Robust Motion Segmentation

Team Members:

1. Pınar Erbil
2. Prachi Kedar

Introduction

- The task of detecting several moving objects in a set of photographs is known as **motion segmentation.** Usually, this is cast as a clustering problem, where the goal is to group points that belong to the same moving item together.
- The purpose of this study is to **improve** the approach in [1] by strengthening its resistance to outliers.

Related Work

Factorization methods [3]: Work with fully independent motions, not good when we have a moving camera

Generalized Principal Component Analysis (GPCA) methods [4]: Algebraic method, does not work well under complex data

Spectral Clustering methods: Uses affinity matrix to calculate embeddings.

- **"Motion Segmentation by Exploiting Complementary Geometric Models" (CVPR 2018) by Xu et al. [1]** proposes a multi-view approach combining fundamental matrix and homography with outlier detection and model selection for improved robustness.
- **"Robust Motion Segmentation From Pairwise Matches" (ICCV 2019) by Arrigoni et al. [5]** tackles motion segmentation without prior track knowledge, focusing on pairwise correspondences and introduces an averaging method for noise reduction
- **"Temporal Consistency Based Hierarchical Motion Clustering for Action Recognition" (CVPR 2020) by Wang et al.** leverages non-consecutive frames and enforces temporal consistency constraints during clustering to improve action recognition performance

Proposed Work - Base Method [1]

- **Geometric Model Hypothesis:** Randomly sample **p** points visible in both frames and use these **p** points to fit the model. We have affine (**p**=3), homography (**p**=4) and fundamental matrix (**p**=8).
- **Computing Ordered Residual Kernek (ORK):** Calculates the residuals between each data point automatically without a need for a threshold.
- **Spectral Clustering:** There are 3 single-view and 3 multi-view spectral clustering approaches presented.
    - **Single-view:** Affine, Homography, Fundamental
    - **Multi-view:** Kernel Addition, Co-regularization, Subset Constrained

Proposed Work

Aim of the project: Making the motion segmentation [1] robust to outliers.

1. We implemented different clustering algorithms
    a. Hierarchical
    b. DBSCAN
    c. Fuzzy C-mean
2. We run the algorithm with non-consecutive pairs and different parameters
    a. 2-framegap pairs
    b. 3-framegap pairs
3. We implemented 2 different fitting homography algorithms
    a. MSAC
    b. LMeds

Proposed Work - Clustering Algorithms

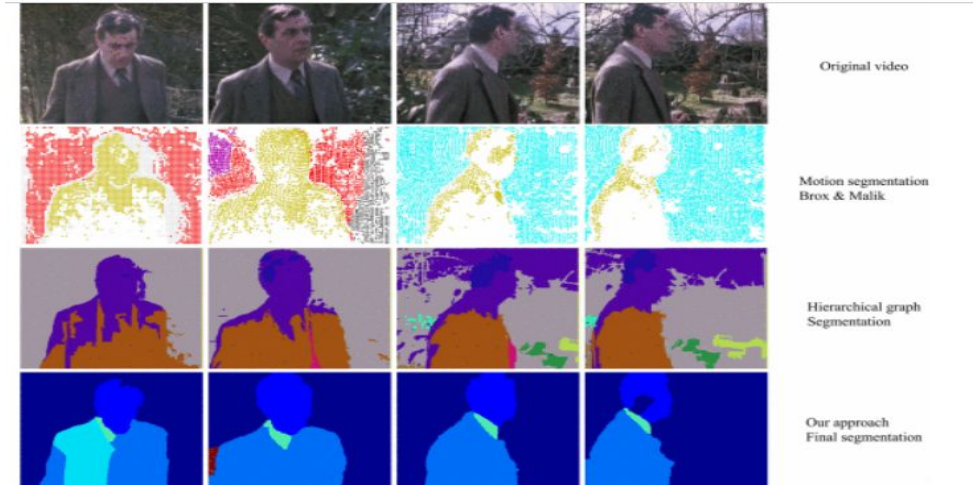
The method implemented in the previous approach was based on K-Means clustering algorithm which has some limitations such as:

- **Fixed Number of Clusters**: K-Means inputs the number of clusters and this can be challenging in real-life scenarios where one may not know the number of clusters beforehand.
- **Sensitivity to Noise and Outliers:** K-Means is sensitive to outliers and noise, which is the problem in our situation.
- **Assuming Spherical Clusters:** K-Means assumes clusters are spherical, this can especially in real-world data can create problems because we mostly have non-spherical objects.

