ISSN:0975-9646
R. Muthukrishnan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 8293-8296

MCD based Principal Component Analysis in Computer Vision

R. Muthukrishnan , E.D. Boobalan, R. Mathaiyan

Department of Statistics, Bharathiar University
Coimbatore, Tamilnadu, India

Abstract— **Principal component analysis has been widely used in computer vision tasks. In image processing the outliers typically occur within the sample due to pixels that are corrupted by noise, alignment error, occlusion etc. The conventional PCA is based on the least square approach. However the least squares approach fails to account the outliers and produce the unreliable results. Many robust alternatives are proposed such as M estimator, MCD and S estimators. This paper makes an attempt to perform principal component analysis with most widely used these robust procedures. Also it is proposed a method which is based on MCD approach. The accuracy of the proposed method has been studied with help of an image along with existing algorithms.**

Keywords — **PCA, Least squares, MCD, Robust PCA, Computer Vision.**

I. INTRODUCTION

Multivariate analysis is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. Multivariate Analysis consists of a collection of methods that can be used when several measurements are made on each individual or object in one are more samples. For many years the applications lagged behind the theory because the computations were beyond the power of the available desktop calculators. However, with modern computers, virtually any analysis one desires, no matter how many variables or observations are involved, can be quickly and easily carried out. The essence of multivariate thinking is to expose the inherent structure and meaning revealed within these sets of variables through application and interpretation of various statistical methods.

Principal Component Analysis is probably the oldest and best known of the techniques of multivariate analysis. It was first introduced by Pearson (1901), and developed independently by Hotelling (1933). Like many multivariate methods, it was not widely used until the advent of electronic computers, but it is now well entrenched in virtually every statistical computer package. The PCA is based on classical mean vectors, covariance matrix with lease square principle. But these tools get affected if the data contains extreme observations/noise and thus produces unreliable results. The established robust methods overcome these limitations to some extent, but still it’s a challenging task of the research communities to extract the complete information from the data with/without contamination.

PCA is arguably the most widely used statistical tool for data analysis and dimensionality reduction today. However, its brittleness with respect to grossly corrupted observations often puts its validity in jeopardy-a single grossly corrupted entry in data. Unfortunately, gross errors are now ubiquitous in modern applications such as image processing, web data analysis, and bioinformatics, where some measurements may be arbitrarily corrupted (due to occlusions, malicious tampering, or sensor failures) or simply irrelevant to the low-dimensional structure. A number of natural approaches to robustifying PCA have been explored and proposed in the literature over the past decades.

The study of computer vision is strongly interdisciplinary. The purpose of computer vision is to develop theories and algorithms to automatically extract and analyze useful information from an observed image, image set, or image sequence. The appearance based approaches to vision problems have recently received the attention in the vision community due to their ability to deals with such a problems like shape, reflection and some illumination conditions. The principal component analysis is a well-known and widely used technique in this context.

Many computer vision communities have been used principal component analysis for image processing. The major drawback of this traditional method is that it requires normalized (aligned) samples in the training data. On the other hand, the standard PCA approach is not robust, where the term robustness refers to the fact that the results remain stable in the presence of various types of noise and can tolerate a certain portion of outliers. For that, the researchers have been developed many kind of robust of principal component analysis procedures based on the different kind of robust estimators.

This paper presents principal component analysis with various types to perform computer vision task specifically image processing along with an experimental results. Section 2 presents the brief discussion and the algorithm for the conventional PCA. The robust PCA and its algorithm have been discussed in section 3 and the algorithm for weighted PCA is presented in section 4. The proposed algorithm namely modified robust PCA is presented in the section 5. Finally, the last section of this paper provides experimental results which are based on the proposed algorithm along with existing algorithms in the context of image processing.

II. CLASSICAL PRINCIPAL COMPONENT ANALYSIS

The PCA is a way of identifying patterns in data and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimensional, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data. The main advantage of PCA is that, by reducing the number of dimensions, without much loss of information. The algorithm for classical PCA is summarized given below.

Assume that the data matrix is X of size N x d. To transform X into an N x m (m<d) matrix Y,

- Centralized the data (subtract the mean).
- Calculate the d x d covariance matrix: $C = \frac{1}{N-1} X^T X$
  - $C_{i,j} = \frac{1}{N-1} \sum_{q=1}^{N} X_{q,i} X_{q,j}$
  - $C_{i,i}$ is the variance of the variable i.
  - $C_{i,j}$ is the covariance between variables i and j.
- Calculate the eigen vectors of the covariance matrix (orthonormal)
- Select m eigenvectors that correspond to the largest m eigenvalues to be the new basis.

