

ROBUST FACE RECOGNITION USING TRIMMED LINEAR REGRESSION

Jian Lai and Xudong Jiang

School of Electrical and Electronic Engineering,
Nanyang Technological University, Singapore 639798

{jlai1, exdjiang}@ntu.edu.sg

ABSTRACT

In this work, we focus on the problem of partially occluded face recognition. Using a robust estimator, we detect and trim the contaminated pixels from query sample. The corresponding pixels in the training samples are trimmed as well. The linear regression is applied to the trimmed images. Finally, the query image is labeled to the class with minimum normalized reconstruction error. Extensive experiments on benchmark face datasets demonstrate that the proposed approach is much more robust than state-of-the-art methods in dealing with occluded faces.

Index Terms— Biometrics, disguise, face recognition, partial occlusion, robust linear regression.

1. INTRODUCTION

Face recognition, as a very widely used application, has been received attention from numerous researchers around the world. In the last few decades, many methods [1, 2, 3, 4, 5] are proposed for building a reliable face recognition system. Current frameworks can well handle the face captured under-controlled environment. However, lots of work still need to be done before it reaches a certain level of maturity in the uncontrolled condition. Occlusion, generated by disguise (e.g. scarf, sunglasses, and hair) or blocked by other objects, is a very common and difficult problem in the real world for face recognition.

As the face is occluded, popular holistic algorithms, such as Eigenfaces [6] and Fisherfaces [7], can not be applied as the extracted features are contaminated. Wright *et al.* [8] try to model the occlusion by the sparse combination of columns of the identity matrix. However, the approach in [8] breaks down for large occlusion. Alternative approaches are based on the local feature. Face image is divided into several local patches and each patch is processed independently [9, 10]. The final decision is based on fusing the classification results of all patches. In [8], the sparse representation coding (SRC) is applied to each patch and majority voting is employed for making decision. The work in [11] handles each part by linear regression classification (LRC) and classifies query face

by labeling it to the subject with the minimum representation error in all patches. However, single patch has limited discriminative power. Recently, to tackle the occlusion problem of holistic approach and insufficient information of local approach, Lai and Jiang [12] advance both the global and modular algorithms by a combination of the both. As the occluded part may partially covers many patches for a fixed pre-partition of the patches, its performance crucially depends on how to partition the images.

In LRC [11], authors cast face recognition as a class specific linear regression problem. The least square (LS) is used to estimate the regression coefficients. LRC works well for clean images. However, as the imposing of outliers, the within class variation is enlarged, while the between class variation is reduced. Moreover, the LS is very sensitive to outliers [13]. Therefore, the performance of LRC decreases sharply as the face contaminated by outliers. To address the instability of LRC with outliers, the robust linear regression classification (RLRC) [14] is proposed to replace the LS with Huber's M-estimator [13]. Compared with the LS, the Huber's M-estimator weighs the large residuals more lightly. As a result, the outliers have less significant affection on the estimated coefficients. Although RLRC attenuates the problem of outliers in some degree, it is still unreliable under the large occlusion which is a common scenario (e.g. about 40% of face is occluded for a man wearing scarf).

As the test sample is occluded, the assumption of faces from the same class lying on a linear subspace is violated. However, if we trim the outliers from the test and the corresponding training samples, the mechanism still can work. In this paper, we propose a method, named trimmed linear regression classification (TLRC), for face recognition with partial occlusion. Outliers are detected and trimmed based on the residuals of least trimmed square (LTS) [15]. Then, the LS is applied to the trimmed samples. The query image is labeled to the class with minimum normalized reconstruction error.

2. TRIMMED LINEAR REGRESSION CLASSIFICATION

Given a set of training images from c different subjects. Each subject contains n_i training samples, $i = 1, 2, \dots, c$. Let

$\mathbf{A}^i = [\mathbf{a}_1^i, \mathbf{a}_2^i, \dots, \mathbf{a}_{n_i}^i] \in \mathbb{R}^{m \times n_i}$, where the column vector $\mathbf{a}_j^i$ represents the j th training image of the i th subject.