Proposed Work - Hierarchical Clustering

Hierarchical Clustering: Hierarchical clustering creates a hierarchy of clusters, allows to explore different granularities and potentially identifying hidden substructures within the motion patterns. This is helpful in understanding the complexity of the scene and uncovering finer details.A cluster tree or dendrogram is created using hierarchical clustering to group data across several scales. The tree is a hierarchical hierarchy, with clusters at one level joining to form clusters at the next level. This allows you to choose the level or scale of clustering that is most suited to the application.

'Long term video segmentation through pixel level spectral clustering on GPUs [7]' uses hierarchical graph segmentation. technique to retrieve 3 objects with less than 10% error compared to other objects



Proposed Work - DBSCAN

DBSCAN: DBSCAN automatically identifies clusters based on density, eliminating the need to predefine the number of clusters. This flexibility might be useful if the clusters have varying densities or irregular shapes.This makes it a good candidate for motion segmentation tasks, where the goal is to segment moving objects from the background based on their motion patterns.

‘Detection of moving foreground objects in videos with strong camera motion’ [6]

Uses clustering method DBSCAN in a mixed motion and colour space to extract foreground objects.

Proposed Work - Fuzzy C-Means (FCM)

Fuzzy C-Means (FCM): FCM allows data points to belong to multiple clusters with varying degrees of membership, making it more flexible for overlapping or ambiguous data. This could be beneficial if the motion patterns are not clearly distinct.

It is based on minimization of the following objective function.where m is any real number greater than 1, uij is the degree of membership of xi in the cluster j, xi is the ith of d-dimensional measured data, cj is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center.

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \quad , \quad 1 \le m < \infty$$

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j

Proposed Work - Non-consecutive pairs

2-framegap pairs: Fit the hypotheses using the pair (Frame 1, Frame 3)

3-framegap pairs: Fit the hypotheses using the pair (Frame 1, Frame 4)

Frame 1

Frame 2

Frame 3

Frame 4

Proposed Work - Non-consecutive pairs

For non-consecutive pairs, we also alter some of the model parameters.

a. **Hypotheses count:** The original method was generating 500 hypotheses for each frame with consecutive pairs, we test for range [300, 400, 500, 600, 700]
b. **Lambda value:** Co-regularization method is using a lambda = 1e-2. We experiment with [1e-4 1e-3 1e-2 1e-1]
c. **Gamma value:** Similarly, subset method uses a gamma value = 1e-2. We experiment with [1e-4 1e-3 1e-2 1e-1]

$$\min_{\{\mathbf{U}_v\}} \sum_v tr(\mathbf{U}_v^\top \mathbf{L}_v \mathbf{U}_v) - \lambda \sum_v \sum_{w, w \neq v} tr(\mathbf{U}_v \mathbf{U}_v^\top \mathbf{U}_w \mathbf{U}_w^\top),$$
$$s.t. \mathbf{U}_v^\top \mathbf{U}_v = \mathbf{I}$$

Co-regularization algorithm

$$\min_{\{\mathbf{U}_v\}} \sum_v tr\left(\mathbf{U}_v^\top \mathbf{L}_v \mathbf{U}_v\right) - \gamma tr(\mathbf{U}_v^\top \mathbf{Q}_v \mathbf{U}_v), \quad s.t. \mathbf{U}_v^\top \mathbf{U}_v = \mathbf{I},$$
$$\mathbf{Q}_v = \begin{cases} \mathbb{1}\left(\hat{\mathbf{K}}_{v+1} < 0\right) \circ \hat{\mathbf{K}}_{v+1}, & v = 1 \\ \mathbb{1}\left(\hat{\mathbf{K}}_{v-1} > 0\right) \circ \hat{\mathbf{K}}_{v-1} + \mathbb{1}\left(\hat{\mathbf{K}}_{v+1} < 0\right) \circ \hat{\mathbf{K}}_{v+1}, & v = 2 \\ \mathbb{1}\left(\hat{\mathbf{K}}_{v-1} > 0\right) \circ \hat{\mathbf{K}}_{v-1}, & v = 3 \end{cases}$$

