Generative models using Probabilistic Principal Component Analysis

Group - 2
Bhavya Patwa - 1401063, Harsh Mehta - 1401086
Vidit Shah - 1401078, Yash Turakhia - 1401118
School of Engineering and Applied Science
Ahmedabad University

***Abstract*—Principal Component Analysis is used for dimension reduction and data analysis. However, this model may not be used for corrupted or incomplete data. For this we use an idealistic version of PCA that are Robust PCA and Probabilistic PCA. Here, in this report we compare both the techniques on parameters such as root mean square error and execution time on recovered image.**

I. Introduction And Theory

A. Probabilistic Principal Component Analysis

Principal component analysis is a technique used for data analysis and processing. However it is not based upon a probabilistic model. Here, We are determining observed data vectors using maximum-likelihood estimation of parameters in a latent variable model which is closely related to factor analysis. We are using EM algorithm for estimating the principal subspace iteratively. The results are based on the algorithm discussed in [1] and [2].

B. Robust Principal Component Analysis

PCA is a widely used statistical tool for data analysis and dimension reduction. However, PCA fails when large number of data are corrupted or missing. In Robust PCA we are given a data matrix which is decomposed into low-rank and sparse component. Our goal is to recover the low-rank matrix from the corrupted and incomplete data matrix. Detailed description of this algorithm for corrupted data and incomplete data is described in [2]. We have used the same algorithm for our analysis.

C. Comparison between Robust PCA and Probabilistic PCA

Here, we are comparing performance of the Robust PCA and Probabilistic PCA based on generated data. For comparison we take the corrupted image as well as incomplete image where comparison parameters are Root mean square error and Execution Time.

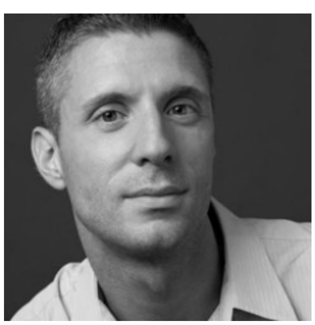

Figure 1.1 : Input Image

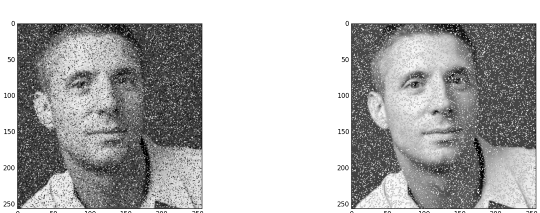

Figure 1.2: Image with corrupted and missing entries

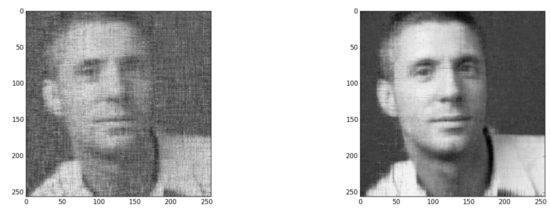

Figure 1.3: Comparison between RPCA and PPCA in Corrupted Image

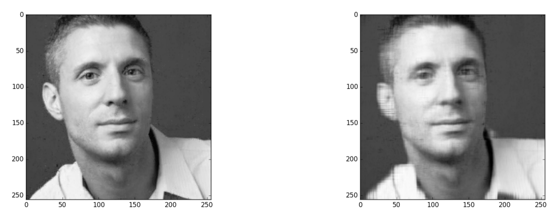

Figure 1.4: Comparison between RPCA and PPCA in Incomplete Image

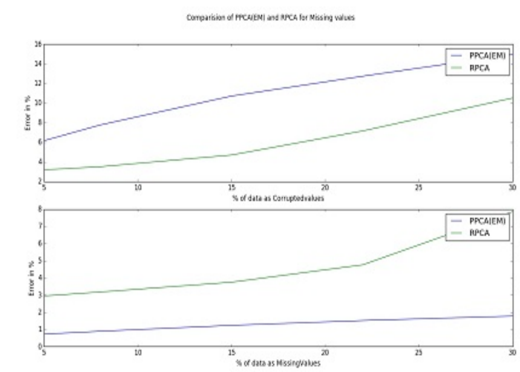

Figure 1.5 : RMSE Comparison of RPCA and PPCA on missing and corrupted respectively

	Missing		Corrupted	
%missing & corrupted	RPCA	PPCA	RPCA	PPCA
5	3.44	0.36	3.17	1.23
8	3.53	0.39	2.8	1.22
15	3.13	0.43	2.16	1.84
22	3.36	0.49	2.26	1.18
30	4.55	0.53	2.32	1.19

Table 1: Performance Analysis based on Execution Time (secs)

References

[1] Cands, E. J., & Recht, B. (2009). Exact Matrix Completion via Convex Optimization. Foundations of Computational Mathematics, 9(6), 717-772. doi:10.1007/s10208-009-9045-5

[2] Tipping, M. E., & Bishop, C. M. (1999). Probabilistic principal component analysis. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3), 611-622.

[3] Josep Porta, Jakob Verbeek, and Ben Krose. Active appearance-based robot localization using stereo vision. IvI - Instituut voor Informatica