Robust Principal Component Analysis Based On L_{1-2} Metric

Fanlong Zhang, Zhangjing Yang, Minghua Wan, Guowei Yang
School of Technology, Nanjing Audit University
211815, Nanjing, China
csfzhang@126.com, yzj@nau.edu.cn, wmh36@sina.com, ygw_ustb@163.com

Abstract

Robust principal component analysis (RPCA) is a new emerging method for exact recovery of corrupted low-rank matrices. Given a data matrix, RPCA can decompose it into the sum of a low-rank matrix and a sparse matrix exactly by minimizing a weighted combination of the nuclear norm and the L_1 norm. It assumes that the error matrix is sparse, and metric it by L_1 norm. However, L_1 norm often leads to bias estimation and the solution is not as accurate as desired. Recently the difference of L_1 and L_2 norms, called L_{1-2} metric, is proposed as the approximation to the L_0 norm. Motivated by the L_{1-2} metric's better approximation to the L_0 norm than the convex L_1 norm, this paper presents a method called robust principal component analysis based on L_{1-2} metric (RPCA-L_{1-2}) for recovering the corrupted data. This method measures the data error by the L_{1-2} metric. Moreover, RPCA- L_{1-2} is solved by DC (difference of convex functions) programming. Extensive experiments on removing occlusion from face images and background modeling from surveillance videos demonstrate the effectiveness of the proposed methods

1. Introduction

Principal component analysis (PCA) [1] is widely investigated and applied in pattern recognition and machine learning for subspace learning and feature extraction. PCA, however, is sensitive to outliers. To overcome the limitations of PCA, a surge of robust principal component analysis methods have been proposed. Wright et al. recently established a robust principal component analysis (RPCA) [2] [3] method, which assumes the error matrix is sparse and the clean data matrix is low rank. Under the restricted isometry property (RIP) condition, RPCA can decompose the corrupted data into the sum of a low-rank matrix and a sparse matrix exactly by minimizing a weighted combination of the nuclear norm and the L_1 norm. As an important extension of RPCA, the low-rank representation (LRR) [4], [5] was presented to segment subspace from a union of multiple linear subspaces. LRR sought the lowest rank representation among all the candidates that represent all vectors as the linear combinations of the basis vectors in a dictionary. Like RPCA, LRR also assumes the error term is sparse.

Most of the abovementioned methods characterize the error via L_1 or L_2 norm, which both are convex regularizers. However, convex regularizers often lead to inaccurate solution [6]. As a result, many nonconvex regularizers are designed, such as capped-L_1 norm [6], L_p norm [7], and log-sum-penalty [8]. Recently the difference of L_1 and L_2 norms [9] [10], called L_{1-2} metric, is proposed as the nonconvex regularizer to the L_0 norm. L_{1-2} metric is nonconvex yet Lipschitz continuous. The computation results [10] show that even if the RIP condition is unsatisfying, the L_{1-2} metric can work well than existing nonconvex regularizers.

Inspired by the L_{1-2} metric's better approximation to the L_0 norm, this paper presents a method called robust principal component analysis based on L_{1-2} metric (RPCA-L_{1-2}) for recovering the corrupted data. The RPCA-L_{1-2} measures the data error by the L_{1-2} metric, instead of L_1 metric in RPCA.

Although the L_{1-2} metric is non-convex, it can be decomposed into the difference of two convex functions. Then the DC programming [11] [12] can be employed to solve our model. The "DC" means "difference of convex functions". DC programming is a special kind of optimization method, whose objective function can be decomposed into the difference of two convex functions. In [10], the DC programming is also employed for solving compressed sensing based on L_{1-2} metric.

The contributions include two aspects. (1) A robust data recovery model is proposed, called RPCA-L_{1-2}. The motivation is that the L_{1-2} metric is better approximation to the L_0 norm than L_1 norm. (2) The DC programming is employed to solve proposed model. DC algorithm decomposes the original problem into series RPCA problems, which can be solved efficiently by inexact augmented Lagrange multiplier algorithms (inexact ALM) [11].