Subset algorithm

Proposed Work - Fitting Homography

Since current architecture uses RANSAC to fit an homography we implemented other two methods such as MSAC and LMeds to see how to works to fit an homography model.

1. **MSAC:**

More robust to partially corrupted data due to its robust cost function.

Often more efficient for large datasets due to its faster convergence compared to RANSAC.

2. **LMedS:**

   Offers even stronger outlier resistance thanks to its median-based approach.

   Less sensitive to parameter choices like the inlier threshold compared to RANSAC.

Experiments - Dataset

- KITTI 3D Motion Segmentation Benchmark (KT3DMoSeg) [1] is used for the dataset.
- Has 22 video clips from the KITTI dataset [2] with 10-20 frames.

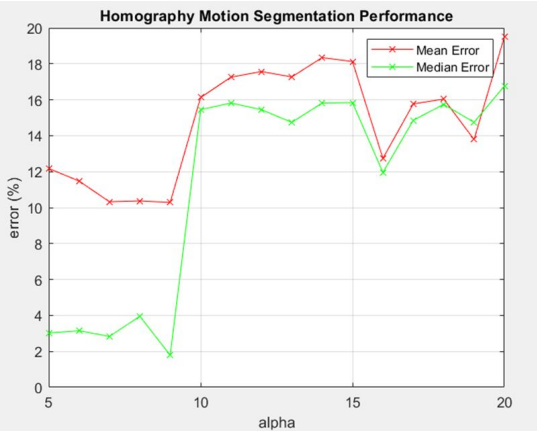
Experiments - Results for Classification Methods

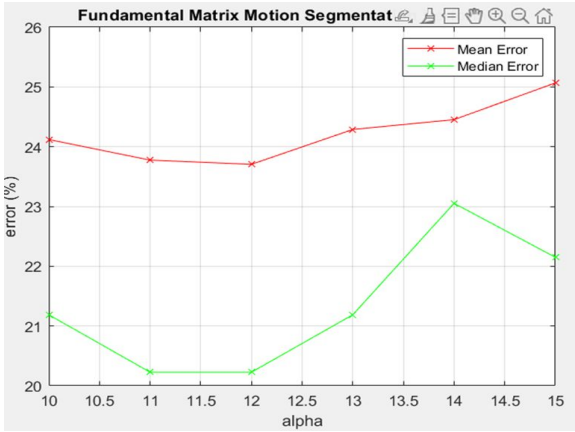
Models	K means	Hierarchical	DBSCAN	Fuzzy C Mean
Affine	15.45	24.90%	21.58%	22.56%
Fundamental	20.19	23.58%	25.07%	20.56%
Homography	14.26	21.93%	18.13%	24.53%
AHF_Subset	9.45	27.11%	20.39%	27.96%
AHF_KerAdd	8.78	30.98%	19.65%	17.53%
AHF_CoReg	7.81	28.60%	20.17%	24.60%

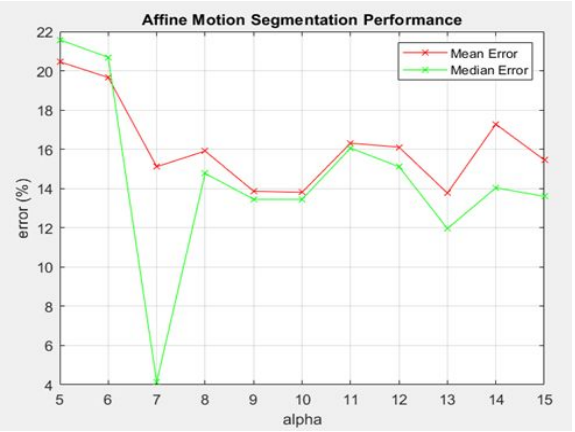
In the above table we have computed **Mean Error** for Affine , Fundamental & Homography and **Overall Misclassification Error** for the AHF_Subset, AHF_KerAdd, AHF_CoReg.

