

Robust Non-Negative Matrix Factorization Procedures for Analyzing Tumor and Face Image Data

Jiayang Sun,* Yifan Xu,† Kenneth K. Lopiano,‡ S. Stanley Young§

Abstract

In this paper we introduce a set of robust non-negative factorization (rNMF) algorithms for analyzing large and corrupted data. The central idea is to introduce a penalized criterion that incorporates a trimming component, so that the rNMF procedure is flexible enough to handle different types of noise that may arise in a data source, and simultaneously control the sparsity of the decomposition. Multiple algorithms are developed by varying the way that outliers are determined and subsequently handled. The resulting algorithms work well when compared to existing NMF algorithms commonly used in practice and an existing robust NMF procedure. The rNMF algorithms developed here include fully automatic controls and semi-supervised controls to address more difficult areas that an unsupervised algorithm cannot properly handle. We illustrate the proposed methodology using simulated tumor image data and images of faces subject to different types of corruptions.

Key Words: Non-negative matrix factorization, robustness, corrupted data, facial image, image cleaning, data compression

1. Introduction

Non-negative matrix factorization (NMF) is a method for obtaining low dimensional structure from high dimensional data. NMF has been studied under the name “positive matrix factorization” by Paatero and Tapper (1994), and became widely known after Lee and Seung (1999) showed a dramatic example and proposed simple algorithms. Given a data matrix representing high dimensional data, traditional dimension reduction methods include principle component analysis (PCA), singular value decomposition (SVD), and vector quantization (Hastie et al. 2009). When a data matrix consists of non-negative entries, NMF enforces non-negative constraints and relaxes the orthogonality constraints enforced by PCA. It is more likely to capture parts-based underlying structures of the data by inducing an additive representation of the original data matrix (Fogel et al. 2013). Like many methods, NMF is sensitive to outliers. As illustrated in section 3.1, unwanted corrupted values in data sets are often captured by regular NMF as relevant features, and as a result compromises representations of real features.

Some robust NMF procedures have been proposed recently. See for example Zhang et al. (2011) and Rapin et al. (2012). The essence of the existing work involves adding a sparse matrix of the same dimension of the original matrix to NMF model, the purpose of which is to exactly capture a few sparse outliers, deterministically, leading to essentially the same approach in both papers.

We approach the problem differently by using statistical trimming procedures while incorporating sparsity constraints. Our robust NMF, called rNMF, offers seven variations in determining and handling outliers to treat different types of data sets and corruptions. The main idea uses an alternating non-negative least squares updating rule. Outliers in rows, columns and/or cells are detected in each iteration and trimmed or smoothed accordingly.

*Case Western Reserve University

†Case Western Reserve University

‡Statistical and Applied Mathematical Sciences Institute (SAMSI)

§National Institute of Statistical Sciences (NISS)

Euclidean distances are used to measure distance. In addition to a traditional NMF application, where low dimensional structure is learned from multiple data vectors, our robust NMF procedure can also be used to compress a single corrupted image and decompose it to reveal its part-based structure. Combining the ability of rNMF to obtain a clean compression with a supervised smoothing algorithm, we also develop a semi-automatic procedure that cleans different types of outliers in a heavily corrupted image. Adding to the seven variations of the rNMF approach, by this semi-supervised approach, we have a set of total eight algorithms. All algorithms are programmed in R (R Core Team 2013).

The structure of the paper is as follows. In Section 2 we introduce our main robust NMF model and algorithms (rNMF). In Section 3 we apply our rNMF to two types of image data, tumor and face images. Multiple image examples using the rNMF algorithms developed here are reserved for a future paper, where it is shown that our rNMF procedure does not suffer the averaging effect that some existing robust procedure exhibits. In Section 4 we develop a multi-level semi-supervised rNMF procedure (MrNMF) for compressing and cleaning images with complex corruptions. Some discussions and conclusions are given in Section 5.