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 presents our model and corresponding algorithm. Section 4 reports experimental results. Section 5 offers conclusions.

2. Related works

Given a data set $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_s]$, where each $\mathbf{x}_i$ is a sample. The nuclear norm of the matrix $\mathbf{X}$ is defined by $\|\mathbf{X}\|_*=\sum_i \sigma_i$, which is the sum of the singular values of $\mathbf{X}$. Besides, the L_2 and L_1 norms of a matrix $\mathbf{X}$ are defined by $\|\mathbf{X}\|_F=\sqrt{\sum_{i,j}(\mathbf{X}_{ij})^2}$, $\|\mathbf{X}\|_1=\sum_{i,j}|\mathbf{X}_{ij}|$, respectively, where $\mathbf{X}_{ij}$ means the (i,j)-th entry.

2.1. RPCA

The data $\mathbf{X}$ is usually corrupted. RPCA tries to decompose $\mathbf{X}$ into two matrices $\mathbf{D}$ and $\mathbf{E}$, where the matrix D is supposed to have low rank and $\mathbf{E}$ is supposed to be sparse. The decomposition model is given by

$$\min_{\mathbf{D},\mathbf{E}} \|\mathbf{D}\|_* + \lambda\|\mathbf{E}\|_1 \quad \text{s.t.} \quad \mathbf{X}=\mathbf{D}+\mathbf{E}. \tag{1}$$

There exist many algorithms for solving the RPCA model, such as inexact augmented Lagrange multiplier algorithms (inexact ALM) [16] and accelerated proximal gradient approach [11].

2.2. L1-2 metric

The vector's $l_{1\text{-}2}$ metric was first addressed in [9] in the context of nonnegative least squares problems, and then applied to compressed sensing problems in [10]. For a vector $\mathbf{x}$, it□s $l_{1\text{-}2}$ metric is given by $\|\mathbf{x}\|_1-\|\mathbf{x}\|_2$. Similarly, for matrix X, one can define its $L_{1\text{-}2}$ metric by:

$$\|\mathbf{X}\|_{1-2} \triangleq \|\mathbf{X}\|_1 - \|\mathbf{X}\|_F. \tag{2}$$

2.3. DC programming

The "DC" means "difference of convex functions". DC programming is a special kind of optimization, whose objective function can be decomposed into the difference of two convex functions. DC programming takes the form:

$$\min\{f(x)=g(x)-h(x): x\in R^n\} \tag{3}$$

where g, h are convex functions. Such a function f is called DC function, and g-h a DC decomposition of f while g and h are DC components of f. The construction of DC algorithm involves the DC components g and h but not the function f itself. The DC algorithm is summarized in Algorithm 1, in which the first-order approximation is used to substitute the non-convex part.

3. Robust principal component analysis based on L1-2 metric

3.1. Model

Inspired by the $l_{1\text{-}2}$ metric of vector better approximation to the l_0 norm in compressed sensing problems [10], this paper extends the $l_{1\text{-}2}$ metric of vector to $L_{1\text{-}2}$ metric of matrix and apply into recovering the corrupted data. The proposed method is called robust principal component analysis based on $L_{1\text{-}2}$ metric (RPCA-$L_{1\text{-}2}$) formulated by (4).

$$\min_{\mathbf{D},\mathbf{E}} \|\mathbf{D}\|_* + \lambda\|\mathbf{E}\|_{1-2} \quad \text{s.t.} \quad \mathbf{X}=\mathbf{D}+\mathbf{E} \tag{4}$$

The RPCA-$L_{1\text{-}2}$ measures the data error by the $L_{1\text{-}2}$ metric, instead of L_1 metric in RPCA.