Experiments - Results for Classification Methods

Hierarchical Clustering

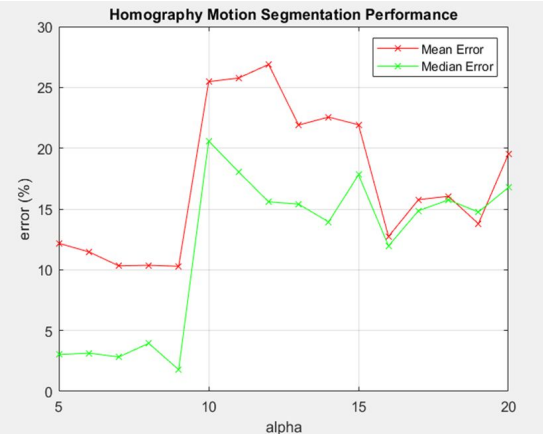


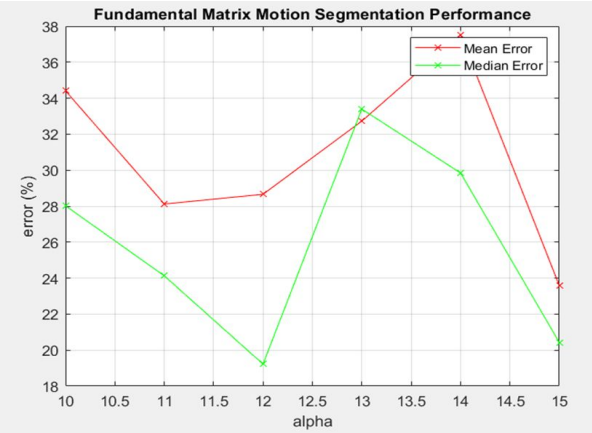


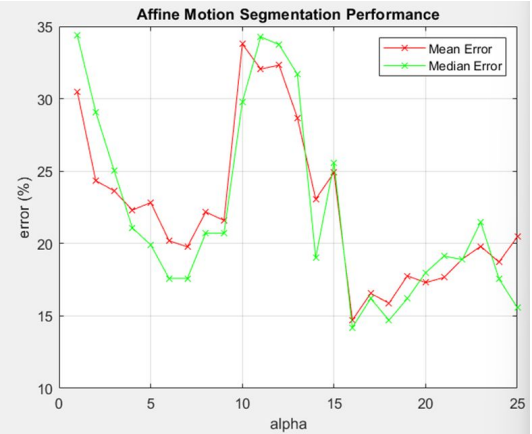


Experiments - Results for Classification Methods

DBSCAN

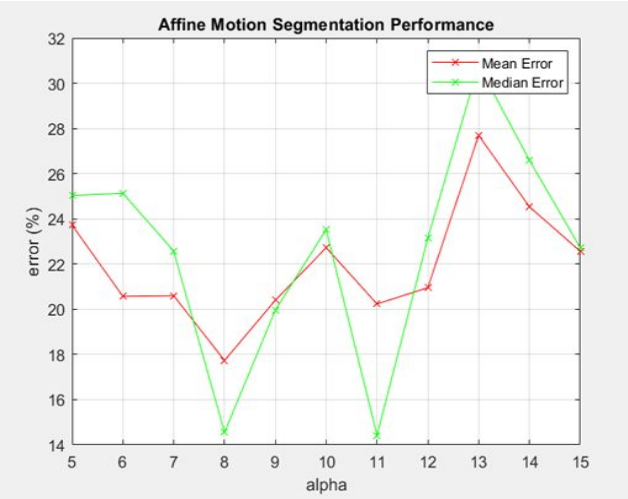


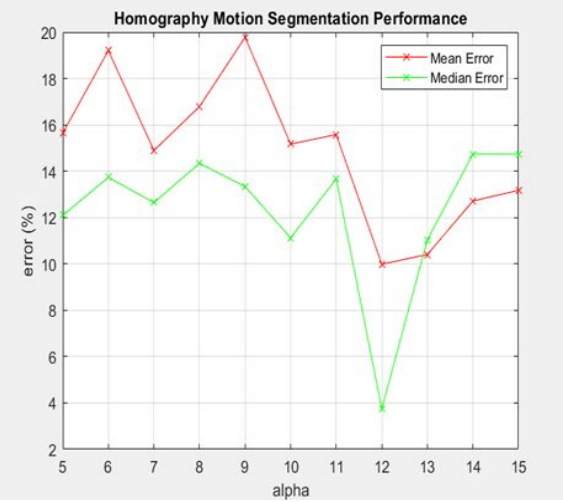


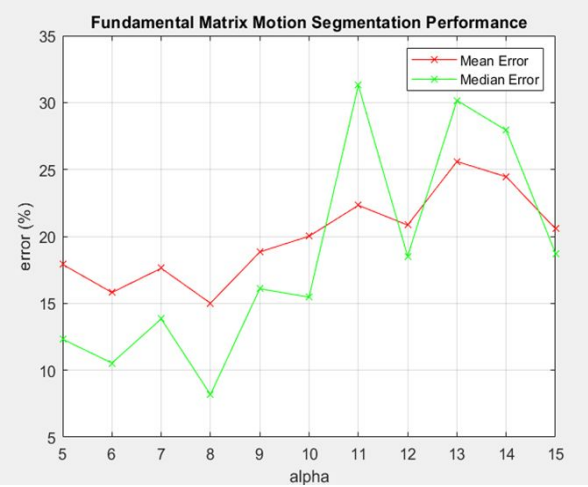


Experiments - Results for Classification Methods

Fuzzy C-Mean







Experiments - Results for Non-consecutive Pairs

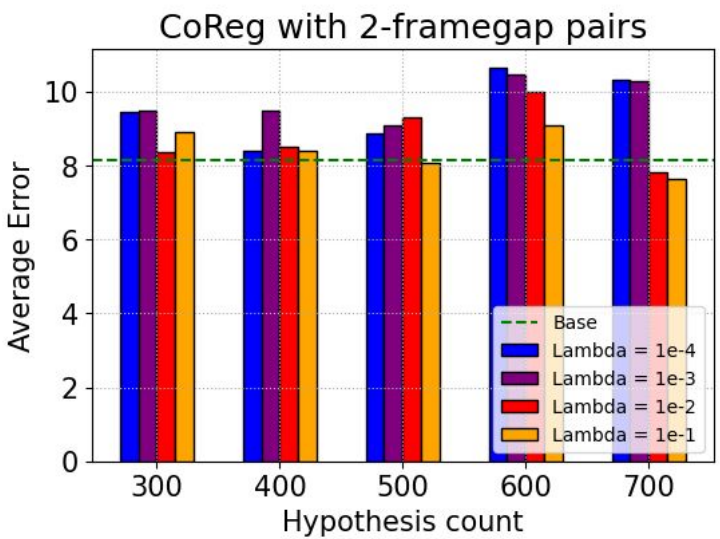
Model	Base		Number of Hypothesis Generated									
			300		400		500		600		700	
	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
Affine	18.85	17.3	17.7	16.7	**14.13**	14.34	**14.15**	**9.3**	15.7	15.3	15.3	13.7
Homography	12.9	10.2	**8.3**	4	8.7	**2.3**	10	4.3	14	12.7	10.7	6.7
Fundamental	15.2	10.8	12.7	6.3	**10**	**4.3**	12.3	6	11	5	11	4.7
KerAdd	8.5	1.9	**7.7**	**1**	9	2	8	**1**	9.3	2.7	9.3	2
CoReg	8.7	2.4	8.7	**1**	**8.3**	**1**	8.7	**1**	10	2	**8.3**	**1**
Subset	11	2.1	12.3	1.3	10.7	**1**	10	**1**	12.7	**1**	**9.7**	**1**