2. Robust Non-negative Matrix Factorization (rNMF)

2.1 Regular NMF

Consider a $p \times n$ non-negative matrix \mathbf{X} , representing an image matrix. Let $x_{ij} \geq 0, i = 1, \dots, p, j = 1, \dots, n$ denote the i, j -th entry of \mathbf{X} . The goal of NMF is to factor \mathbf{X} into two low rank non-negative matrices whose product is as close to \mathbf{X} as possible with respect to a certain measure. \mathbf{X} is expressed as

$$\mathbf{X} = \mathbf{WH} + \mathbf{E} \quad (1)$$

where \mathbf{W} is $p \times k$ and \mathbf{H} is $k \times n$ matrices with $k \ll \min(p, n)$, and all entries of \mathbf{W} and \mathbf{H} are non-negative. In other words, \mathbf{W} and \mathbf{H} are chosen so that a distance measure between \mathbf{X} and \mathbf{WH} , namely $D(\mathbf{X}, \mathbf{WH})$ is minimized. Frequently used measures include the Frobenius distance defined by

$$D(\mathbf{X}, \mathbf{WH})_F := \sqrt{\sum_{i,j} e_{ij}^2}$$

where e_{ij} denotes the i, j -th entry of $\mathbf{X} - \mathbf{WH}$. Another measure is the generalized Kullback-Leibler divergence defined by

$$D(\mathbf{X}, \mathbf{WH})_{KL} := \sum_{i,j} \left(x_{ij} \log \frac{x_{ij}}{y_{ij}} - x_{ij} + y_{ij} \right)$$

where y_{ij} denotes the i, j -th entry of \mathbf{WH} . We use the Frobenius distance in this paper, but the algorithms can be easily modified to use the generalized Kullback-Leibler divergence. Frobenius distance is optimal if the elements in \mathbf{E} are normally distributed, while the generalized Kullback-Leibler divergence is optimal if \mathbf{E} has a Poisson distribution.

2.2 Proposed rNMF

Given a data matrix \mathbf{X} , the goal of our rNMF is to obtain

$$\mathbf{X} \approx \mathbf{WH}, \text{ where } \mathbf{X} \in \mathbb{R}_+^{p \times n}, \mathbf{W} \in \mathbb{R}_+^{p \times k}, \mathbf{H} \in \mathbb{R}_+^{k \times n}$$

such that \mathbf{W} and \mathbf{H} are resistant to outliers in \mathbf{X} , while controlling certain attributes of \mathbf{W} and \mathbf{H} , such as sparsity. To achieve the goal, we propose a trimmed, penalized criterion

$$(\mathbf{H}, \mathbf{W}) = \arg \min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \left\{ \|\mathbf{X} - \mathbf{WH}\|_{F,trim}^2 + \alpha \|\mathbf{W}\|_F^2 + \beta \sum_{j=1}^n \|h_{\cdot,j}\|_1^2 \right\} \quad (2)$$

where $\|\cdot\|_1$ stands for the L_1 norm and $h_{\cdot,j}$ denotes the j -th column of \mathbf{H} . α bounds $\|\mathbf{W}\|_F^2$, and β controls the sparsity of \mathbf{H} . $\|\cdot\|_{F,trim}$ will be defined below. We may further normalize \mathbf{W} by requiring $\|w_{\cdot,j}\|^2 = 1$ while holding α to be 0. Our criterion (2) can be viewed as a generalization of the criterion proposed by Kim and Park (2007). It is reduced to the latter by setting the trimming percentage in $\|\mathbf{X} - \mathbf{WH}\|_{F,trim}^2$ to zero. Furthermore, if α and β are also set to zero then (2) becomes the objective function of a regular NMF procedure.

How to trim the difference matrix $\mathbf{X} - \mathbf{WH}$ is actually challenging. One natural approach would be to adapt the idea of a trimmed regression: 1. Estimate \mathbf{W} and \mathbf{H} using a NMF algorithm. 2. Remove points with extreme residuals. 3. Refit NMF with the remaining points. However this approach does not yield good results, as in most cases NMF algorithms also identify outliers as real features (see Figure 1).

