Generative models using Probabilistic Principal Component Analysis

Akash Soni (1401047), Jay Joshi (1401005), Mihir Gajjar (1401076), Nishi Shah (1401099), Shruti Agrawal (1644007)

***Abstract*—The main aim of this project is to be able to build Generative models using the concept of Probabilistic Principal Component Analysis or PPCA. Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Robust PCA and Probabilistic PCA are techniques to recover missing data a matrix of lower dimension using different approaches and estimate principal subspace iteratively.**

***Index Terms*—Probabilistic PCA, Robust PCA, Latent Variables, Expectation Maximization.**

I. Introduction

In this report, a comparison between the Robust PCA and PPCA with EM algorithm is done. Experiments are conducted with corrupted and missing entries in the original image. The error is found with RMSE(Root Mean Square Error) and the time is the real time obtained by the time command in linux.

II. Robust Principle Component Analysis

Robust Principal Component Analysis (RPCA) is a modification of principal component analysis (PCA) which works well with respect to grossly corrupted observations. Robust PCA aims to recover a low-rank matrix L_0 from highly corrupted measurements $M = L_0 + S_0$. Then the optimization is done by Augmented Lagrange Multiplier Method(ALM).

III. Probablilistic Principle Component Analysis - Expectation Maximization Algorithm

Probabilistic PCA is an example of the linear-Gaussian framework. Latent Model Representation by Factor analysis is represented as $x = Wz + \mu + \epsilon$, where z is an M-dimensional Gaussian latent variable, μ is the mean and ϵ is a D-dimensional zero-mean Gaussian-distributed noise variable with covariance σ^2I. The EM Algorithm is based on two steps - the E-Step (Expectation Step) that guesses a probability distribution over completions of missing data in the current model; and the M-Step (Maximisation Step) that re-estimates the model parameter using these completions.

IV. Implementation Result

Figure 1.1 : Input Image

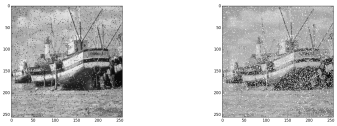

Figure 1.2: Images with corrupted and missing entries respectively

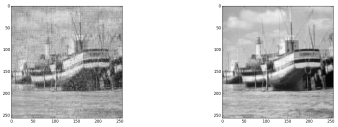

Figure 1.3: Comparison between PPCA and RPCA in Image with corrupted entries respectively

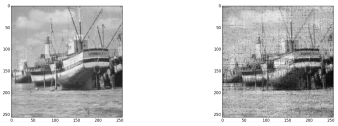

Figure 1.4: Comparison between PPCA and RPCA in Image with missing entries respectively

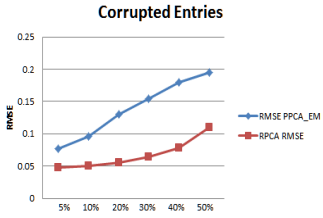

Figure: 1.5

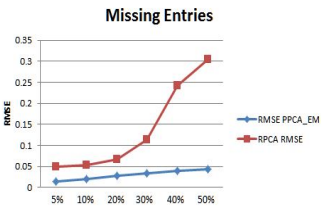

Figure:1.6

	Missing		Corrupted	
%missing & corrupted	RPCA	PPCA	RPCA	PPCA
5.00 %	2.910s	1.049s	3.012s	2.319s
10.00 %	2.861s	1.118s	3.073s	2.366s
20.00 %	3.068s	1.221s	2.861s	2.314s
30.00 %	4.253s	1.243s	2.820s	2.277s
40.00 %	2.932s	1.288s	3.259s	2.340s
50.00 %	2.859s	1.293s	3.742s	2.304s

Table 1: Performance Analysis based on Execution Time

V. Interpretation and Conclusion

- PPCA is a probabilistic method that uses Linear Gaussian model of Latent Variables and EM algorithm to increase likelihood estimation.
- RPCA uses Augmented Langrange multiplier Algorithm for tractable convex optimization to retrieve data in grossly corrupted image of low rank matrix.
- **Time Analysis :** For increasing data loss, Time to compile is much less for PPCA as compared to RPCA.
- **Error Rate :** For corrupted data, PPCA is less efficient while for missing data RPCA is less efficient.

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.