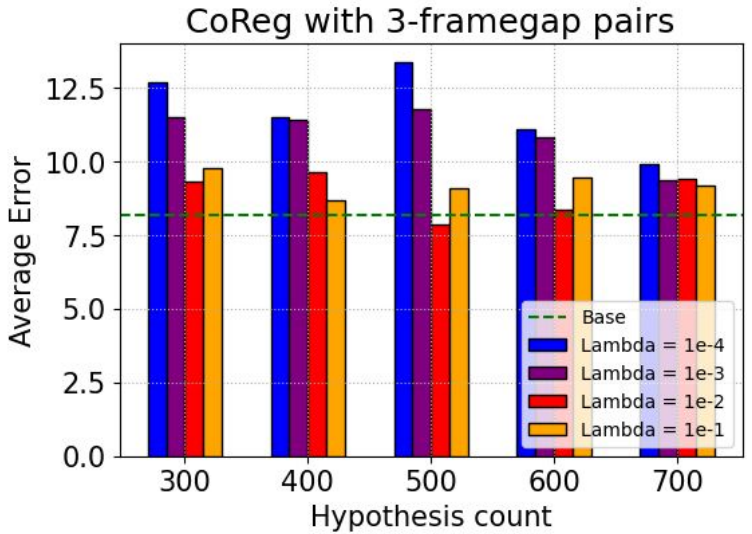
Table 2. Average and Median Errors (%) for 2-framegap Pairs. The best values are in bold, for base we are reporting the results for consecutive-pairs with hypothesis number = 500

Model	Base		Number of Hypothesis Generated									
			300		400		500		600		700	
	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
Affine	18.9	17.3	19	**16**	18	16.3	19	17.3	18.7	17.7	**17.7**	18.3
Homography	12.9	10.2	12.3	11	13.3	7.3	**10.3**	6.3	**10.3**	7.3	12.3	**6**
Fundamental	15.2	10.8	**12.3**	9.7	13.7	11.3	14.3	8.3	15.3	9	12.7	**7**
KerAdd	**8.5**	1.9	9.3	**1**	9.3	2.3	10	3.7	11	3.7	11.3	3.7
CoReg	**8.7**	2.4	9.3	3.3	9.3	2	**8.7**	**1.3**	**8.7**	2.3	10	2
Subset	**11**	2.1	12.3	5	12	4.7	13.3	**2**	13	2.3	13	**2**

Table 3. Average and Median Errors (%) for 3-framegap Pairs. The best values are in bold, for base we are reporting the results for consecutive-pairs with hypothesis number = 500