If a query image $\mathbf{y} \in \mathbb{R}^m$ belongs to the i th class, it can be predicted as a linear combination of the training images from the same class, i.e.

$$\hat{\mathbf{y}} = \mathbf{A}^i \boldsymbol{\beta}^i. \quad (1)$$

where $\boldsymbol{\beta}^i$ is the regression coefficients. Since real data are noisy, it may not be possible to reconstruct the test image exactly. The query image can be explicitly modeled as the sum of $\hat{\mathbf{y}}$ and noise by writing

$$\begin{aligned} \mathbf{y} &= \hat{\mathbf{y}} + \mathbf{e} \\ &= \mathbf{A}^i \boldsymbol{\beta}^i + \mathbf{e} \end{aligned} \quad (2)$$

where $\mathbf{e} \in \mathbb{R}^m$ is an error term following the Gaussian distribution with zero mean and σ standard deviation. LRC performs well under this ideal case.

2.1. Detection of Outliers

Unfortunately, when query image is with occlusion, the above model is violated. It is impossible to fit all pixels well. What could be expected is that the majority of image can be represented accurately. For this purpose, since the location of the outliers is unknown, LS is not applicable. In this work, we choose the least trimmed squares (LTS) [15] as the robust estimator. As the proportion of occlusion is uncertain, we assume that at most half of the image is occluded, which is also the highest breakdown point that can be achieved by all robust estimators [13]. Therefore, the robust regression coefficients is estimated as

$$\boldsymbol{\beta}_{lts}^i = \arg \min_{\boldsymbol{\beta}} \sum_{j=1}^{m/2} (e^2)_{j:m}, \text{ s.t. } \mathbf{e} = \mathbf{y} - \mathbf{A}^i \boldsymbol{\beta} \quad (3)$$

where $(e^2)_{1:m} \leq (e^2)_{2:m} \leq \dots \leq (e^2)_{m:m}$ are the squared residuals in ascending order. (3) minimizes sum of the squares of the smallest half residuals instead of all. If the outliers exist, LTS will achieve a much more robust result than LS. Meanwhile, LTS will generate a close estimate as LS when the data is clean.

As the faces of some subjects are similar, there could be more than 50% pixels of images from different subjects matching well. In this case, the result of trimming half pixels is unreliable. Therefore, contrary to directly using the result of (3), we trim pixels as few as possible by the following procedures.

After generating $\boldsymbol{\beta}_{lts}^i$, we reconstruct the test sample as

$$\mathbf{y}_{lts}^i = \mathbf{A}^i \boldsymbol{\beta}_{lts}^i. \quad (4)$$

The error term between $\mathbf{y}$ and $\mathbf{y}_{lts}^i$ is

$$\mathbf{e}_{lts}^i = \mathbf{y} - \mathbf{y}_{lts}^i. \quad (5)$$

$\mathbf{e}_{lts}^i$ is an estimate of $\mathbf{e}$ in (2), whose real median is zero. Therefore, the median absolute deviation (MAD) from the real median is derived as

$$\begin{aligned} mad(\mathbf{e}_{lts}^i[s_0^i]) &= median(|\mathbf{e}_{lts}^i[s_0^i] - 0|) \\ &= median(|\mathbf{e}_{lts}^i[s_0^i]|) \end{aligned} \quad (6)$$

here $median(\cdot)$ denotes the median function and s_0^i is the set including indices of all pixels, that is $s_0^i = \{1, 2, \dots, m\}$.

The robust estimate of standard deviation is derived by [16]

$$\hat{\sigma}_0 = mad(\mathbf{e}_{lts}^i[s_0^i])/0.6745. \quad (7)$$

For the Gaussian distribution, the probability of a sample beyond the range from -2σ to 2σ is only 2.28%. As the estimated $\hat{\sigma}_0$ is larger than the real σ , there are more than 97.72% clean pixels with residuals residing in $-2\hat{\sigma}_0$ to $2\hat{\sigma}_0$. Any sample with residual lying outside this range is very likely contaminated. Therefore, we trim these pixels from the test image and update the indices set $\mathbf{s}^i$ as

$$\mathbf{s}_1^i = \{k | -2\hat{\sigma}_0 \leq e_{lts}^i[k] \leq 2\hat{\sigma}_0\}. \quad (8)$$

However, as the outliers exist, the estimated $\hat{\sigma}_0$ tends to be larger than the real σ . Therefore, there are still some indices of outliers in $\mathbf{s}_1^i$ set. To address this problem, we recalculate $\hat{\sigma}_1$ by replacing $\mathbf{s}_0^i$ with $\mathbf{s}_1^i$ in (6) and (7). The new $\hat{\sigma}_1$ is employed to trim the pixels in (8). This procedure (6)-(8) is repeated until $\mathbf{s}^i$ is convergent, that is $\mathbf{s}_{t+1}^i = \mathbf{s}_t^i$.

