Comparision on Probabilistic Principal Component Analysis and Robust Principal Component Analysis

Hemang Jethva, Jainil Vachhani, Nand Parikh, Palash Hariyani, Parth Shah

School of Engineering and Applied Science
Ahmedabad University

***Abstract*—Principal component analysis(PCA) is a well perceived technique for dimensionality reduction, data analysis. Various PCA techniques developed and proposed has the capability to recreate the original image even when the image matrix is either corrupted or is incomplete.Here we propose a comparision between two of such PCA technique viz. Robust Principal Component Analysis(RPCA) and Probabilistic Principal Component Analysis(PPCA).A comparision on corrupted and incomplete image matrix is done, which judges the performance on time taken to recreate the image and error rate when either data is corrupted or missing.**

Keywords: RPCA, Factor Analysis, PPCA, Expectation Maximization (EM).

I. Robust PCA

Robust PCA(RPCA) decomposes the data matrix into a lower rank matrix and a sparse matrix, where the lower rank matrix would be a lower dimensional recreated matrix. RPCA works on the essence of classical PCA. Principal Component Pursuit (PCP) algorithms have been proposed in [1], which sucessfully retrives the lower rank matrix and a sparse matrix from a corrupted or incomplete image. The assumptions, claim and algorithms for achieving these lower rank and sparse matrix is comprehensively explained in [1].

II. Probabilistic PCA

Probabilistic PCA (PPCA) model absorbs the essense of factor analysis[2] for finding the pricipal axes of a set of observed data vectors using maximum-likelihood estimation of the latent variables's parameter. Expectation Maximization (EM) algorithms, an iterative approach to recreate the original matrix from either corrupted or incomplete matrix are used for the comparision with RPCA. An EM approach for corrupted data matrix is proposed in [2], similarly [3] has given an EM algorithm for completing a matrix where values are missing.

III. Comparing RPCA and PPCA

Performance of RPCA and PPCA is judged on 2 parameters, accuracy at different number of Principle components (PCs) taken and time required for recreating the image. Both the scenarios matrix corrution and matrix completion is taken into consideration.Standard *Cameraman* is taken as our sample image. The figures below states that for corrupted values RPCA performs better(fig1-right), but at the same time table1 shows that time taken by RPCA is very huge. For missing values PPCA outperfoms RPCA on both the parameters, hence from the observed results we come to a conclusion that PPCA with EM is a better approach then PPCA.

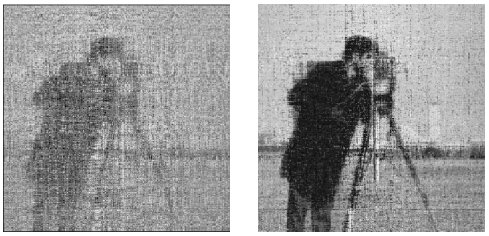

Fig1:Corrupted PPCA and RPCA

Fig2:Missing Values PPCA and RPCA

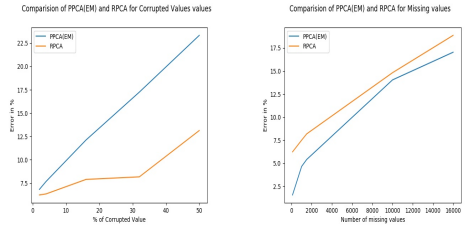

Fig3:Comparision for error rate

corruption(%tage)	RPCA(sec)	PPCA(sec)
2	1.483	2.793
4	1.491	2.69
16	1.523	2.904
32	1.66	2.992
50	1.72	3.315

Table1: Time taken for Corruption Matrix

number of missing value	RPCA(sec)	PPCA(sec)
100	0.688	2.603
1000	0.715	2.69
1500	0.728	2.71
10000	0.858	2.77
16000	0.950	2.81

Table2: Time taken for Missing Value

References

[1] E.J Candes, Xiaodong Li, Yi Ma, and John Wright. Robust Principal Component Analysis.

[2] Tipping, M. E., and C. M. Bishop. Probabilistic Principal Component Analysis. Journal of the Royal Statistical Society. Series B (Statistical Methodology), Vol. 61, No.3, 1999, pp. 611622.

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