In our rNMF procedures, different actions are taken for different types of data: tumor images, face images, or other types of biological data. Our central idea in finding a solution to (2) is to combine and generalize Kim and Park (2007)'s alternating non-negative least square (ANLS) algorithm for sparse NMF, with a trimmed least squares procedures in a tactical way. Specifically, our rNMF back-bone algorithm is:

1. Randomly initialize \mathbf{H} .
2. With the fixed \mathbf{H} , update \mathbf{W} by solving

$$\mathbf{W} \leftarrow \arg \min_{\mathbf{W} \geq 0} \left\| \begin{pmatrix} \mathbf{H}^T \\ \sqrt{\alpha} I_k \end{pmatrix} \mathbf{W}^T - \begin{pmatrix} \mathbf{X}^T \\ 0_{k \times p} \end{pmatrix} \right\|_F^2. \quad (3)$$

- 2a. Declare outliers by evaluating residuals $\mathbf{R} = \mathbf{X} - \mathbf{WH}$ (see details below).
- 2b. Trim or smooth outliers (see details below) .
- 2c. Refit \mathbf{W} with trimmed/modified \mathbf{X} and \mathbf{H} by (3).
3. With the fixed \mathbf{W} , update \mathbf{H} by solving

$$\mathbf{H} \leftarrow \arg \min_{\mathbf{H} \geq 0} \left\| \begin{pmatrix} \mathbf{W} \\ \sqrt{\beta} e_{1 \times k} \end{pmatrix} \mathbf{H} - \begin{pmatrix} \mathbf{X} \\ 0_{1 \times n} \end{pmatrix} \right\|_F^2 \quad (4)$$

- 3a. Declare outliers by evaluating residuals $\mathbf{R} = \mathbf{X} - \mathbf{WH}$ (see details below).
- 3b. Trim or smooth outliers (see details below) .
- 3c. Refit \mathbf{H} with trimmed/modified \mathbf{X} and \mathbf{W} by (4).
4. Go to 2, unless convergence or the maximum number of iterations is reached.

Note that \mathbf{W} and \mathbf{H} are updated alternatively until convergence, and trimming occurs at either or both updating steps. Care is needed in deciding which data cells or vectors are outliers (step 2a and/or 3a), and how much related information needs to be trimmed (step 2b and/or 3b). Different criteria for declaring and trimming outliers are summarized in the following. The degree of the trimming percentage is a user option denoted by $0 < \gamma < 1$. Let $w_{\cdot,i}$ and $h_{\cdot,j}$ denote the i -th row vector of \mathbf{W} and j -th column vector of \mathbf{H} , respectively. All products are matrix multiplications.

1. The options for step 2a & 2b are the following:
 - (i) Calculate $r_{\cdot j}^2$, the column residual sum of squares, for $j = 1, \dots, n$. Declare columns that have γ largest $r_{\cdot j}^2$ to be the outlier columns; Remove outlier columns from \mathbf{X} and \mathbf{H} .
 - (ii) Calculate $r_{i \cdot}^2$, the cell residual sum of squares, for $i = 1, \dots, p, j = 1, \dots, n$. Declare columns that contain cells corresponding to γ largest $r_{i \cdot}^2$ to be the outlier columns; Remove outlier columns from \mathbf{X} and \mathbf{H} .
 - (iii) Calculate $r_{i j}^2$, as that in (iii), for $i = 1, \dots, p, j = 1, \dots, n$. Declare cells that contain γ largest $r_{i j}^2$ to be the outlier cells; Smooth outlier cells in \mathbf{X} with averages of nearest neighbors (See the Smoothing Subroutine in Section 4).
 - (iv) Do nothing.
2. The options for step 3a & 3b are similar to those of 2a & 2b, but for rows i , leading to four options, denoted by (I), (II), (III) and (IV).

