# Performance of Robust Two-dimensional Principal Component for Classification

Dyah E. Herwindiati, Sani M. Isa, and Janson Hendryli
*Faculty of Information Technology, Tarumanagara University*
Email: herwindiati@untar.ac.id, sani.fti.untar@gmail.com, jansonhendryli@gmail.com

***Abstract* — The robust dimension reduction for classification of two dimensional data is discussed in this paper. The classification process is done with reference of original data. The classifying of class membership is not easy when more than one variable are loaded with the same information, and they can be written as a near linear combination of other variables. The standard approach to overcome this problem is dimension reduction. One of the most common forms of dimensionality reduction is the principal component analysis (PCA). The two-dimensional principal component (2DPCA) is often called a variant of principal component. The image matrices were directly treated as 2D matrices; the covariance matrix of image can be constructed directly using the original image matrices. The presence of outliers in the data has been proved to pose a serious problem in dimension reduction. The first component consisting of the greatest variation is often pushed toward the anomalous observations. The robust minimizing vector variance (MVV) combined with two dimensional projection approach is used for solving the problem. The computation experiment shows the robust method has the good performances for matrix data classification.**

***Keywords*:** 2DPCA, PCA, outlier, robust, sensitivity, vector variance, wishart distribution

## I. Introduction

CLASSIFICATION is one technique of data mining to predict an object to a certain class based on information in one or more characteristics of data. As with most data mining solutions, a classification usually comes with a degree of certainty. It might be the probability of the object belonging to the class or it might be some other measure of how closely the object resembles other examples from that class. This paper discusses the new measure of classification by combining of two advantages from two approaches; the two-dimensional (2D) projection approach and the robust approach.

The principal components analysis (PCA) is primarily a data analytic technique describing the variance covariance structure through a linear transformation of the original variables, Jollife [4]. The technique is the most popular among the dimension reduction analysis which is used to transform the original set of variables into a smaller set of linear combinations that accounts for most of the original set variance. The first principal component is the combination of variables that explains the greatest amount of variation. One disadvantage of PCA is the high computation.

Yang et. al [6] proposed the two-dimensional Principal Component (2DPCA) for reducing computational time of standard PCA. The 2DPCA is often called as a variant of principal component (PCA). In the 2DPCA, the image matrices were directly treated as 2D matrices; the images do not need to be transformed into a vector so that the covariance matrix of image can be constructed directly using the original image matrices. Compared with PCA, 2DPCA is more efficient.

The decomposed information variation of classical PCA and 2DPCA becomes pointless if outliers are present in the data. The decomposed classical covariance matrix is very sensitive to outlying observations. The first component consisting of the greatest variation is often pushed toward the anomalous observations. Anscombe [2] categorized outliers into two majors: those arising from errors in the data and those arising from the inherent variability of the data. The several causes of data errors are the experimental error, human error, and instrument error. An outlier is often difficult to be identified through visual inspection without the analytic tools. The difficulty becomes harder when data size is in larger dimension.

The classical estimates such as the sample mean and covariance are very sensitive to outlier, even by a single outlier. One or more outliers can significantly shift the mean and increase the dispersion of variance. The presence of outliers can lead to inflated error rates and substantial distortions of parameter. Robust approach is one method believed to be able to detect outliers well. In this paper, author introduces the robust 2DPCA for handling outlier in the process of 2D projection.

The robust method deals with a very real problem in statistical applications, the robust estimator provide a good solution when the data contain outliers. The word 'robust' is loaded with many–sometimes inconsistent–connotations. Major goal of robust statistics is to develop methods that are robust against the possibility that one or several unannounced outliers may occur anywhere in the data, Hampel [3].

 978-1-4799-8075-8/14/$31.00 

There are some robust criteria proposed to get an effective estimator. The most well known criterion is to minimize the volume of ellipsoid of a parallelotop. The minimum covariance determinant (MCD) is a robust high break down point method using minimum volume ellipsoid, Rousseuw [8]. MCD has an important role in the application of data mining, but the one lack property of MCD is the determinant of covariance matrix equal zero is not certainly implies that a random vector $\vec{X}$ is of degenerate distribution in the mean vector $\vec{\mu}$. MCD approach requires a condition that the covariance matrix must be non singular. Herwindiati et. al [1] proposed robust minimum vector variance to overcome the difficulties of MCD.

