

CS283 Fall 2017 Final Project

Nao Ouyang

December 13, 2017

Abstract

plicate their results.

1 Introduction

According to this paper, motion segmentation is the problem of determining which

2 Motivation

In the decade before this paper, people did a lot of work with static scenes where the camera moved. However, there was not as much work done on dynamic

3 Methods

3.1 Problem Formulation

- n = number of objects aka independent motions aka subspaces
- P = number of (x, y) point correspondences
- F = number of frames

Let us stack all the frames vertically so that we get a matrix W

$$W = MS^T$$

where M is motion matrix $\in \mathbb{R}^{2F}$ and where S is shape matrix (point correspondences).

Additionally, we have the world coordinates (x, y, z) as

$$X = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

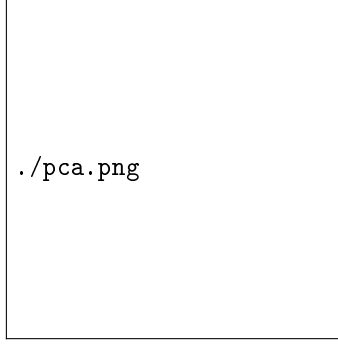


Figure 1: If we imagine the blobs as clouds of points in 3D (so imagine them going at some angle through space), the problem with picking a single subspace is that the data lies more naturalies in two subspaces, e.g. the two lines going through the gray and dark gray clusters.

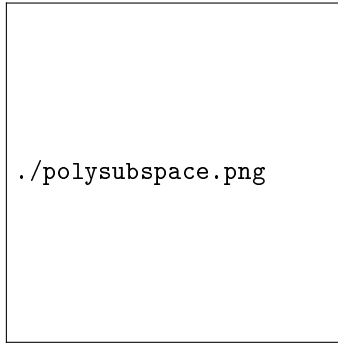


Figure 2: Here the intersection, or combined zero-set, of two 3D planes are used to define a line (2D subspace) in \mathbb{R}^3 .

$$x = \begin{bmatrix} x \\ y \end{bmatrix} \quad (2)$$

The affine camera matrix translating between a given world and image coordinate is written in terms of rotations and translations as $A \in \mathbb{R}^{2 \times 4}$

$$A = \begin{bmatrix} R1 & R2 & R3 & T1 \\ R3 & R4 & R5 & T2 \end{bmatrix} \quad (3)$$

3.1.1 Representing motion subspaces with polynomials

The gist of this section is that, if all n subspaces (each representing one object) have dimension 4 in \mathbb{R}^5 , then a single linear polynomial of degree 5 is enough to represent them. To see this, consider that a plane may be represented by a linear polynomial of three variables, $ax + by + cz = 0$. Then, we can fit a polynomial defined as the product of the n planes, which will have at most degree n and have three variables. We can also describe lower dimension subsubspaces as a the "common zero set" of multiple hyperplanes. For instance, a line can be described as the intersection of two hyperplanes.

Aside: an algebraic variety is roughly like a "shape" that fits the points. Here we are saying that if this polynomial "shape" fits all the datapoints in n subspaces, then we should be able to factor the algebraic variety nto n polynomials. Co-dimension means the number of dimensions not occupied by a subspace, e.g. a subspace of dimension 4 in 5 dimensional space has a co-dimension of 1.

4 Results

I used the dataset Hopkins 155 provided by JHU. It is available at <http://www.vision.jhu.edu/data/>.

I created a MATLAB file which plots the datapoints in the .mat files against the video. This file may be found as . In order to run

My code appeared to have an error rate of 35% for the 'arm' dataset in Pt 6 of Hokins 155. This was unfortunately much larger than the reported 5% error for most of the datasets.

I did not get around to plotting the points by color onto the video to debug what was going on.

5 Conclusion

Powerfactorization was not implemented, and the code appears to still have issues. To Debug, I should assign colors to my labels and check what might be causes of the issue. Future work includes playing around with RANSAC with outliers.

6 References

R. Vidal and R. Hartley, "Motion segmentation with missing data using PowerFactorization and GPCA," Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., doi: 10.1109/CVPR.2004.1315180. [Paper link](#)

Hartley R.I. and Tron, R., & Vidal, R. (2007). Multiframe Motion Segmentation with Missing Data Using PowerFactorization and GPCA. International Journal of Computer Vision, 79, 85-105. doi: 10.1007/s11263-007-0099-z. [Paper link](#)