The Performance of LTS-based Regression Methods in Face Recognition with Occluded Images

Nur Azimah Abdul Rahim*, Norazan Mohamed Ramli and Nor Azura Md Ghani
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia; nurazimah1809@gmail.com, norazan@tmsk.uitm.edu.my, azura@tmsk.uitm.edu.my

Abstract
In this paper, we compared the performance of several Least Trimmed Squares (LTS) based methods for face image recognition. The focus was on the problem of severely occluded face images. The performance of random LTS, Fast LTS and GA-LTS (which is based on genetic algorithm) to a benchmark dataset with occluded query images was examined. The best method was the one being least affected by the occluded images and produces highest recognition rates. The AT&T and Yale Data were used to assess the methods in performing face recognition. The query images were contaminated with salt and pepper noise and the recognition rates were measured when the contaminated images were used as query image in the context of linear regression. Results show that the random LTS outperforms the rest in dealing with occluded images with highest recognition rates.

Keywords: Face Recognition, Least Trimmed Squares with Genetic Algorithm, Occluded Images

1. Introduction
Several studies have been conducted on face recognition methods [1-2]. The face recognition methods were applied in the real-world settings due to its vast applications in our life [3]. There exist various frameworks and algorithms for a face recognition system. However, most of these frameworks are only reliable when the face captured under controlled environment. Even though the frameworks are reliable enough to be used, there are still lots of improvements can be done to these frameworks [4]. Face recognition method is very much affected by noise or occlusion which can be seen as grain in film and pixel variations if in digital images [5-6] and their presence caused variations intensity in the image pixels instead of true pixel values [7]. For most face recognition algorithms, partial occlusions affect the performance of the algorithm [8]. This study aims to investigate the performance of Least Trimmed Squares (LTS) based regression methods in face recognition. This paper is organized as follows. The LTS based methods for face recognition is presented in Section 2. This is followed by results and analysis in Section 3. The paper finally concludes the findings in Section 4.

2. LTS-based Regression Methods for Face Recognition
Face recognition can be represented as a class specific linear regression problem [9-11]. In modeling the face recognition model structure, \( N \) is assigned for distinguished classes, while \( p_i \) are the prototype images for the \( i \)-th class, \( i = 1, 2, ..., N \). Every single prototype (or training image) is denoted as a column vector, \( x_i(p) \in \mathbb{R}^{q \times 1} \) (\( p = 1, 2, ..., p_i \)) and \( q \) is the dimension of feature space for all images. The developed class-specific model of the \( i \)-th class is denoted as

\[
A_i = \begin{bmatrix}
x_i^{(1)} & x_i^{(2)} & \cdots & x_i^{(p)}
\end{bmatrix}
\]

where \( A_i \in \mathbb{R}^{q \times p} \). A query image, \( y \in \mathbb{R}^{q \times 1} \), can be predicted as a linear combination of training images from the same class

\[
\hat{y} = A_i \beta^p
\]

here \( \beta^{p \times 1} \) is the regression coefficients. The query image can be explicitly modeled as the sum of \( \hat{y} \) and noise by
where $e \in \mathbb{R}^q$ is the error term which follows the Gaussian distribution with zero mean and $\sigma$ standard deviation. Hence, if a test image $y \in \mathbb{R}^{q \times 1}$ is belong to the $i$th class, the following equation should be satisfied as $y$ lies either on or is closed to the subspace of its own class

$$d_i = \|y - A_i\beta^p\|_2 = 0$$

In the nearest subspace classification introduced by\(^{11}\), a class-specific model that can give a test image the best prediction as a Least Squares (LS) regression problem can be formulated as:

$$\beta^p = (A_i^TA_i)^{-1}A_i^Ty$$

where $A_i^TA_i \in \mathbb{R}^{p \times p}$. By comparing the predicted vector and the original test image in the Euclidean distance as shown in Equation (6), we can rule in favor of the class which has a minimum $d_i$.

$$d_i = \|y - \hat{y}\|$$

The following general algorithm describes the necessary steps of face recognition:

**Step 1:** Input matrixes of training set $A$ and a query image $y$.

**Step 2:** For each subject $i$, solve the problem as a linear regression problem as written in Equation (3).

**Step 3:** Compute the distance as in Equation (6).

**Step 4:** Identify the subject with the minimum distance.

The LS method uses the entire set of images pixel and thus can be easily affected by the presence of noise. Hence, a robust method like the LTS-based is more preferred when dealing with noise or occlusion. The method was introduced by\(^{12}\) and is expected to perform well when outliers are present. The LTS regression coefficients and residuals are respectively formulated as follows:

$$\beta_{lts}^p = \arg \min_{\beta} \sum_{p=1}^{q/2} (e^2)_{p;q} e = y - A_i\beta^p$$

where $(e^2)_{1;q} \leq (e^2)_{2;q} \leq \cdots \leq (e^2)_{q/2;2}$ represent the squared residuals in ascending order. The LTS method will generate a close estimate as the LS when the data is clean but is expected to be more resistant when outliers are present.