The minimum vector variance (MVV) is a robust method that uses the minimum of a square of length of a parallelotope diagonal to estimate the location and scatter. MVV is robust high breakdown point generated from vector variance (VV) as multivariate dispersion [1]. The objective of paper is to propose the robust minimizing vector variance in 2D projection process for classification of $m \times p$ arbitrary matrix data. The aspect of theoretical distribution for sensitivity is also discussed to see the robustness of measure.

## II. Classification of Matrix Data using the Classical 2DPCA

Two dimensional Principal Component (2DPCA) was proposed by Yang et.al [6]. The method using the projection technique is developed for the gray scale face recognition. Though the 2DPCA is often called as a variant of principal component (PCA), the 2DPCA has two important benefits over PCA: it is easier to evaluate the covariance matrix and it uses less time for determining the eigenvectors. In the 2DPCA, the image matrices were directly treated as 2D matrices; the images do not need to be transformed into a vector so that the covariance matrix of image can be constructed directly using the original image matrices.

Consider $X_1, X_2, \cdots$ is a $m \times p$ random image matrix, let $\vec{V}$ is an $p$-dimensional unitary column vector, the idea of 2DPCA is to project $X$ onto $\vec{V}$ by linear transformation

$$\vec{Y} = X\vec{V} \qquad (1)$$

Define the image covariance matrix $S_M = E\left[(X - EX)^T (X - EX)\right]$ which is a $p \times p$ non negative definite matrix. The covariance matrix of projected feature of sample is defined as $S_X = \vec{V}' E\left[(X - EX)' (X - EX)\right]\vec{V} = \vec{V}' S_M \vec{V}$.

Suppose there are $N$ image matrices $\{X_i\}$, $i = 1, 2, \cdots$ and denote the average image as $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$, then $S_M$ can be evaluated by

$$S_X = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^T \left(X_i - \bar{X}\right) \qquad (2)$$

To have the optimal projection direction of 2DPCA, $S_X$ has the important rule, the $\vec{V}_{opt}$ is the eigenvector of $S_X$ corresponding to the largest eigenvalue. A set orthonormal projection directions $\vec{V}_1, \vec{V}_2, \cdots$ are the orthonormal eigenvector of $S_X$ corresponding to the $d$ largest eigenvalues, i.e. $\vec{V}_{opt} = \lfloor \vec{V}_1, \vec{V}_2, \cdots \rfloor$. Projecting a matrix X onto $\vec{V}_{opt}$ is

$$\vec{Y}_k = X\vec{V}_k, \quad k = 1, 2, \cdots \qquad (3)$$

The descriptions of formula (1) until formula (3) give us the comprehension that the 2DPCA takes to less time than PCA for classification, because the size of $S_X$ is only $p \times p$.

Principal component analysis (PCA) is well established dimension reduction technique. To differ from 2DPCA, all of the 2D data must be previously transformed into 1D vector before the data will be processed by PCA approach. The transformation leads to a high dimensional vector space. Consider $X_1, X_2, \cdots$ is a $m \times p$ random image matrix, the $N$ image matrices were transformed into 1D vector $1 \times mp$. The dimensional of PCA covariance matrix $S_C$ is $mp$ by $mp$. The large size covariance matrix $S_C$ makes the computation becomes time consuming.

## III. Classification of Matrix Data using the Robust 2DPCA

In this section author will discuss the robust 2DPCA using the measure minimizing vector variance (MVV). The robust 2DPCA is primarily a robust approach describing the variance covariance structure through a linear transformation of the original variables. The technique is a useful device for representing a set of variables by a much smaller set of composite variables that account for much of the variance among the set of original variables. The data reduction based on the classical approach becomes unreliable if outliers are present in the data. The decomposed classical covariance matrix is very sensitive to outlying observations. The first component consisting of the greatest variation is often pushed toward the anomalous observations.