Experiments - Effect of Lambda

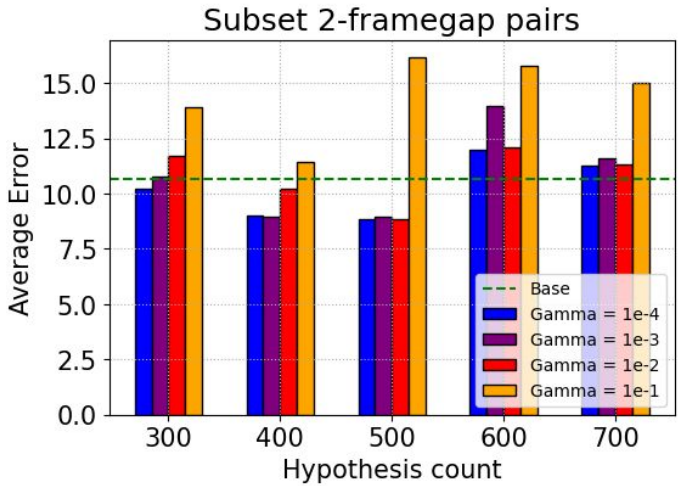


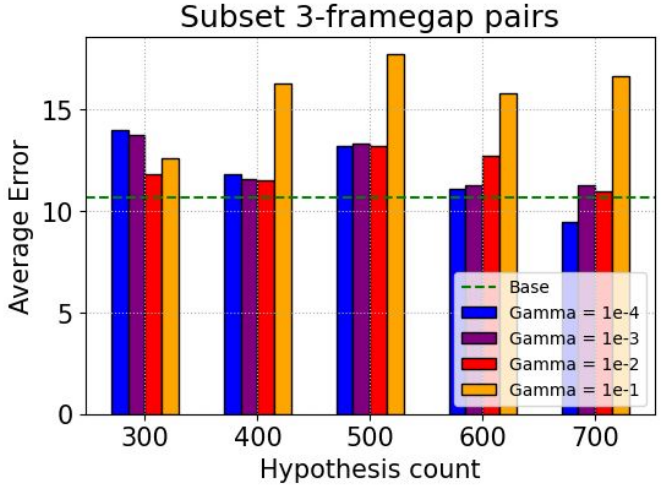


$$\min_{\{\mathbf{U}_v\}} \sum_v tr(\mathbf{U}_v^\top \mathbf{L}_v \mathbf{U}_v) - \lambda \sum_v \sum_{w, w \neq v} tr(\mathbf{U}_v \mathbf{U}_v^\top \mathbf{U}_w \mathbf{U}_w^\top),$$

$$s.t. \mathbf{U}_v^\top \mathbf{U}_v = \mathbf{I}$$

Experiments - Effect of Gamma





$$\min_{\{\mathbf{U}_v\}} \sum_v tr\left(\mathbf{U}_v^\top \mathbf{L}_v \mathbf{U}_v\right) - \gamma tr(\mathbf{U}_v^\top \mathbf{Q}_v \mathbf{U}_v), \quad s.t. \mathbf{U}_v^\top \mathbf{U}_v = \mathbf{I},$$

$$\mathbf{Q}_v = \begin{cases} \mathbb{1}\left(\hat{\mathbf{K}}_{v+1} < 0\right) \circ \hat{\mathbf{K}}_{v+1}, & v = 1 \\ \mathbb{1}\left(\hat{\mathbf{K}}_{v-1} > 0\right) \circ \hat{\mathbf{K}}_{v-1} + \mathbb{1}\left(\hat{\mathbf{K}}_{v+1} < 0\right) \circ \hat{\mathbf{K}}_{v+1}, & v = 2 \\ \mathbb{1}\left(\hat{\mathbf{K}}_{v-1} > 0\right) \circ \hat{\mathbf{K}}_{v-1}, & v = 3 \end{cases}$$

Experiments - Results for Homography Model

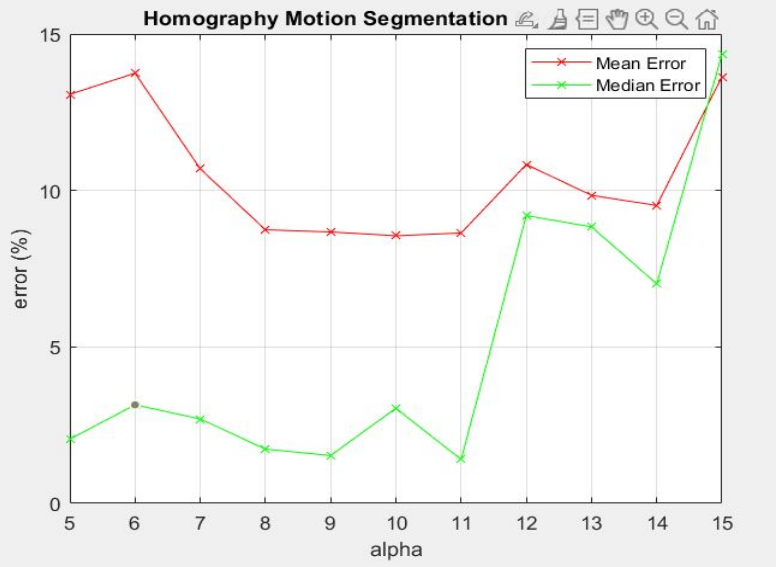
Models	MSAC	LMeds	RANSAC
Mean Error	10.12%	15.39%	14.26%

Experiments - Fitting Homography using MSAC

- **MSAC :** is a robust estimation algorithm commonly used in tasks like motion segmentation, especially when dealing with **outliers in data**. It iteratively fits a model to a small set of randomly selected data points ("samples") and then evaluates how well the model fits the remaining points.