2.2. Trimmed Least Squares

The convergent result of (8) is denoted as $\mathbf{s}^i$ and the $\bar{\mathbf{s}}^i = \{1, \dots, m\} \setminus \mathbf{s}^i$. The number of elements in $\mathbf{s}^i$ is indicated by l^i . $\tilde{\mathbf{y}}^i$ and $\tilde{\mathbf{a}}_j^i$ are obtained by trimming the pixels corresponding to $\bar{\mathbf{s}}^i$ from $\mathbf{y}$ and $\mathbf{a}_j^i$, respectively. The trimmed training matrix of the i th class is generated as $\tilde{\mathbf{A}}^i = [\tilde{\mathbf{a}}_1^i, \tilde{\mathbf{a}}_2^i, \dots, \tilde{\mathbf{a}}_{n_i}^i]$. We consider to represent the $\tilde{\mathbf{y}}^i$ as a linear combination of the columns of $\tilde{\mathbf{A}}^i$, i.e.

$$\tilde{\mathbf{y}}^i = \tilde{\mathbf{A}}^i \boldsymbol{\beta}^i. \quad (9)$$

As $\tilde{\mathbf{y}}^i$ is produced by trimming outliers, $\boldsymbol{\beta}^i$ can be estimated by the LS estimator:

$$\hat{\boldsymbol{\beta}}^i = (\tilde{\mathbf{A}}^{iT} \tilde{\mathbf{A}}^i)^{-1} \tilde{\mathbf{A}}^{iT} \tilde{\mathbf{y}}^i. \quad (10)$$

The predicted response vector for the i th class is represented as

$$\hat{\mathbf{y}}^i = \tilde{\mathbf{A}}^i \hat{\boldsymbol{\beta}}^i. \quad (11)$$

2.3. Classification Procedure

The distance between reconstructed image $\hat{\mathbf{y}}^i$ and $\tilde{\mathbf{y}}^i$ is calculated for each class. The most common metric to measure the distance between two vectors is l_2 -norm (i.e. Euclidean

distance). However, the l_2 -norm is not robust to measure the similarity. It amplifies the large components and reduces the small ones. Even for two very similar vectors except a few elements with large differences, they will be far away by measuring in l_2 -norm metric. As discussed in [17, 18], the l_1 -norm metric (i.e. Manhattan distance) is more preferable than the l_2 -norm metric in high dimension data application. Therefore, in this work, l_1 -norm is used to measure the distance between $\tilde{\mathbf{y}}^i$ and $\hat{\mathbf{y}}^i$. Since the number of trimmed pixels is different for various classes, a normalizer is needed. As a natural choice is l^i , the average pixel distance is obtained by

$$\hat{d}(\mathbf{y}, i) = \|\tilde{\mathbf{y}}^i - \hat{\mathbf{y}}^i\|_1 / l^i. \quad (12)$$

In case that two classes are with the same $\hat{d}(\mathbf{y}, i)$, the one with more pixels is more likely to be correct. Therefore, to encourage the subject with larger l^i , the normalized distance between $\tilde{\mathbf{y}}^i$ and $\hat{\mathbf{y}}^i$ is calculated as

$$d(\mathbf{y}, i) = \hat{d}(\mathbf{y}, i) / \log l^i. \quad (13)$$

The class label of test sample $\mathbf{y}$ is given by

$$i^* = \arg \min_i d(\mathbf{y}, i) \text{ for } 1 \leq i \leq c. \quad (14)$$

Algorithm 1: Trimmed Linear Regression Classification (TLRC)

Input: Matrixes of training sample set $\mathbf{A}^1, \mathbf{A}^2, \dots, \mathbf{A}^c$ for c classes and $\mathbf{A}^i \in \mathbb{R}^{m \times n_i}$. A query image $\mathbf{y} \in \mathbb{R}^m$.