Minimum Vector Variance (MVV) is method by using the minimization of vector variance (VV)

 978-1-4799-8075-8/14/$31.00 

criteria to identify the outliers. The estimator MVV for the pair $(\vec{\mu}, \Sigma)$ is the pair $(T_{MVV}, C_{MVV})$ giving minimum vector variance. The MVV estimator can be computed by the following description, given random samples $\vec{X}_1, \vec{X}_2, \cdots$ of dimension $n$ taken from a $p$-variate distribution of location parameter $\vec{\mu}$ and a positive definite covariance matrix $\Sigma$. Suppose $T_{MVV}$ and $C_{MVV}$ are MVV estimators for location parameters and covariance matrix. Both estimators are defined based on a set $H \subseteq X$ consist of $h = \left[\frac{n+p+1}{2}\right]$ data points which gives covariance matrix $C_{MVV}$ of minimum $Tr(C^2_{MVV})$ among all possible $h$ data, see Herwindiati et.al [1]. Then,

$$T_{MVV} = \frac{1}{n}\sum_{i \in H} \vec{X}_i \qquad (4)$$

$$C_{MVV} = \frac{1}{n}\sum_{i \in H}(\vec{X}_i - T_{MVV})(\vec{X}_i - T_{MVV})' \qquad (5)$$

The algorithm of MVV robust 2DPCA has no significant difference with MVV robust PCA except for the criterion projection. The proposed method is not focused on face detection, the paper is purposed to classify a general problem on a matrix data. The algorithm of the MVV robust 2DPCA has three stages. Suppose $X_1, X_2, \cdots$ is a $m \times p$ random image matrix.

Stage 1 Start with a construction the covariance matrix by using the $N$ original two dimensional (2D) matrices. Find the orthonormal eigenvectors corresponding to the $d$ largest eigenvalues $S_X$, $\vec{V}_{opt} = [\vec{V}_1, \vec{V}_2, \cdots]$. Projecting a matrix X onto $\vec{V}_{opt}$ is $\vec{Y}_k = X\vec{V}_k$, $k = 1, 2, \cdots$.

Stage 2 Estimate the location and covariance matrix of projected matrix $X_{m \times d}$ using MVV robust approach.

1. Let $H_{old}$ be an arbitrary subset containing $h = \left[\frac{n+k+1}{2}\right]$ matrix data points. Compute the average matrix as $\bar{X}_{H_{old}}$ and covariance matrix $S_{H_{old}}$ of all observations belonging to $H_{old}$. Then calculate $B_{m \times k} = (X - \bar{X}_{H_{old}})$, $k = 1, 2, \cdots$
2. Compute $d^2_{H_{old}}(i) = \vec{D}' S^{-1}_{H_{old}} \vec{D}$, for all $i = 1, 2, \ldots, N$ where $\vec{D}_{1 \times d}$ is defined as mean of $m$ rows in each $k$ column $k = 1, 2, \cdots$
3. Sort these distances in increasing order
4. Define $H_{new} = \{\vec{X}_{\pi(1)}, \vec{X}_{\pi(2)}, \cdots$
5. Calculate $\bar{X}_{H_{new}}$, $S_{H_{new}}$, and $d^2_{H_{new}}(i)$
6. If $Tr(S^2_{H_{new}}) = 0$, repeat step 1 to 5

   If $Tr(S^2_{H_{new}}) = Tr(S^2_{H_{old}})$, the process is stopped.

   Otherwise, the process is continued until the $r$-th iteration if

$$Tr(S_1^2) \geq Tr(S_2^2) \geq Tr(S_3^2) \geq \cdots \quad Tr(S_r^2) = Tr(S_{r+1}^2)$$

Thus, we get

$$Tr(S_1^2) \geq Tr(S_2^2) \geq Tr(S_3^2) \geq \cdots \quad Tr(S_r^2) = Tr(S_{r+1}^2)$$

Stage 3 Classify the matrix data based on robust MVV distance

$$d^2_{MVV}(i) = \vec{D}_{MVV} S^{-1}_{H_{old}} \vec{D}_{MVV}, \text{ for all } i = 1, 2, \ldots, N \qquad (6)$$

## IV. Comparison Result of Classic 2DPCA and Robust 2DPCA

To compare the classification process of classics and MVV robust 2DPCA, we do several experiments.

### A. The Classification of Two Objects

Starting with the classification of two object types, there are (30x30) pixels of 50 grass images and 25 ocean images.

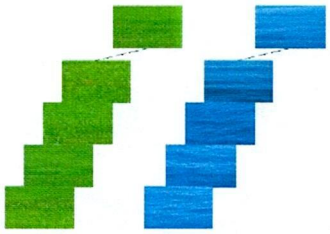

Fig. 1. Two Objects for Classification: Grass and Ocean.

The extraction of object features based on RGB color spaces are to be used as elements in the classification. The classical classification of 2DPCA is unable to hold the two significant variations of grass and ocean; consequently, the objects are not separated well (see Figure 2A). The classic 2DPCA is not robust to outlier. The occurrence of one or more outliers can shift a data center $\vec{\bar{X}}$ to keep away from a location of main data, so that the masking effect is not avoidable.