In principle, each combination of the above criterion for Step 2a & 2b and Step 3a & 3b defines one unique trimming procedure, but not all choices are sensible. The actual choices of combinations should depend on the nature of the data set \mathbf{X} . For example, if \mathbf{X} consists of pixel intensities of a single facial image with sparse corruptions, approach (iii) and (III) are appropriate (See Section 3.2). If \mathbf{X} consists of column vectors each of which represents an observation with p features, approach (i) or (ii) and (IV) generally yield better results. In summary, the sensible combinations would lead to 7 variations in our rNMF:

1. ``colcol``: (i) + (IV).
2. ``cellcol``: (ii) + (IV).
3. ``rowrow``: (iv) + (I).
4. ``cellrow``: (iv) + (II).
5. ``vecall``: (i) + (I).
6. ``cellall``: (ii) + (II).
7. ``smooth``: (iii) + (III).

where the rationals for variations 1-6 are obvious, and the rational for variation 7 is given in Section 3.2.

Remark 1. *Regardless of the choices of row, column and/or cell trimming, trimming and modification of \mathbf{X} , \mathbf{W} and \mathbf{H} take place in each iteration, so that the descent of the objective function towards its local minimum is not affected by corrupted rows/columns/cells. As shown in Figure 2.*

We conclude this section with the following proposition showing the convergence of rNMF algorithm. We leave the proof and discussion of the proposition to the full journal paper.

Proposition 1. *Under the conditions defined in the beginning of this section, (2) converges to its local minima by the rNMF algorithm.*

3. Corrupted Image Compression and Cleaning by rNMF

In this section we illustrate rNMF with examples of different types of corrupted data sets.

3.1 rNMF with Tumor Data

First we apply rNMF to a single corrupted simulated tumor image. As shown in Figure 1, regular NMF compressions shown in the first five little images are compromised by outliers, while rNMF shown in the last little image effectively removes outliers. Figure 2 compares histograms of fitting errors of rNMF and other 5 popular NMF methods. The results for regular NMF are obtained from the “NMF” R package (Gaujoux and Seoighe 2010).

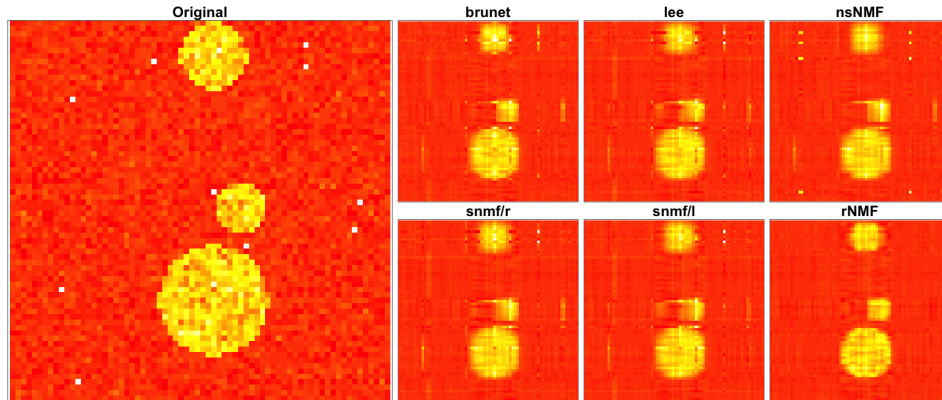


Figure 1: The dimension of the simulated tumor image is 70 by 70. Pixel intensities range from 0 to 1, and 10 outliers at intensity 2 are added at random locations. Results from five regular NMF procedures (Kim and Park 2007, Gaujoux and Seoighe 2010) are compared with the rNMF result. Outliers are effectively removed during the compression by rNMF.