III. ROBUST PRINCIPAL COMPONENT ANALYSIS

The standard PCA for estimating the principal components are not robust to outliers that are common in training data and that can arbitrarily bias the solution. This happens because all the energy functions and the covariance matrix are derived from a least-squares framework. One approach replaces the standard estimation of the covariance matrix, with a robust estimator of the covariance matrix. In a real world environment, especially in computer vision application it is happened that the images may contain various outliers (occlusions, motion, etc.) whose exact positions in an image are not known. Since the standard PCA is intrinsically sensitive to non-Gaussian noise, such disturbances may considerably degrade the results of the visual learning and recognition. The algorithm for robust PCA is as follows

Input: Data matrix D, number of principal axes to be estimated K.

Output: Mean vector μ, eigenvectors U, eigenvalues λ, coefficients A.

- Repeat
- Perform standard PCA on D and obtain μ', $U' \in {}^{M\times K'}$ and $A' \in {}^{K'\times N}$.
- Reconstruct the training images using μ', U' and A' and calculate the reconstruction error.
- Detect outliers considering reconstruction errors.
- Treat outliers as missing pixels and perform PCA using an algorithm for learning from incomplete data to obtain μ, U, **λ** and A from inliers only.
- Reconstruct the training images using μ, U, and A. Replace the missing pixels in D with reconstructed values.
- Until the change in the outlier set is small.

IV. WEIGHTED PRINCIPAL COMPONENT ANALYSIS

The potential drawback of the conventional PCA method is mainly based on least squares method which is not a robust one. Scokaj et al. (2005) proposed a weighted principal component analysis method to solving the problems for vision communities by considering the robustness. The algorithm for the weighted PCA is as follows.

- Estimate the weighted mean vector:

$$\mu_i = \sum_{j=1}^{N} \omega_{ij} d_{ij} \Big/ \sum_{j=1}^{N} \omega_{ij}, \; i = 1,2,\ldots, M$$

- Centre the input data around the mean:

$$\hat{X} = X - \mu 1_{1\times N}$$

- Set the elements of $U \in \Re^{M\times K}$ to random values.
- Repeat.
- E-Step:

$$a_j = \left(\left(\sqrt{W_j 1_{1\times k}}\right) U\right)\left(\sqrt{W_j} \circ \hat{X}_j\right), \; j = 1,2,\ldots., N$$

- M-Step:

$$U_i = \left(\sqrt{W_i} \circ \hat{X}_{i:}\right)\left(\left(\mathbf{1}_{k\times 1}\sqrt{W_{i:}}\right) \circ A\right), \; i = 1,2,.., M$$

- Until convergence.

V. MODIFIED PRINCIPAL COMPONENT ANALYSIS

The robust PCA uses M-estimator to detect outliers (De La Torre et al. (2001). It involves more number of iterations and computation time. To reduce the number of iteration and time, it is proposed a new algorithm which is based on Minimum Covariance Determinant estimator. The proposed method named as modified robust PCA. The main description of modified robust PCA (MRPCA) approach is, instead of M-estimator in the robust PCA algorithm, apply MCD estimator and then follow the usual algorithmic steps as done in the case of robust PCA. This MRPCA algorithm has the capability of detecting outliers in training images during eigen space learning. These outliers are then treated as missing pixels and the principal subspace is estimated from the inliers only.

A high breakdown estimation procedure, Minimum Covariance Determinant (MCD) estimator was proposed by Rousseeuw (1984). It is obtained by finding the half set of multivariate data points that gives the minimum value determination of the covariance matrix. The resulting estimator of location is the sample mean vector of the points that is the half set and the estimator of the dispersion is the sample covariance matrix of the points multiplied by an appropriate constant. The MCD estimators are intuitively appealing because a small value of the determinant corresponds to near linear dependencies of the data in the p-dimensional space that is because a small d determinant corresponds to a small Eigenvalue which suggests a near linear dependency that suggests that there is a group of points that are similar to each other. Normally in RPCA case, median and MAD (Median Absolute Deviation) taken as location and scale. But in this case of our method, we split data into groups, each groups contains 50 variables and calculate value of location and scale based on MCD procedure, after that join the location and scale value of each group together. This is the main modification in this procedure. Apart from this same procedure was carried out like RPCA.

VI. EXPERIMENTAL RESULTS

This section presents the performance of classical PCA, robust PCA, weighted PCA and the proposed MRPCA procedures on image processing.

The experiment was performed on an image sequence. The image sequence contains 506 images of size 120 × 160 pixels each in the data set which was described by Torre and Black (2001). Here, it is considered the more clarity of 20 images out of it. The objective of this experiment is to model the background, capturing the gradual illumination changes, while excluding the people that appear in the images. The original images and the extracted images with outliers under the various PCA methods along with the proposed method are displayed in the appendix.