 978-1-4799-8075-8/14/$31.00 

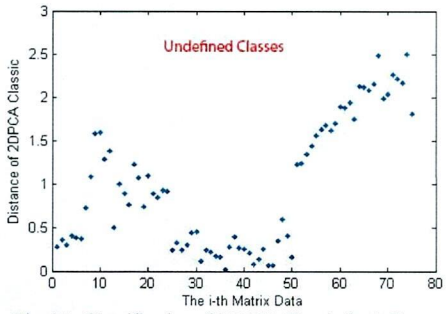


Fig. 2A. Classification of 2DPCA Classic for 2 Classes

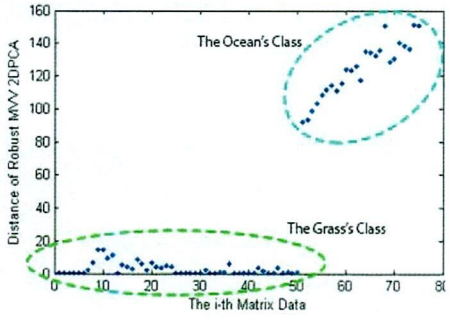


Fig. 2B. Classification of Robust 2DPCA for 2 Classes

The MVV robust 2DPCA is going to be used for improving classification process. The MVV robust 2DPCA gives a better classification (see Figure 2B).

### *B. The Classification of Three and Four Objects*

More than two object types are to be tried for classification. The authors want to know how the number of classes affects the classification process.

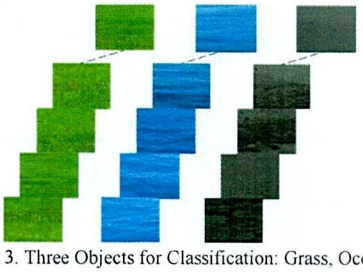

Fig. 3. Three Objects for Classification: Grass, Ocean and Sand

The three and four objects have no significantly influence over MVV 2DPCA in the classification process. The new characteristics are still able to be classified definitely (see Figure 5B and 6B).

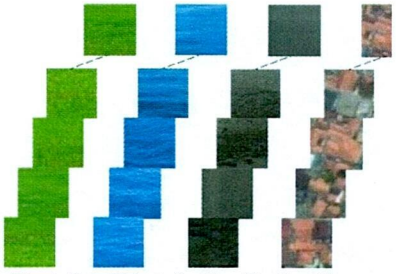

Fig. 4. Four Objects for Classification: Grass, Ocean, Sand and Roof

In contradiction with the approach, we see undefined class, see Figure 5A and 6A. The dispersion of classics 2DPCA becomes bigger after the characteristics of new objects are added in a data set.