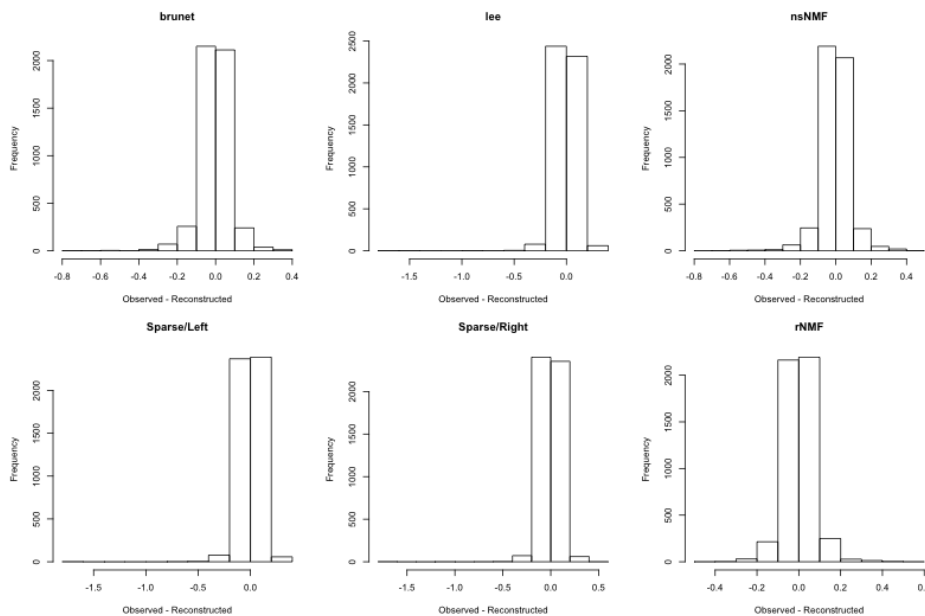


Figure 2: Comparisons of fitting residuals of rNMF and regular NMF procedures. Fitting residuals are the differences between the reconstructions and the original clean image.

One key advantage of NMF over other dimension reduction techniques is the ability to obtain a parts-based reconstruction. In Figure 3 we plot reconstructions of the tumor image by all possible subsets of the columns of the first matrix \mathbf{W} and corresponding rows of the second matrix \mathbf{H} . From the sequential plotting we can clearly see that column 1, 2 and 3 of the first matrix captured most of the information in the original tumor data. Furthermore, each column explains one tumor.

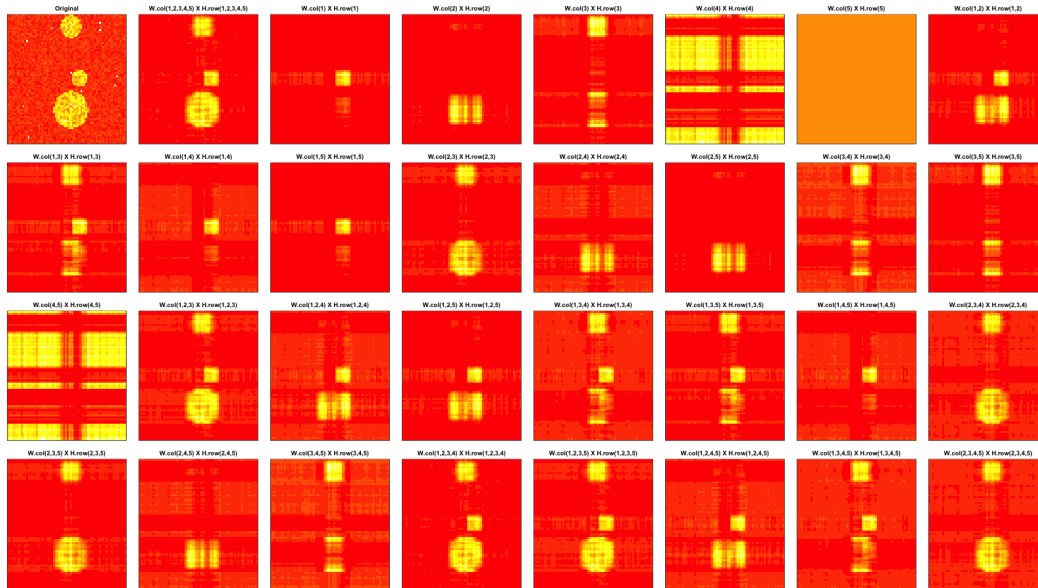


Figure 3: Sequential plotting of tumor image rNMF reconstruction.

3.2 rNMF with YaleB face image

In this section we apply rNMF on facial image with different types of corruptions. A gray scale image from the Yale Face Database B (Georghiades et al. 2001) is manually contaminated with three types of corruptions. Pixel intensities of the original image are normalized to range $[0, 1]$. Corruptions are added at intensity 0.7.