### 2.1 Random LTS

This method with a low computational complexity in practice compared to the LTS estimator was introduced\(^{13}\). This algorithm is simple in terms of implementation and has a solid theoretical foundation. Besides, by using this method, we can get a regression estimator with high breakdown point. LTS formula can be also written as

$$\min_{\beta} \frac{1}{m} \sum_{i=1}^{m} e_i^2_{N(\gamma)} = \min_{\beta} \frac{1}{m} \sum_{i=1}^{N} w_i ((y_i) - A_i\beta)^2$$

where $Q = \{\bar{w} = (w_1, \ldots, w_m)\}$, $w_i = 0.1$, $\sum_{i=1}^{N} w_i = m$, $J(\bar{w}) = \frac{1}{m} \|W[A(A^TWA)^{-1}A^TW\hat{y} - \hat{y}]^2\|$. From this, we can see that the calculation of $\beta$ is reduced to finding the best weighting vector $\bar{w} \in Q$. To evaluate $J(\bar{w})$, arrange it in order as below:

$$\min_{\bar{w} \in Q} J(\bar{w})$$

The goal of the calculation is to find $J(\bar{w})_{1;j}$ and the corresponding $\bar{w} \in Q$. In another study by\(^{14}\), LTS has similar asymptotic efficiency as an M-estimator with trimming at the quartiles. As for the maximal breakdown value, LQS and LTS have similar value. It can be obtained when $\text{floor}(n+p)/2 < \text{quantile} < \text{floor}(n+p+1)/2)$.

If $n$ is large and $p$ is small, the intercept for each trial fit need to be adjusted. Hence, the residuals also need to be arranged leads to extra computation. There are few different opinions over choosing the value of $p$ sample among scholars. In\(^{15}\) only consider $p$ while\(^{14}\) claimed $p+1$ and suggested that with large samples we will get better results. The computations are exact for a model with just an intercept and adjustment, and for LQS for a model with an intercept plus one regressor and exhaustive search with adjustment. For all other cases the minimization is known to be approximate.

Concerning $n$ is small, it is conceivable to create all subsets of size $H = n/2$ to figure out which one minimizes the LTS criterion. Be that as it may, when $n$ is large, there will be an excessive amount of subsets to be created which is too huge for sensible application. Consequently, the performance of this method dropped drastically when the value of $n$ is large. To address this issue, another strategy in view of LTS was presented.
2.2 Fast LTS

According to[6] the LTS stays resistant when the noise percentage are up to 50% of the data. In[6] proposed FAST-LTS method which performs faster and better than the random LTS. The basic ideas for FAST-LTS are an inequality involving order statistics and sums of squared residuals, and techniques which we call ‘selective iteration’ and ‘nested extensions’. This method performs concentration step (C-Steps) on randomly selected subset. By taking $h$ observations and will return $h$ observations with the smallest absolute residuals. The C-steps are then performed with $h$ observations repeatedly until the final $h$ observations are obtained. During the process, it is anticipated that some outliers exit the basic subset while clean observations enter.

2.3 LTS with Genetic Algorithm

In[17] improved the Fast LTS with genetic algorithm by replacing the random subset selection part to a non-iterative procedure. A classical Genetic Algorithm (GA) draws random candidate solutions (chromosomes) which can be arrays of real values in the case of real-valued problems. With this, the coding or encoding procedures can be omitted and the process might be faster.

3. Results and Analysis

The AT&T database was used to demonstrate the performance these methods. Artificial Salt and Pepper noise with 10%, 20%, 30%, 40%, and 50% occlusion was created to a given query image. All images from both datasets were down sampled to 32x32 pixel. Half of the images were randomly selected for training and the remaining half for testing[18]. Two different levels of image resolution; resolution 72 and resolution 50 were created. Images that are in the test sample set cannot be in the training sample set. Face recognition rate represents the percentage of the total number of correctly matched images when the subject for the test image with the minimum distance is matched with the subject image from the train sample set.

3.1 AT&T Database

Table 1 gives face recognition rates of AT&T database with resolutions 72 and 50. Figure 1 gives examples of images from AT&T Database with different levels of Salt and Pepper Noise. Random LTS performs the best and the FAST-LTS method gives the lowest recognition rate (8.5%) when noise level reached 50%. Surprisingly, the recognition rates of LS method are as high as other robust methods. Overall performance dropped significantly as the noise in image pixels increased.