3.2. Algorithm

The optimization (4) is equivalent to

$$\min_{\mathbf{E}} F(\mathbf{E})-G(\mathbf{E}), \tag{5}$$

where $F(\mathbf{E})$ and $G(\mathbf{E})$ are two convex functions defined by

$$\begin{cases} F(\mathbf{E}) = \|\mathbf{X}-\mathbf{E}\|_* + \lambda\|\mathbf{E}\|_1 \\ G(\mathbf{E}) = \lambda\|\mathbf{E}\|_F \end{cases}. \tag{6}$$

Here F, G are convex functions. The DC algorithm can be employed for solving (5). Corresponding algorithm is summarized in Algorithm 2.

Algorithm 1 General form of DC algorithm

Input: x^0, k=0;
Repeat
1 Compute: $y^k \in \partial h(x^k)$;
2 Compute: $x^{k+1} \in \arg\min\{g(x)-h(x^k)-\langle x-x^k, y^k\rangle : x\in R^n\}$
Until convergence.
3 Output: x^{k+1}.

Algorithm 2 DC algorithm for RPCA-$L_{1\text{-}2}$

Input: $\mathbf{E}^0$, k=0;
Repeat
1 Compute:

$$\partial\lambda\|\mathbf{E}^k\|_F = \begin{cases} \lambda\frac{\mathbf{E}^k}{\|\mathbf{E}^k\|_F}, & \text{if } \mathbf{E}^k \neq 0 \\ 0, & \text{if } \mathbf{E}^k = 0 \end{cases}; \tag{7}$$

2 Compute:

$$\mathbf{E}^{k+1} = \arg\min_{\mathbf{E}} F(\mathbf{E}) - \mathrm{tr}\left(\mathbf{E}^T\partial\lambda\|\mathbf{E}^k\|_F\right) \tag{8}$$

Until convergence.
3 Output: $\mathbf{E}^{k+1}$.

In step k, if $\mathbf{E}^k = 0$, then (8) is the RPCA, which can be solved by the inexact augmented Lagrange multiplier algorithms (inexact ALM).

If $\mathbf{E}^k \neq 0$, the problem (8) also can be solved by inexact ALM. In this case, the augmented Lagrange function of (8) is:

$$\begin{aligned} L_\mu(\mathbf{D},\mathbf{E},\mathbf{Y}) &= \|\mathbf{D}\|_* + \lambda\|\mathbf{E}\|_1 - \mathrm{tr}\left(\mathbf{E}^T \partial\lambda \|\mathbf{E}^k\|_F\right) \\ &+ Tr(\mathbf{Y}^T(\mathbf{X}-\mathbf{D}-\mathbf{E})) + \tfrac{\mu}{2}\|\mathbf{X}-\mathbf{D}-\mathbf{E}\|_F^2 \end{aligned}, \quad (9)$$

where $\mathbf{Y}$ is the Lagrange multiplier and μ is the penalty parameter. Then (8) can be solved by alternately updating $\mathbf{D},\mathbf{E},\mathbf{Y}$, respectively. The updating of $\mathbf{D}$ and $\mathbf{Y}$ coincide with RPCA. When fix $\mathbf{D}$ and $\mathbf{Y}$, the $\mathbf{E}$ can be updated by optimizing augmented Lagrange function. The optimization can be written as (10).

$$\begin{aligned} \min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 &- Tr\left[\left(\mathbf{Y}+\partial\lambda\|\mathbf{E}^k\|_F\right)^T \mathbf{E}\right] \\ &+ \tfrac{\mu}{2}\|\mathbf{X}-\mathbf{D}-\mathbf{E}\|_F^2 \end{aligned}, \quad (10)$$

which can be equivalently rewritten as:

$$\min_{\mathbf{E}} \lambda\|\mathbf{E}\|_1 + \tfrac{\mu}{2}\left\|\mathbf{X}-\mathbf{D}+\tfrac{1}{\mu}\left(\mathbf{Y}+\partial\lambda\|\mathbf{E}^k\|_F\right)-\mathbf{E}\right\|_F^2 . \quad (11)$$

The solution of (11) can be given by

$$S_{\frac{\lambda}{\mu}}\left(\mathbf{X}-\mathbf{D}+\tfrac{1}{\mu}\left(\mathbf{Y}+\partial\lambda\|\mathbf{E}^k\|_F\right)\right), \quad (12)$$

where $S_\varepsilon(\cdot)$ is soft-thresholding operator defined by

$$S_\varepsilon(x) \triangleq \begin{cases} x-\varepsilon, & \text{if } x>\varepsilon, \\ x+\varepsilon, & \text{if } x<-\varepsilon, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

4. Experiments

The proposed RPCA-$L_{1\text{-}2}$ is applied in recovering corrupted image and background modeling for performance evaluation. For comparison, we select the standard PCA, EPCA [13], and RPCA.