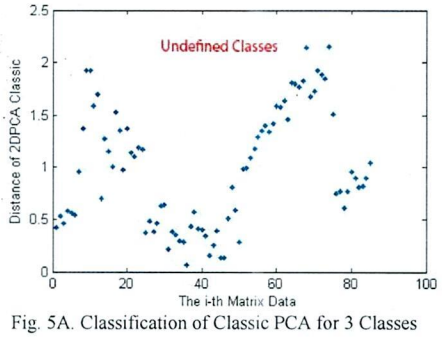


Fig. 5A. Classification of Classic PCA for 3 Classes

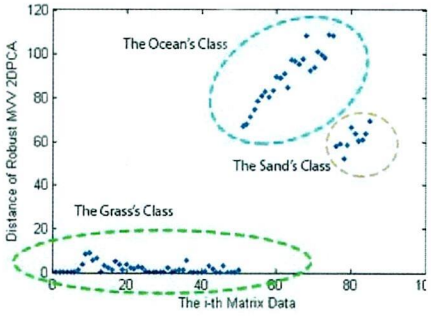


Fig. 5B. Classification of Robust PCA for 3 Classes

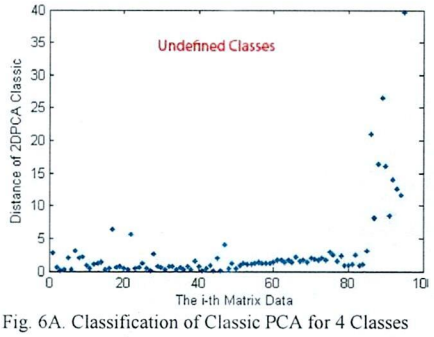


Fig. 6A. Classification of Classic PCA for 4 Classes

 978-1-4799-8075-8/14/$31.00 

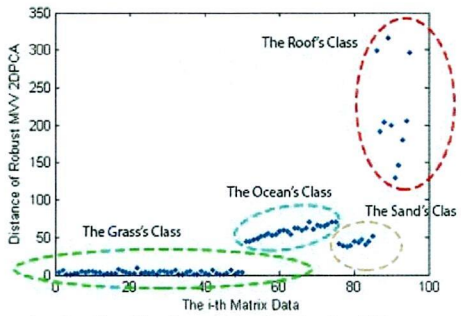


Fig. 6B. Classification of Classic PCA for 4 Classes

The outcomes of experiment tell us that the robust MVV 2DPCA is a powerful approach for classification, even when new objects are added in the dataset.

## V. The Sensitivity of Classical and Robust Method to Outlier

The good performance of robust methods is exhibited in Section IV. The main problem of classical method is that the location estimator shifts closer to outliers. The occurrence of one or more outliers shifts the mean vector toward outliers and the covariance matrix becomes inflated.

Outlier can be considered as an influential observation. An observation is called influential if its deletion would cause major changes in estimates. The influential observation can significantly change an estimator.

The estimator is said to be insensitive if there is no significant change due to removal of outlier. There are many ways to measure the sensitivity; this paper brings simple discussion, both on computation and the theoretical distribution.

**Theorem:**

Suppose $\vec{X}_1, \vec{X}_2, \cdots$ are random sample of size $n$ of a probability distribution having mean $\vec{\theta} \in$ where $p \geq 2$ is an integer and the covariance matrix $\Sigma$ is of positive definite. Then the random vector

$$\vec{Y}_n = \sqrt{n}\mathbf{C}^{-1}\left(\vec{\bar{X}}_n - \vec{\mu}\right) \rightarrow \sim N_p\left(\vec{0}, \mathbf{I}_p\right) \qquad (7)$$

where $\vec{X}_n = \sum \vec{X}_j$, $\mathbf{CC^t} = \mathbf{\Sigma}$.

Consider data set $X_n = \left\{\vec{X}_1, \vec{X}_2, \cdots\right\}$ of $p$-variate, the scatter matrix of sample $\boldsymbol{A}$ is

$$\boldsymbol{A} = \sum_{j=1}^{n}\left(\vec{X}_j - X\right)\left(\vec{X}_j - X\right) \qquad (8)$$

where $\vec{\bar{X}} = \frac{1}{n}\sum \vec{X}_j$, $\vec{\bar{X}}$ is the sample's mean vector.

From equation (5), the scatter matrix $\mathbf{A}$ is of Wishart distribution with parameter $\mathbf{\Sigma}$ and the degree of freedom $n-1$, written as $\mathbf{A} \sim W_p\left(\mathbf{\Sigma}, n-1\right)$, $\boldsymbol{A}$ is independent of $\vec{\bar{X}}$.

Define $\boldsymbol{A}_{-i}$ the scatter matrix removing the $i^{\text{th}}$ observation, say the $i^{\text{th}}$ observation is an outlying observation. The scatter matrix $\boldsymbol{A}_{-i}$ is formulated as

$$\boldsymbol{A}_{-i} = \sum_{i \neq j=1}^{n}\left(\vec{X}_j - X_{-i}\right)\left(\vec{X}_j - X_{-i}\right) \qquad (9)$$

where $\vec{\bar{X}}_{-i} = \frac{1}{n-1}\sum_{i \neq j=1}^{n} \vec{X}_j$.

The scatter matrix $\boldsymbol{A}_{-i}$ is of Wishart distribution with parameter $\mathbf{\Sigma}$ and the degree of freedom $n-2$, $\boldsymbol{A}_{-i} \sim W_p\left(\mathbf{\Sigma}, n-2\right)$. Based on two formulas, the ratio of scatter matrix as the consequence of removal the $i^{\text{th}}$ observation is given by

$$\boldsymbol{R}_i = \frac{|\boldsymbol{A}_{-i}|}{|\boldsymbol{A}|} = \frac{|\boldsymbol{A}_{-i}|}{|\boldsymbol{A}_{-i} + \vec{b}\vec{b}'|} \qquad (10)$$

$$\vec{b}_i = \sqrt{\frac{\cdot\cdot}{h^*}}\left(\vec{\bar{X}}_i - X\right)$$

and $\boldsymbol{R}_i$ can be shown of distribution $beta\left(\frac{n-p-1}{2}, \frac{p}{2}\right)$. The ratio $\boldsymbol{R}_i$ is close to 1 means that no significant change due to the removal of that observation.