The images considered are displayed in the first column, the reconstructed images under various methods PCA, RPCA, WPCA and the proposed MRPCA are displayed in columns (b), (c), (e), (g) respectively. The outliers detected under the methods RPCA, WPCA and the proposed MRPCA are displayed in the columns (d), (f) and (h) respectively.

The number of iterations involved and the time taken to extract the images under various PCA methods are shown in the following table.

TABLE I
TIME AND ITERATIONS INVOLVED FOR THE RECONSTRUCTION OF IMAGES

Methods	No. of Iterations	Time (in Min.)
PCA	0	0
RPCA	454	28.63
WPCA	78	12.00
MRPCA	276	17.98

It is observed that, robust and weighted principal component analysis is performing well. Also, the modified robust principal component method is equally good with RPCA, but it took minimum time to extract the image. The computational steps involved in MRPCA are tedious but it is an efficient method to extract the image with more clarity while compared with the other PCA methods. It is noted that, the number of iterations and time taken by the classical PCA is very less but very poor performance, while considering the quality of the extracted images and detecting outliers. The weighted PCA has also less in number of iterations and the time taken to extract the image while compared with the proposed method MRPCA but the extracted images are still blurred. It is concluded that the proposed method MRPCA performs well by detecting outliers/noise and extracting the image with high quality by taking less amount of time and number of iterations when compared with RPCA.

VII. CONCLUSION

PCA is one of the most widely used techniques in the context of computer vision task such as image processing, pattern recognition, signal processing and etc. The conventional PCA is based least square approach and it doesn't provide the reliable results. This paper reviewed the robust alternatives and proposed a new algorithm to perform PCA. The efficiency of the proposed algorithm, MRPCA has been performed and compared with the methods, classical PCA, robust PCA and weighted PCA on image processing, specifically, the experiment was performed on an image sequence. The objective of the experiment is to model the background, capturing the gradual illumination changes, while excluding the people that appear in the images. From the experiments, it is concluded that the proposed method MRPCA performs well by detecting outliers/noise and extracting the image with high quality by taking less amount of time when compared with RPCA. But it is still challenging of the research communities to reduce the number of iteration and timing while considering the reconstruction of the images.

ACKNOWLEDGMENT

First author convey his sincere thanks to University Grants Commission, New Delhi, India for providing financial assistance under the major research project [F.N.40-247/2011 (SR)] scheme awarded at the department of statistics, Bharathiar University, Coimbatore - 641046, Tamilnadu, India.

REFERENCES

[1] M.J. Black and A.D. Jepson, Eigen tracking: robust matching and tracking of objects using view-based representation, *International Journal of Computer Vision*, Vol. 26, Issue - 1, pp. 63-84, 1998.

[2] F. De La Torre and M.J. Black, Robust principal component analysis for computer vision. *International Conference on Computer Vision*,pp. 362-369, 2001.

[3] F. De La Torre and M.J. Black, Robust parameterized component analysis: theory and applications of 2D facial modeling, *In European Conference on Computer Vision*, pp. 653-669, 2002.

[4] F. De La Torre and M.J. Black, A frame work for robust subspace learning. *International Journal of Computer Vision*, pp. 117-142, 2002.

[5] K.R. Gabriel and S. Zamir, Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, Vol. 21, pp. 489-498, 1979.

[6] H. Hotelling, Analysis of a complex of statistical variables Into principal components. *Journal of Educational Psychology*, Vol. 24, pp. 417-441 and 498-520, 1933.

[7] P.J. Huber, *Robust Statistics*. Wiley, New York, 1981.

[8] K. Pearson, On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Vol. 2, pp. 559-572, 1901.

[9] S. Roweis, EM- Algorithms for PCA and SPCA. *In Neural Information Processing Systems*, pp. 626-632, 1997.

[10] P.J. Rousseeuw, Least median of squares regression. *Journal of the American Statistical Association*, Vol. 79, pp. 871-880, 1984.

[11] P.J. Rousseeuw and K. Van Driessen,A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, Vol. 41, pp. 212-223, 1999.

[12] F.H. Ruymagaart, A robust principal component analysis. Journal of Multivariate Analysis, Vol. 11, pp. 485-489, 1981.

[13] D. Skocaj, A. Leonardis and H. Bischof, Weighted and robust learning of subspace representations. *Pattern Recognition*, pp. 1556-1569, 2006.

[14] M. Tipping and C.M. Bishop, Probabilistic principal component analysis. *Journal of the Royal Statistical Society B*, Vol. 61, pp. 611-622, 1999.

[15] L. Xu and A. Yuille, Robust principal component analysis by self-organizing rules based on statistical physics approach. *IEEE Transactions on Neural Networks*, pp. 131-143, 1995.

[16] T.N. Yang and S.D. Wang, Robust algorithms for principal component analysis. *Pattern Recognition Letters*, 1999.

APPENDIX

Fig 1: Original image, reconstructed images under the various PCA methods along with outliers