4.1. Recovering corrupted face image

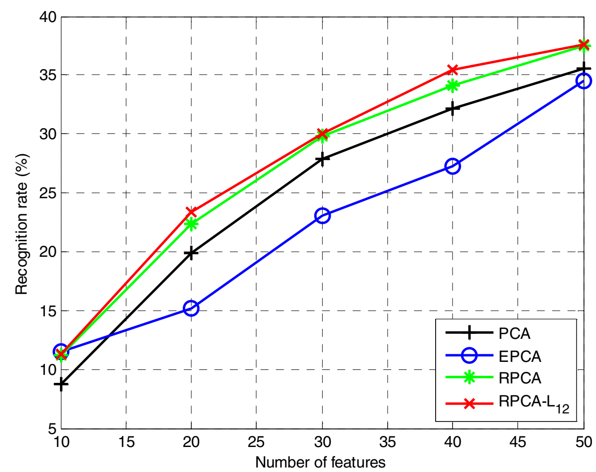
In the Extended Yale B database [14], there are 38 subjects. For each subject, there are 64 images under different lighting conditions. For each subject, 16 images are randomly selected for training, and the rest 48 images for testing. Each sample of the training set is resized to 24x21. Half of the training samples are corrupted by white block with 30% rate and 20% rate. Some samples are shown in Fig 1.

In the training phase, the PCA, EPCA, RPCA, and RPCA-$L_{1\text{-}2}$ are used for recovering the image corrupted by block. One can obtain a projection matrix $\mathbf{P}$ after PCA or EPCA. For RPCA and RPCA-$L_{1\text{-}2}$, one can obtain the recovered data $\mathbf{D}$. Then we pursue a projection $\mathbf{P}$ by PCA process for the recovered data $\mathbf{D}$.

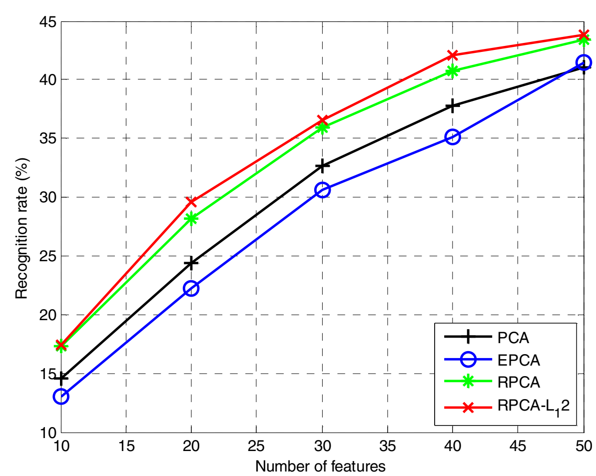
In the testing phase, the $\mathbf{P}$ can be projected on the testing data, and the NN classifier and SVM are employed for classification on projected testing data.

Figs. 2 and 3 shows recognition rates of all methods with varying features under the NN classifier and SVM. It can be seen that RPCA-L_{12} outperforms other methods in many cases.