In this case, the estimator is said to be insensitive to an outlier when $\boldsymbol{R}_i > beta\left(\frac{n-p-1}{2}, \frac{p}{2}\right)$.

In application on data mining, it is often found problems of more than one outlier, so the masking effect is unavoidable. This section discussed the sensitivity of estimator when there is $k$ outliers ($k > 1$).

Suppose the group consists of $k$ outliers, the scatter matrix $\boldsymbol{A}_{-I}$, as a consequence of the removal of $\boldsymbol{I}^{\text{th}}$ group, is of distribution $\boldsymbol{A}_{-I} \sim W_p\left(\mathbf{\Sigma}, m\right)$. Matrix $\boldsymbol{A}$ can be decomposed as $\boldsymbol{A} = \boldsymbol{A}_{-I} + \boldsymbol{B}$ and $\boldsymbol{B} = \vec{X}_k \vec{X}_k'$. The distribution of $\boldsymbol{B}$ is $W_p\left(\mathbf{\Sigma}, k\right)$.

Similar with the case of single outlier, the ratio of scatter matrix as a consequence of removal of the observation on the group $\boldsymbol{I}$ can be formulated as

 978-1-4799-8075-8/14/$31.00 

$$\boldsymbol{R}_I = \frac{|\boldsymbol{A}_{-I}|}{|\boldsymbol{A}|} = \frac{|\boldsymbol{A}_{-I}|}{|\boldsymbol{A}_{-I} + \boldsymbol{B}|} \tag{11}$$

Mardia et al [7] stated that $\boldsymbol{R}_I$ has Wilk's Lambda distribution with parameter $p, m, k$, and $m = n-(k+1)$ or $\boldsymbol{R}_I \sim \Lambda(p,m,k)$. The Wilk's Lambda distribution can be approximated by

$$\Lambda(p,m,k) \sim \prod_{i=1}^{h} u_i$$

$$u_i \sim beta\left(\frac{m-p+i}{2}, \frac{p}{2}\right), \tag{12}$$

$$i = 1, 2, \ldots, k$$

$\boldsymbol{R}_I$ close to 1 means that there is no significant change due to the removal of $k$ observations on the group $\boldsymbol{I}$. The estimator is said to be insensitive to $k$ outliers when

$$\boldsymbol{R}_I > \prod_{i=1}^{h} u_i \tag{13}$$

The distribution of classical approach is well known and it is different with the robust approach. The distribution of robust is not easy to be composed. Usually we have to do the simulation approach to get the distribution. In the section will be discussed the sensitivity and the approximated distribution of robust approach.

Let dataset $X_n = \{\vec{X}_1, \vec{X}_2, \cdots \vec{\ }_{,}$ of $p-$ variate observations. If observations taken from it a subset $H \subseteq X$ consist of $h$ data points, then $\vec{X}_1, \vec{X}_2, \cdots \vec{\ }_{..}$ are random sample of size $h$ and of distribution $N_p(\vec{\mu}, \Sigma)$, $h$ assumed as $h = \left[\frac{n+p+1}{2}\right]$. The location and scale estimator can be computed as,

$$\vec{\bar{X}}^R = \frac{1}{h}\sum_{i \in h} \vec{X}_i \tag{14}$$

$$S^R = \sum_{i \in h}\left(\vec{X}_i - X\right)\left(\vec{X}_i - X\right)$$

Based on limit central theory, if $\vec{X}_1, \vec{X}_2, \cdots \vec{\ }_{..} \sim N_{p\ ,\ \ }$ then the distribution of $S^R$ can be approximated by $m\,c^{-1}S^R \sim W(m, \Sigma)$ [6]. It means that

$$\boldsymbol{A}^R = \sum_{j \in h}\left(\vec{X}_j - X^{\wedge}\right)\left(\vec{X}_j - X^{\wedge}\right) = c^{-1}\frac{S^R}{m} \tag{15}$$

$$\text{and } A^R \sim \frac{1}{m} W(m, \Sigma) \tag{16}$$

The $c$ can be approximated by 1. Hardin and Rocke [5] predicted the values of $m$ by simulation approach The predictions are listed in the Table I.