Face images have more complicated connected features than the simulated tumor image shown in the previous section. After outliers are determined in step 2a and/or 3a, trimming entire rows or columns may remove an entire portion of real facial feature. Therefore, these outliers are smoothed out by averaging values of neighboring pixels (variation 7 `'smooth'`). Various corruptions of the face were considered (see Figure 4): salt & pepper corruptions in a1, stream corruptions in b1, and block corruptions in c1. rNMF works very well with sparse corruptions (see a1, a2, b1 and b2 in Figure 4) but not clustered ones (see c1 and c2 in Figure 4). This is expected, as without supervision the third type of outliers in an image is recognized as a feature in much the way other clustered features are identified, e.g., eyes, mouth, etc. It generates small residuals in each fitting iteration, therefore it is not trimmed out by rNMF. Motivated by this limitation, we develop a semi-supervised rNMF procedure in the next section.

4. MrNMF: Semi-supervised Compression and Cleaning by rNMF, a Multi-level Algorithm

To distinguish the regular blockwise outliers with the true features (such as the whites in eyes as illustrated in Section 3.2), a fully automatic algorithm would have to be com-

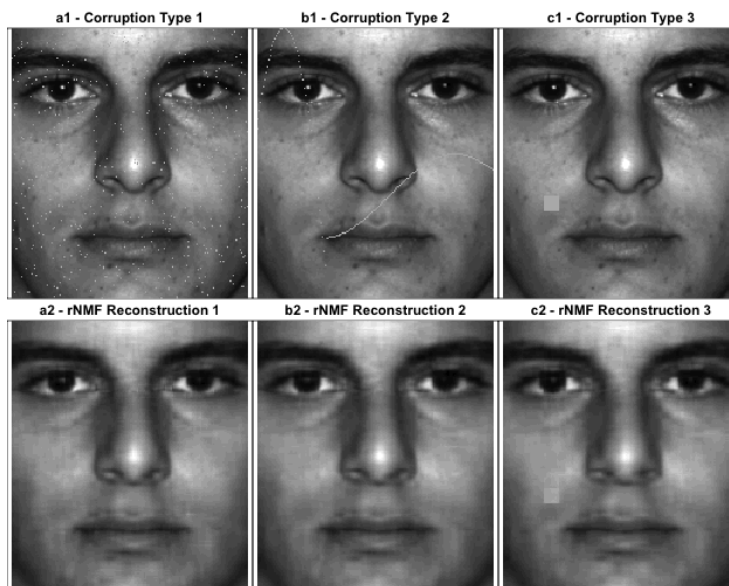


Figure 4: We apply rNMF to compress a single image contaminated by three types of outliers. The dimension of one image is 192 by 168. `variation = 'smooth'`, `k = 15`.

plex and costly to incorporate various types of blocks, specifying whether they represent true outliers or special features in an image, while a human can distinguish the difference of these two types of blocks easily. Hence, we introduce a semi-supervised procedure which let both computer and human do their best quickly and easily (a.k.a allowing semi-supervised cleaning by simple clicks inside an automation). Before introducing the multi-level semi-supervised rNMF (MrNMF), we need the following simple smoothing subroutine.

Procedure 1 (Smoothing Subroutine).

1. Assume a certain amount of points in a data matrix are labeled as outliers, and the rest are labeled as regular points.
2. For each outlier, search for regular points within a small radius of a predetermined length r . Increase r if there are no regular points in the neighborhood. Repeat until some regular points are found.
3. Replace the outlier with the average of regular points in the neighborhood.
4. Repeat 1,2 and 3 until all outliers are smoothed.

The smoothing subroutine is used in each iteration in the seventh variation of rNMF introduced in Section 2. This variation of our rNMF procedure successfully removed sparse outliers such as salt & pepper or stream noise. The smoothing subroutine can also be used independently to clean images, given that outliers are labeled in advance. Such situation includes images with missing values or outliers with extreme values (See Figure 5).

On the other hand, blocks of outliers in single images are much more difficult to detect automatically without involving complex algorithms that reveal data-specific low-dimensional structures. Motivated by this, we introduce the following fast semi-supervised detection procedure for block outliers.