3.2 Yale Database

The overall result for Yale data set is shown in Table 3. Table 3 presents recognition rates for Yale data set images. If the test images are clean from the Salt and Pepper noise, the LS and LTS methods produces the highest recognition rate compared to the other three methods. The presence of noise in test image affect the recognition rate for all methods as we can see that the recognition rate for each method dropped significantly. As the level of the noise in test images increases, we can see that the performance of all methods getting better except for the LS method. The highest recognition rate can be achieved by LS method is

| Noise  | LS     | FAST_LTS | RANDOM_LTS | GA_LTS |
|--------|--------|----------|------------|--------|
| 0%     | 87.50% | 86.00%   | 88.50%     | 88.00% |
| 10%    | 82.00% | 81.50%   | 84.00%     | 81.50% |
| 20%    | 78.50% | 77.00%   | 76.50%     | 77.50% |
| 30%    | 69.50% | 71.50%   | 69.50%     | 69.00% |
| 40%    | 45.00% | 49.00%   | 59.50%     | 50.50% |
| 50%    | 25.50% | 8.50%    | 46.50%     | 34.50% |

Figure 1. Images from AT&T database with different levels of salt and pepper noise.

| Noise  | LS     | FAST_LTS | RANDOM_LTS | GA_LTS |
|--------|--------|----------|------------|--------|
| 0%     | 87.50% | 84.50%   | 84.50%     | 68.50% |
| 10%    | 85.00% | 84.00%   | 85.50%     | 84.50% |
| 20%    | 76.50% | 77.50%   | 75.50%     | 77.00% |
| 30%    | 71.50% | 73.50%   | 68.00%     | 69.50% |
| 40%    | 55.50% | 49.50%   | 59.50%     | 55.50% |
| 50%    | 30.50% | 7.00%    | 52.00%     | 45.50% |
Table 3. Face recognition rates for yale database images with size 32*32

| Noise | LS     | FAST_ LTS | RANDOM _LTS | GA_LTS |
|-------|--------|-----------|-------------|--------|
| 0%    | 80.00% | 75.56%    | 75.56%      | 80.00% |
| 10%   | 13.30% | 17.78%    | 44.44%      | 17.78% |
| 20%   | 4.44%  | 13.38%    | 37.70%      | 24.44% |
| 30%   | 6.67%  | 15.56%    | 55.56%      | 22.22% |
| 40%   | 4.44%  | 8.89%     | 75.56%      | 33.33% |
| 50%   | 4.44%  | 22.22%    | 66.67%      | 46.67% |

4.4%. The recognition rate Dynamic LTS-GA method is higher than the LTS-GA especially when the noise level in the images is 50%. Random LTS method yields the highest recognition rate when the test images are half occluded.

4. Conclusion

This paper compares the performance of LTS based methods in face recognition for occluded query image. The AT&T and Yale datasets were used to measure the face recognition rate the being contaminated with different levels of Salt and Pepper noise. Comparative analysis with existing algorithms clearly reflects the strength of random least trimmed squares approach which perform the best in recognizing images with occlusion.

5. Acknowledgement

We would like to express our appreciation to the Ministry of Higher Education (MOHE) and University Teknologi MARA for financial support under the Grant Scheme (FRGS/1/2014/ST06/UiTM/02/5) and 600-RMI/DANA 5/3/PSI (197/2013).

6. References

1. Etemad K, Chellappa R. Discriminant analysis for recognition of human face images. JOSA A. 1997; 14(8):1724–33.
2. Fidler S, Skočaj D Leonardis A. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006; 28(3):337–50.
3. Phillips PJ, et al. The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000; 22(10):1090–104.
4. Lai J, Jiang X. Robust face recognition using trimmed linear regression. ICASSP; 2013.
5. Rousseeuw PJ, Leroy AM. Robust regression and outlier detection; 2003. p. 360.
6. Rousseeuw PJC. Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association. 1990; 85(411):633–9.
7. Toh KKV, Isa NAM. Noise adaptive fuzzy switching median filter for salt-and-pepper noise reduction. IEEE Signal Processing Letters. 2010; 17(3):281–4.
8. Wagner A, et al. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 34(2):372–86.
9. Jia H, Martinez AM. Face recognition with occlusions in the training and testing sets. 8th IEEE International Conference on Automatic Face and Gesture Recognition FG’08; 2008.
10. Jiang X. Linear subspace learning-based dimensionality reduction. IEEE Signal Processing Magazine. 2011; 28(2):16–26.
11. Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010; 32(11):2106–12.
12. Rousseeuw PJ. Least median of squares regression. Journal of the American Statistical Association. 1984; 79(388):871–80.
13. Bai E-W. A random least-trimmed-squares identification algorithm. Automatica. 2003; 39(9):1651–9.
14. Marazzi A. Algorithms, Routines, and S-Functions for Robust Statistics. Taylor and Francis; 1993.
15. Rousseeuw PJ, Van Driessen K. Computing LTS regression for large data sets. Institute of Mathematical Statistics Bulletin, Citeseer; 1999.
16. Rousseeuw PJ, Van Driessen K. Computing LTS regression for large data sets. Data Mining and Knowledge Discovery. 2006; 12(1):29–45.
17. Satman MH. A genetic algorithm based modification on the LTS Algorithm for large data sets. Communications in Statistics - Simulation and Computation. 2011; 41(5):644–52.
18. Cai D, et al. Learning a spatially smooth subspace for face recognition. 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR’07); 2007.