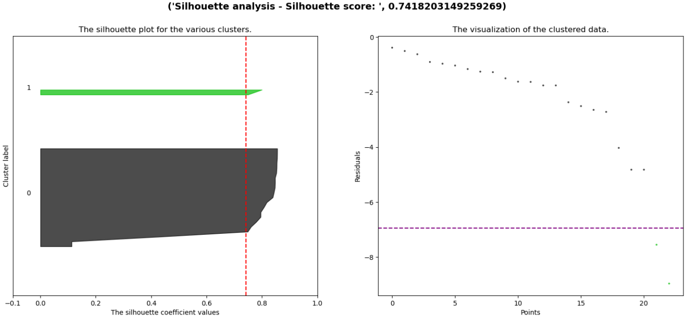


Figure: Silhouette Scores of one model of one image for Homography estimation - inlier threshold

Performance indexes - Silhouette Score (Fundamental Matrices)

- The change point detection methods that are robust to outliers (all but variance based) proved a bad performance for the Fundamental Matrix scenario, so we used only the Variance based approach.

POLITECNICO MILANO 1863

Performance indexes - Influence Function

- Most interesting Influence function results

Homography:

Image-Model	Inliers Avg Score	Outliers Avg Score	Increase Ratio
img3 model0	2.916 (37)	979.683 (1)	336.000
img7 model2	0.009 (290)	55.764 (1)	6358.609
img8 model0	0.043 (106)	19.624 (2)	458.326
img14 model0	0.002 (184)	0.606 (1)	332.099

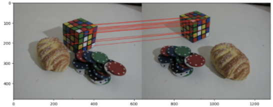
Fundamental Matrix:

Image-Model	Inliers Avg Score	Outliers Avg Score	Increase Ratio
img3 model1	0.000 (66)	0.086 (2)	1736.593
img5 model0	0.000 (32)	0.085 (1)	39708.388
img8 model1	0.000 (57)	0.029 (1)	410982.937
img15 model0	0.000 (76)	0.041 (2)	64894.942

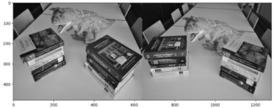
POLITECNICO
MILANO 1863

Ensembling technique

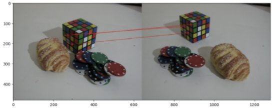
- Ensemble of GC-RANSAC , LMEDS and RANSAC.
- Ensembling improved the quality of the detections, both decreasing the number of false positives and increasing the number of detected outliers.



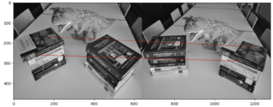
(a) gc-ransac detects some inliers as outliers



(b) gc-ransac doesn't detect outliers



(c) ensembling was able to filter false positives



(d) ensembling detects new outliers

Figure: Results found by gc-ransac and ensembling on outlier detection in FM estimation.

Experimental results

Homography estimation

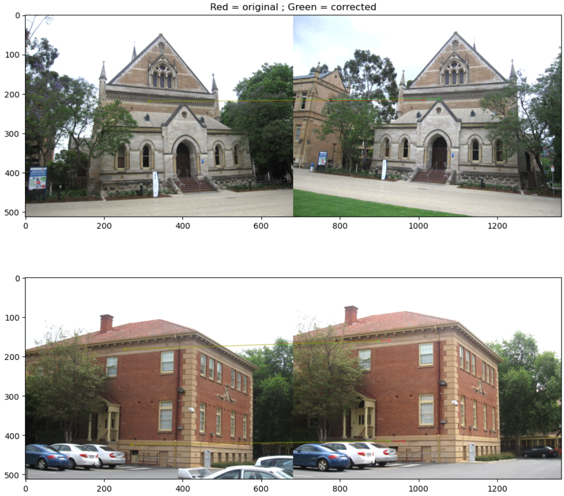


Figure: The two with the highest influence increase ratio

POLITECNICO
MILANO 1863

Fundamental Matrix estimation estimation

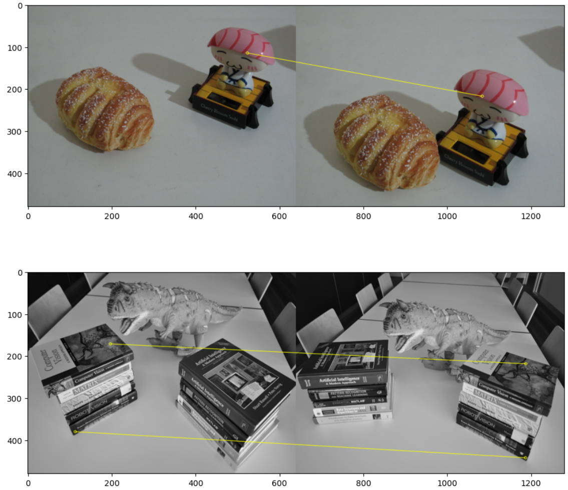


Figure: The two with the highest influence increase ratio

POLITECNICO
MILANO 1863

Open problems

Soft clustering performance index

- Soft Clustering in Multi Model fitting Scenarios.

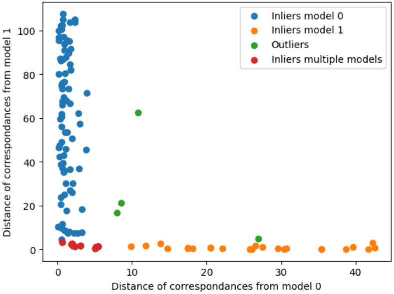


Figure: Example of soft clustering of Multi Model Fitting with two models.

Soft clustering performance index

- How to measure the performance of the final soft clustering? $\rightarrow$ Fuzzy Silhouette allows to measure the performance of standard soft clustering scenario.
    - How to adapt the Fuzzy Silhouette to our scenario?
    - How to define the distance of a point from the cluster of outliers?
    - Are the empirical results consistent and coherent with what is expected?

POLITECNICO
MILANO 1863