TABLE I
THE PREDICTION OF $M$

| Dimension and Size | $m_{pred}$ |
|---|---|
| $p=5, n=50$ | 12.89 |
| $p=10, n=100$ | 33.13 |
| $p=10, n=500$ | 126.71 |
| $p=20, n=1000$ | 298.35 |

Based on the formulas, $\boldsymbol{R}_i^R = \frac{|\boldsymbol{A}_{-i}^R|}{|\boldsymbol{A}^R|}$ approximated by

$$\boldsymbol{R}_i^R \sim \frac{mp}{m(m-p+1)} F_{m,\,m-p+1} \tag{17}$$

The estimator is said to be insensitive to k outliers on the group $\boldsymbol{I}$ when $\boldsymbol{R}_I > \prod_{i=1}^{h} u_i$.

The section illustrates the sensitivity of classical and robust measure $k > 1$ outliers. For illustration, let the multivariate data having size $n = 50$; $p = 5$. Data contain $k = 3$ outliers which are far from a bulk of data. The sensitivities are measured by ratio of scatter matrix $\boldsymbol{R}_I^R = \frac{|\boldsymbol{A}_{-i}^R|}{|\boldsymbol{A}^R|}$. The ratio of classical and robust approaches is computed by simulation as shown in Table II.

TABLE II
RATIO $\boldsymbol{R}_I$ BY REMOVING $K$=3 OUTLIERS

| Value | Method | |
|---|---|---|
| | Classical Method | Robust Method |
| $\boldsymbol{A}$ | 27.1628 | 0.040128 |
| $\boldsymbol{A}_{-I}$ | 0.807614 | 0.039319 |
| $\boldsymbol{R}_{-I}$ | 0.029732 | 0.979824 |
| Cut off | 0.999722 | 0.096965 |
| Sensitivity to outliers | Very sensitive* | Insensitive |

The removing outliers causes a serious problem on the classical estimator. The value of estimator is very sensitive to outliers. It can be seen in the table 2, the estimator becomes to be inflated when the outliers 'present' on the data set.

The reverse of classical sensitivity, the ratio of MVV robust estimator is almost 1, though $k = 3$ outliers are removed.

## VI. REMARK

The MVV robust estimator is not sensitive from 'presenting' or removing outlier. On the classification processes, the MVV robust 2DPCA is an effective method. The outcomes of all experiments show the

 978-1-4799-8075-8/14/$31.00 

MVV 2DPCA is powerful approach to classify the several objects.

## REFERENCES


[1] D. E. Herwindiati, M. A. Djauhari, and M. Mashuri, "Robust multivariate outlier labeling", *J. Communication in Statistics Simulation and Computation*, vol. 36, no. 6, 2007

[2] F. J. Anscombe, "Rejection of outliers", *Technometrics*, 2, pp. 123-147, 1960

[3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseuw, W. A. Stahel, *Robust Statistics*, New York: John Wiley, 1985

[4] I. T. Jolliffe, *Principal Component Analysis*, Springer Verlag, 1986

[5] J. Hardin and D. M. Rocke, "The distribution of robust distance", *J. of Computation and Graphical Statistics*, 14, pp. 928-946, 2005

[6] J. Yang, D. Zhang, A. F. Frangi, J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131-137, 2004

[7] K. V. Mardia, J. T. Kent, J. M. Bibby, *Multivariate Analysis*, London: Academic Press, 1979

[8] M. A. Djauhari, "Improved monitoring of multivariate process Variability", *Journal of Quality Technology*, 37(1), pp. 32-39, 2005

[9] P. J. Rousseuw, "Multivariate estimation with high breakdown point", *Mathematical Statistics and Applications*, B, D. Reidel Publishing Company, pp. 283-297, 1985

[10] P. J. Rousseuw and K. van Driessen. "A fast algorithm for the minimum covariance determinant estimator", *J. Technometrics*, 41, pp. 212-223, 1999

[11] S. S. Wilks, "Multivariate statistical outliers", *J. Sankya A*, 25, pp. 407-426, 1963

[12] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis, 2nd Edition*, New York: John Wiley, 1984

[13] V. Barnett and T. Lewis, *Outliers in Statistical Data, 2nd Edition*, New York: John Wiley, 1984


 978-1-4799-8075-8/14/$31.00 