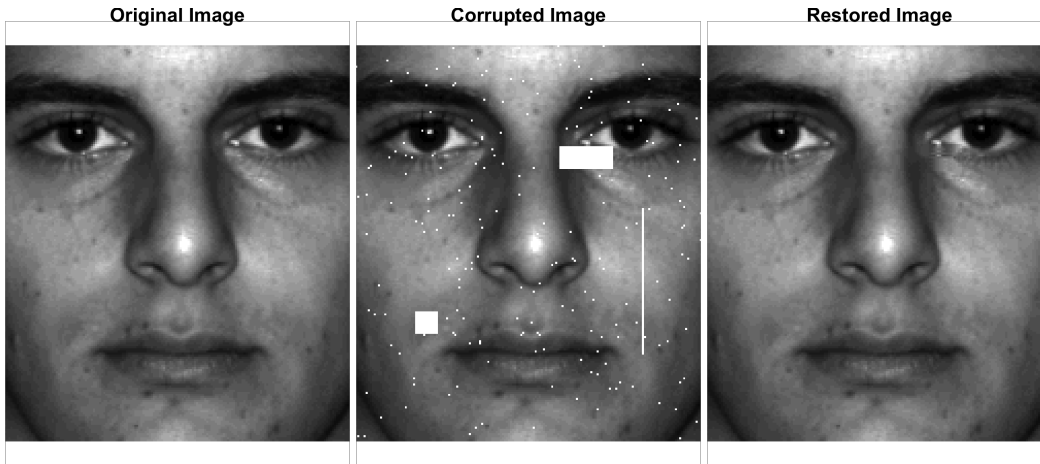


Figure 5: Apply thresholding and the smoothing subroutine with one corrupted Yale database B face image. Pixel intensities of the original image is normalized to $[0, 1]$. Outliers are added at intensity 1.001.

Procedure 2 (Semi-supervised Detection of Block Outliers).

1. *Select rectangular region that covers an outlier cluster.*
2. *Perform k -means clustering to separate the region into two sets, label the set that has more distinctive pixel values than surrounding areas as “outliers”, and label other points “regular points”.*

Combining the above two procedures with our rNMF, we develop a comprehensive multi-level rNMF (MrNMF) which has two important features. First it is capable of extracting clean low dimensional structure from a heavily corrupted image. Second it is a powerful image cleaning algorithm that effectively removes both clustered and sparse outliers at various pixel intensities, while protecting various sensible areas with minimal human interaction. The outline of MrNMF is summarized as follows, and illustrated in Figure 6.

Procedure 3 (Multi-level rNMF, MrNMF).

1. *Denote the original corrupted image by M_1 .*
2. *Threshold and smooth outliers accordingly in M_1 by Procedure 1. Denote the output by M_2 . This step is automatic given threshold parameters.*
3. *(A quick human controlled step.) Apply Procedure 1 and 2 on M_2 . Denote the resulting image by M_3 .*
4. *Apply rNMF to M_3 . Denote the low dimension reconstruction by M_4 .*
5. *Subtract M_4 from M_3 . Denote the difference matrix by M_D .*
6. *Label $\eta\%$ pixels of M_D that have the most extreme values as “outliers”, where $\eta\%$ is a predetermined percentage.*
7. *Smooth “outliers” labeled in the previous step with Procedure 1.*

Remark 2. *A clean low dimension reconstruction is obtained after Step 4 in MrNMF. Step 5 - 7 further generate an uncompressed clean image. If necessary, Step 3 can be reapplied after Step 7.*