Figure.1 Sample images in the training set

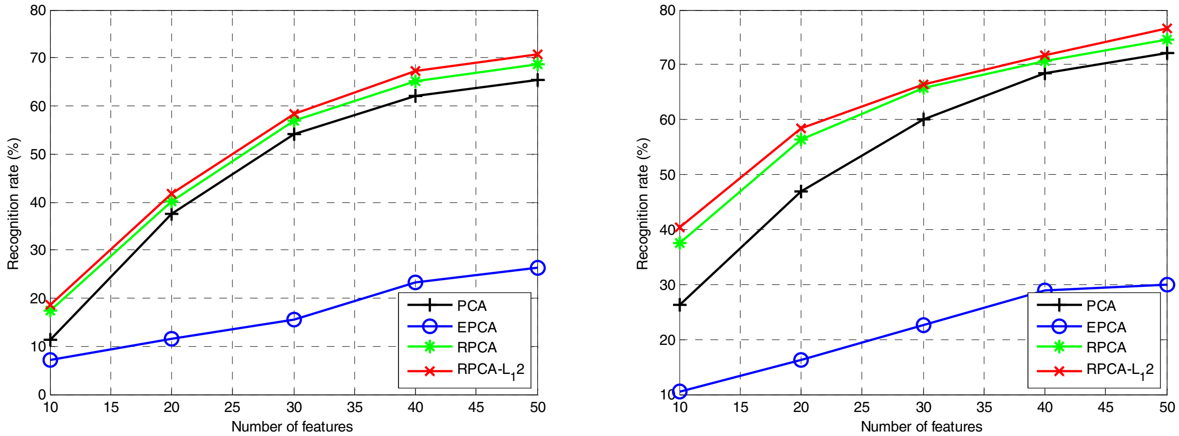


(a) 30% image range corrupted



(b) 20% image range corrupted

Figure.2 Recognition rates with varying feature number on the Extended Yale B database under the NN classifier



(a) 30% image range corrupted (b) 20% image range corrupted

Figure.3 Recognition rates with varying feature number on the Extended Yale B database under SVM.

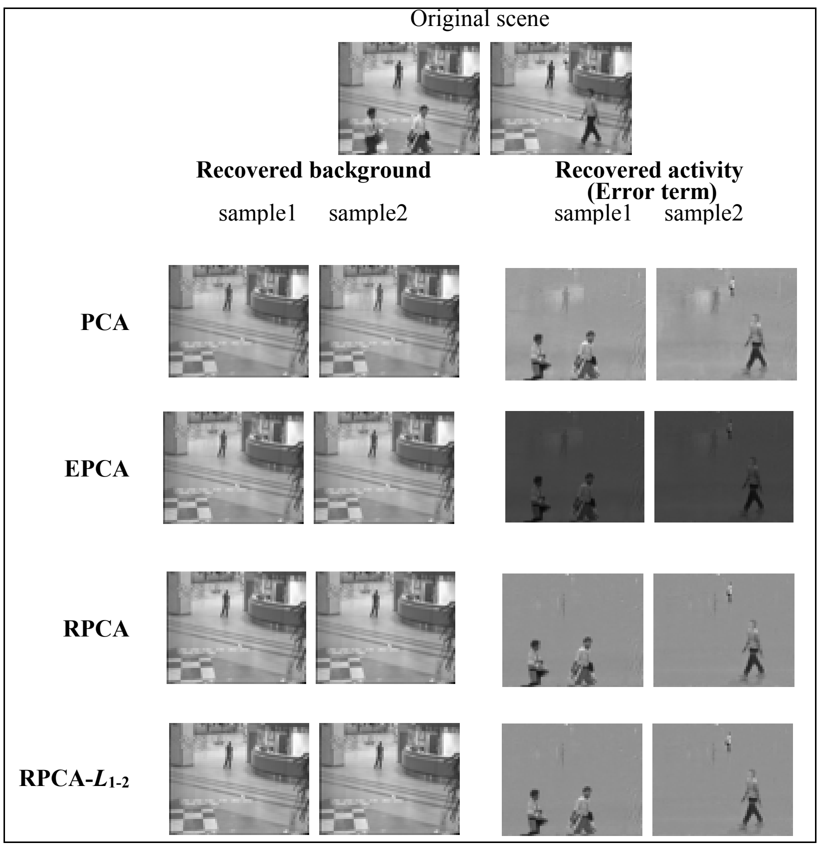


Figure 4 Background modeling from airport hall video

4.2. **Background modeling from surveillance video**

Background modeling from surveillance video plays a key role in event detection and human action recognition. In our experiments, an airport hall video is used [15]. Each frame is of size 85×106.