- 1 **for** each subject i **do**
- 2 Solve the LTS problem as (3).
- 3 Compute the difference between the query image and the reconstructed image as (4), (5) and initialize $\mathbf{s}_0^i = \{1, 2, \dots, m\}$.
- 4 **repeat**
- 5 Estimate the robust standard deviation as (6) and (7).
- 6 Update the $\mathbf{s}_{t+1}^i$ as (8).
- 7 **until** $\mathbf{s}_{t+1}^i = \mathbf{s}_t^i$;
- 8 Trim the occluded pixels from both test and training images as $\tilde{\mathbf{y}}^i = \mathbf{y}[\mathbf{s}^i]$ and $\tilde{\mathbf{A}}^i = \mathbf{A}^i[\mathbf{s}^i, :]$
- 9 Compute the regression coefficients based on the trimmed samples as (10).
- 10 Compute the normalized reconstruction error $d(\mathbf{y}, i)$ as (11), (12) and (13).
- 11 **end**

Output: $\text{identify}(\mathbf{y}) = \arg \min_i d(\mathbf{y}, i)$.

3. EXPERIMENTS

In this section, the proposed method is evaluated on the same databases as in [8]: Extended Yale B [19] and AR

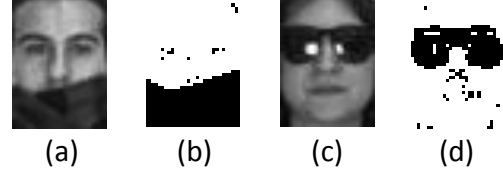


Fig. 1. Estimate of Outliers for faces with disguise. (a) A test face with scarf. (b) The estimated outliers (denoted by black pixels) by the correct subject. (c) A test face with sunglasses. (d) The estimated outliers by the correct subject.

database [20]. In the experiments, we compare the proposed TLRC with the related methods: nearest neighborhood (NN), SRC appended with an identity matrix [8], modular SRC (MSRC) [8], LRC [11], modular LRC (MLRC) [11], and RLRC [14]. Here, we employ the NN to provide a standard baseline for comparison.

3.1. AR Database

The AR database consists of over 4,000 frontal-face images from 126 subjects. For each individual, 26 pictures were taken in two separate sessions. The same as in [8], 50 male subjects and 50 female subjects are selected.

In this experiment, we use 799 images (about 8 per subject) of unoccluded frontal views under varying facial expression as training samples with image size of 42×30 . Two disguised test sets with 200 images for each are evaluated. In the first test set, images of the subjects wearing sunglasses are selected. The second disguise set contains faces with scarves. Fig. 1(a) and (c) show the examples of query images.

Table 1. Face Recognition Rate on the AR Database.

Scenarios	Sunglass	Scarf
NN	42.5%	7.0%
SRC[8]	87.0%	59.5%
MSRC[8]	97.5%	93.5%
LRC[11]	65.5%	12.5%
MLRC[11]	93.5%	73.5%
RLRC[14]	47.5%	9.5%
TLRC	100%	94.5%

Fig. 1(b) and (d) give the estimates of outliers by the correct subject for faces with scarf and sunglasses, respectively. They show that the estimated outliers are very accurate. Although RLRC employs a robust estimator (Huber's M-estimator) to obtain the regression coefficients, it performs even worse than LRC because of using a non-robust measurement (l_2 -norm) for classification for this dataset. Table 1 lists

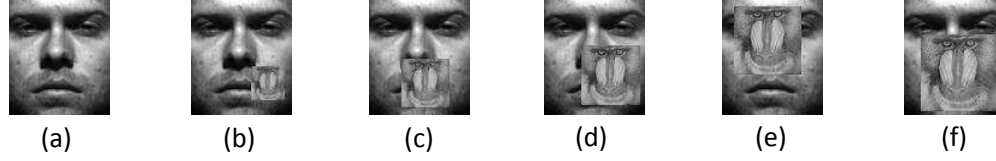


Fig. 2. Query images under varying level of contiguous occlusion from 0% to 50%.

Table 2. Face Recognition Rate on the Random Block Occluded Extended Yale B Database.