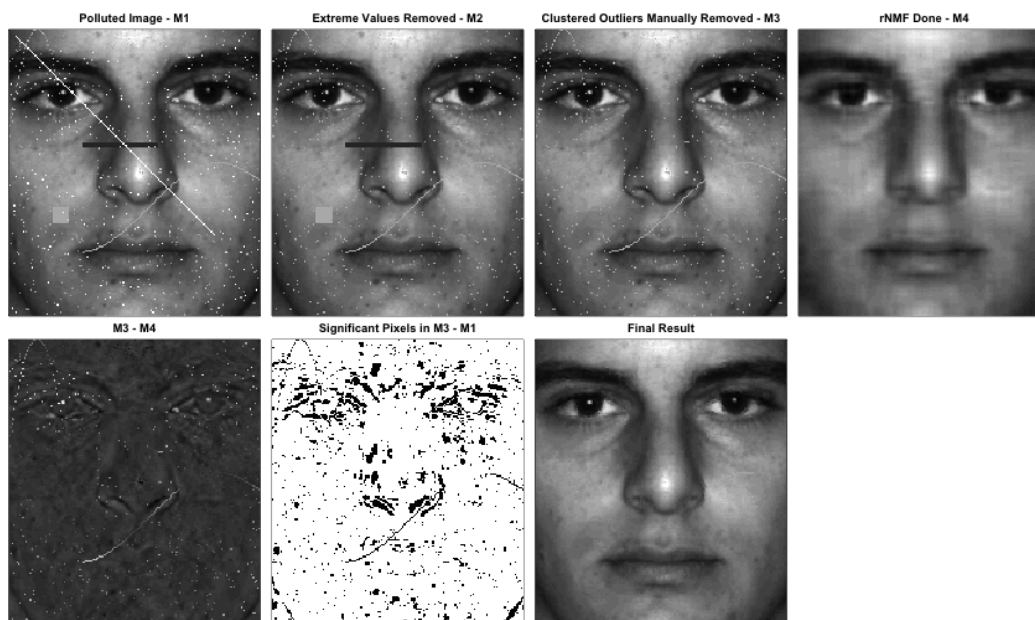


Figure 6: Multi-level rNMF (MrNMF).

5. Discussions and Conclusions

We have developed a set of effective rNMF algorithms and the MrNMF procedure, which achieve data compression in all tested cases while removing outliers and keeping the most important features, such as the whites in eyes, and eye pupils, as shown in Figure 6. We have also applied and extended our procedures to multiple images and compared with an existing robust procedure. We shall present the multiple image result in a full journal paper. The purpose of this proceedings paper is to provide some wonderful ideas, and show what our algorithms can do, signaling a good beginning of our new set of rNMF algorithms, quickly.

REFERENCES

- Fogel, P., Hawkins, D. M., Beecher, C., Luta, G. and Young, S. S. (in press) "A tale of two matrix factorizations" *The American Statistician*
- Gaujoux, R. and Seoighe, C. (2010), "A flexible R package for nonnegative matrix factorization" *BMC Bioinformatics*, 11.1: 367.
- Georghiadis, A. S., Belhumeur, P. N. and Kriegman, D. J. (2001), "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose" *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23.6:643-660
- Hastie, T., Tibshirani, R., and Friedman, J. (2009), "The Elements of Statistical Learning" *Springer-Verlag*
- Kim, H. and Park, H. (2007), "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis" *Bioinformatics*, 23:1495-1502.
- Lawson, C.L., and Hanson, R.J. (1974), "Solving Least Squares Problems" *Prentice-Hall, Englewood Cliffs, NJ*
- Lee, D. D. and Seung, H. S. (1999), "Learning the parts of objects by non-negative matrix factorization" *Nature*, 401:788-791.
- Paatero, P. and Tapper, U. (1994), "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values" *Environmetrics*, 5: 111126.
- R Core Team (2013), "R: A language and environment for statistical computing" *R Foundation for Statistical Computing, Vienna, Austria*
- Rapin, J., Bobin, J., Larue, A. and Starck, J. (2012), "Robust Non-Negative Matrix Factorization for Multispectral Data with Sparse Prior" *Proceedings, Seventh Conference on Astronomical Data Analysis, Cargese, Corsica*

- Shen, B., Si, L., Ji, R. and Liu, B. (2012), "Robust Nonnegative Matrix Factorization via L_1 Norm Regularization" *ArXiv e-prints*, 1204.2311
- Zhang, L., Chen, Z., Zheng, M. and He, X. (2011), "Robust Non-negative Matrix Factorization" *Frontiers of Electrical and Electronic Engineering*, 6.2:192-200