To obtain a quantitative comparison, we manually quote out the activities and use the F-score to measure the accuracy of segmentation, which is defined as:

$$F\text{-score} = 2\frac{\text{precision} \bullet \text{recall}}{\text{precision+recall}}, \tag{14}$$

$$\text{precision} = \frac{|G \cap T|}{|T|}, \text{recall} = \frac{|G \cap T|}{|\text{G}|}, \tag{15}$$

where G is the ground truth mask, and T is the mask output.

One hundred frames are selected randomly as training data and stack each frame into the matrix **X**. Then we recovery it by different methods. The parameters of these comparative methods are tuned to obtain the best F-scores.

Two original frames, the recovered frames and identified activities are shown in Fig.4, respectively. From these figures, we can see that all models can identify the activities from the frames in training set. Table 1 exhibits the F-scores. We can see that in many cases, RPCA-L_{12} outperforms others.

	sample 1	sample 2
PCA	0.8089	0.7578
EPCA	0.8551	0.8678
RPCA	0.8582	0.8562
RPCA-$L_{1\text{-}2}$	**0.8684**	**0.8612**

Table 1 F-scores of two samples in training set

5. Conclusions

This paper proposes a new data recovery model and corresponding optimization algorithm. In the new model, the $L_{1\text{-}2}$ metric is employed for characterizing the noise in the data. The experiments show that the proposed method is efficient in image recognition and background modelling. In further, we will consider how to extend the $L_{1\text{-}2}$ metric into the nuclear norm.

Acknowledgement

This work was supported by the National Science Fund of China under Grant Nos. 61603192.

References

[1] I.T. Jolliffe, Principal component analysis, second edition. New York: Springer-Verlag, 2002.

[2] E. Candès, X. Li, Y. Ma and J. Wright, "Robust principal component analysis?" Journal of the ACM, vol. 58, no. 3, pp.1–37, 2011.

[3] J. Wright, A. Ganesh, S. Rao and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in Proc. Adv. Neural Inform. Process. Syst., 2009.

[4] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," In Proc. Annu. Int. Conf. Mach. Learn., 2010, pp. 1–8.

[5] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp.171-184, Jan. 2013.

[6] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," Journal of Machine Learning Research, vol. 11, pp. 1081–1107,2010.

[7] R. Chartrand andW. Yin, "Iteratively reweighted algorithms for compressive sensing," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 3869–3872.

[8] E. Cand`es, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," Journal of Fourier Analysis and Applications, vol. 14, no. 5-6, pp. 877–905, 2008.

[9] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to nonnegative least squares problems with applications," SIAM Journal on Imaging Sciences, vol. 6, no. 4, pp. 2010–2046, 2013.

[10] P. Yin, Y. Lou, Q. He, and J. Xin, "Minimization of `1-`2 for compressed sensing," SIAM Journal on Scientific Computing, vol. 37, no. 1, pp. A536–A563, 2015

[11] P. Tao and L, An. Convex analysis approach to dc programming: Theory, algorithms and applications, Acta Mathematica Vietnamica, 22(1): 289-355, 1997.

[12] Z Li, Y Lou, T Zeng, Variational multiplicative noise removal by DC programming, Journal of Scientific Computing,68 (3):1200-1216,2016

[13] Liwicki, S., Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Euler principal component analysis. Int. J. Comput. Vis., 101(3), 498-518 (2013)

[14] Lee, K.C., Ho, J., Driegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell., 27(5), 684-698, (2005)

[15] L. Li, W. Huang, I. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground objects detection," IEEE Trans. Image Process, vol. 13, no. 11, pp. 1459–1472, 2004

[16] Lin, Z., Chen, M., Wu, L., .Ma, Y.: The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical report, UIUC Technical Report UILU-ENG-09-2215, (2009)