  It is Faster than RANSAC (another outlier handling algorithm) due to early termination on large consensus sets.

- **The average mean error comes out is 10.12 % which is better than the RANSAC implementation which was 14.26% .**

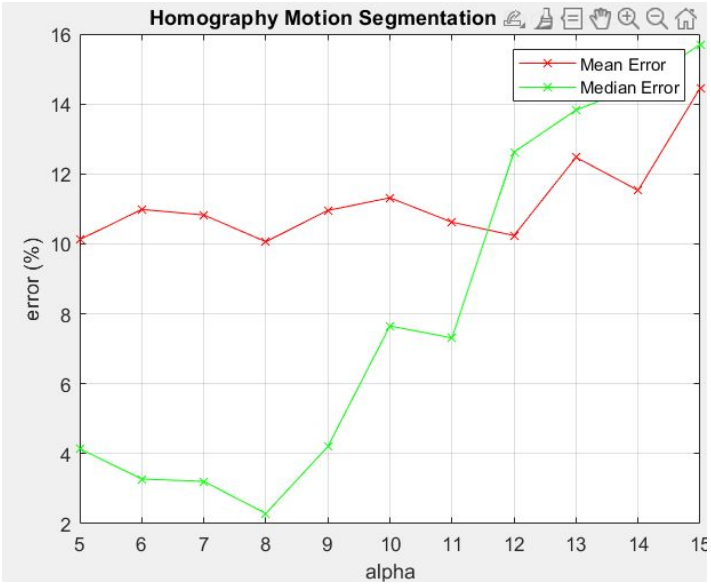


MSAC implementation

Experiments - Fitting Homography using lMeds

- **LMedS (Least Median of Squares):** This is a robust estimation method that finds the median of a set of least squares fits to small subsets of the data. This is less sensitive to outliers than the mean of the least squares fits.

- **The average mean error comes out is 15.39 % which is slightly higher than the RANSAC implementation which was 14.26%**



Lmeds implementation

Conclusion

1. We implemented and compared different clustering algorithms: Hierarchical, DBSCAN and Fuzzy C Means.
    - Perform worse than K-Means
2. Non-consecutive pairs: 2-framegap and 3-framegap pairs
    - We change the hypotheses number range.
    - We did not see drastic improvements; however, 2-framegap pairs outperform the base model in every setting.
3. Fitting the homography model: LMedS and MSAC
    - Base model is using RANSAC with a mean error 14.26% .
    - We showed MSAC outperforms it with a mean error 10.12%.

References

[1] X. Xu, L. F. Cheong, and Z. Li. Motion segmentation by exploiting complementary geometric models. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2859–2867, 2018. 1, 2

[2] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. The International Journal of Robotics Research, 32(11):1231–1237, 2013. 3

[3] J. P. Costeira and T. Kanade. A multibody factorization method for independently moving objects. International Journal of Computer Vision, 29:159–179, 1998. 1

[4] R. Vidal and R. Hartley. Motion segmentation with missing data using powerfactorization and gpca. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., volume 2, pages II–II. IEEE, 2004. 1

[5] F. Arrigoni and T. Pajdla. Robust motion segmentation from pairwise matches. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 671–681, 2019. 1

[6] . M. Y. G. J.-F. D. D. Szolgay, J. Benois-Pineau ,Detection of moving foreground objects in videos with strong camera motion, Pattern Analysis and Application 2011,Volume 14, pages 311–328

[7] Narayan Sundaram, Kurt Keutzer, Long term video segmentation through pixel level spectral clustering on GPUs, 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2012