Occluded	0%	10%	20%	30%	40%	50%
NN	74.7%	78.9%	74.5%	67.3%	56.5%	39.3%
SRC	100%	100%	99.8%	98.5%	90.03%	65.3%
MSRC	98.5%	96.9%	93.6%	86.8%	76.5%	51.4%
LRC	100%	99.6%	93.0%	88.2%	60.4%	43.5%
MLRC	99.6%	99.3%	99.1%	98.0%	95.0%	80.7%
RLRC	100%	98.2%	79.8%	63.7%	49.7%	40.0%
TLRC	100%	100%	100%	100%	99.3%	98.2%

the recognition performances in conjunction with the two disguised variations. For face occluded by sunglasses, our method correctly classifies all subjects. It is 2.5% better than the second best result. On faces covered by scarves, none of the other global algorithms achieves higher than 60%. Although MSRC obtains a competitive result with the proposed method, it is much more time-consuming.

3.2. Extended Yale B Database

The cropped Extended Yale B database consists of 2,414 frontal-face images of 38 subjects with size 192×168 , captured under 64 different lighting conditions. For computational convenience, the face images are resized to 42×30 . As [8], Subsets 1 and 2 (719 images, normal-to-moderate lighting conditions) are selected for training and Subset 3 (455 images, more extreme lighting conditions) is chosen for testing. Different levels of contiguous occlusion, from 0% to 50%, are added to the test images by replacing a randomly located square block of each test image with an unrelated image, as shown in Fig. 2. The location of occlusion is randomly chosen for each image and is unknown to the classifiers.

Table 2 shows the face recognition rates for different levels of occlusion cross various methods. When no or only 10% occlusion is added, all methods but NN achieve almost perfect results. As the occlusion increases to 20% and 30%, the recognition accuracies of LRC, MSRC and RLRC are slightly affected because of their sensitiveness to outliers. MLRC, SRC and TLRC perform much better than others. For the rate of occlusion increases to 40%, the performances of MLRC, SRC, MSRC, LRC and RLRC drop to around 95%, 90%, 75%, 60% and 50%, respectively. The proposed method is hardly interfered by the occlusion. When half of the query image

is occluded, the recognition rate of other methods deteriorate drastically. The performance of TLRC is still around 98%. It confirms that the TLRC is much more robust to handle face with large occlusion.

4. CONCLUSION

In this paper, we propose a novel robust face recognition algorithm based on the trimmed least square. The contaminated pixels in query sample are detected and trimmed. Correspondingly, the pixels in training samples are trimmed as well. The classification is performed on the linear regression of trimmed samples. Different from the LRC and RLRC, we eliminate the affection of contaminated pixels from the final classification. Our method is evaluated on disguised and blocked faces. The experimental results clearly and consistently show that the proposed framework is much more robust than state-of-the-art methods in dealing with severely occluded faces.

5. REFERENCES

- [1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, Apr. 1994.
- [2] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: global versus component-based approach," in *Int. Conf. Computer Vision*, 2001, vol. 2, pp. 688–694.
- [3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.
- [4] X. D. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 5, pp. 931–937, May 2009.

- [5] X. D. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 16–26, Mar. 2011.
- [6] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 1991, pp. 586–591.
- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, July 1997.
- [8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [9] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 1994, pp. 84–91.
- [10] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, June 2002.
- [11] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.
- [12] J. Lai and X. D. Jiang, "Modular weighted global sparse representation for robust face recognition," *IEEE Signal Processing Letters*, vol. 19, no. 9, pp. 571–574, Sep. 2012.
- [13] P. J. Huber, *Robust Statistics*, John Wiley and Sons, Inc., 2005.
- [14] I. Naseem, R. Togneri, and M. Bennamoun, "Robust regression for face recognition," *Pattern Recognition*, vol. 45, no. 1, pp. 104–118, 2012.
- [15] J. R. Peter and M. L. Annick, *Robust Statistics*, Wiley-Interscience, 2003.
- [16] R. Maronna, R. Martin, and V. Yohai, *Robust Statistics: Theory and Methods*, John Wiley, 2006.
- [17] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional spaces," in *Proceedings of the 8th International Conference on Database Theory*, London, UK, 2001, pp. 420–434, Springer-Verlag.
- [18] A. Hinneburg, C. C. Aggarwal, and D. A. Keim, "What is the nearest neighbor in high dimensional spaces?," in *Proceedings of the 26th International Conference on Very Large Data Bases*, San Francisco, CA, USA, 2000, pp. 506–515, Morgan Kaufmann Publishers Inc.
- [19] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643–660, June 2001.
- [20] A. M. Martinez, "The ar face database," CVC Technical Report 24, June 1998.