Vol. 6: 26-32

THE UNIVERSITY OF CENTRAL FLORIDA
UNDERGRADUATE RESEARCH JOURNAL

Published
May 15, 2012

An Iteratively Reweighted Least Square Implementation for Face Recognition

By: Jie Liang
Faculty Mentor: Dr. Xin Li

ABSTRACT: We propose, as an alternative to current face recognition paradigms, an algorithm using reweighted l_2 minimization, whose recognition rates are not only comparable to the random projection using l_1 minimization compressive sensing method of Yang et al [5], but also robust to occlusion. Through numerical experiments, reweighted l_2 mirrors the l_1 solution [1] even with occlusion. Moreover, we present a theoretical analysis on the convergence of the proposed l_2 approach.

KEYWORDS: reweighted, minimization, face recognition, implementation

INTRODUCTION

Since the number of significant features hidden in massive data is frequently much smaller than the dimension of data in a standard representation, data are compressible. Under such assumption, compressive sensing (CS) represents a paradigm in which the number of measurements of high dimensional data is reduced during acquisition so that no additional compression is necessary. The reconstruction from highly incomplete measurement is possible according to Candes, Romberg, and Tao [1] and can be done by constrained l_1 minimization. Using this idea, Yang et al. cast object recognition problems as finding a sparse representation of a given test image with regard to the training images of multiple objects in [5].

Recently, a companion paper of Wright et al. [4] proposed a random l_1 minimization approach on sparse representations, which exploits the fact that the sparse representation of the training image space helps classification and is robust to noises and occlusions. However, the l_1 minimization in [4] has a computational complexity $O(d^2N^{3/2})$, where d is the number of measurements and N is the size of the training image set. This makes computation expensive for large-scale data sets. Moreover, a large dense random matrix with size of d by N has to be generated beforehand and stored during the entire processing period.

Recent literature suggests that there are alternatives to l_1 minimization, which extend the theory and techniques for compressive sensing. We propose an efficient algorithm in detecting disguised faces using reweighted l_2 minimization to replace the l_1 minimization in order to achieve more accurate detection and faster speed. Evaluated on the Yale B data set, the proposed method is much faster than the method in [4] with comparable recognition rate. Furthermore, we show that using reweighted l_2 minimization, our algorithm is more robust to random occlusion, which is critical for detecting disguised faces in a more realistic setting.

In summary, we state our main contributions.

-We provide an alternative algorithm iteratively reweighted least square (IRLS) with improvement of the accuracy and speed.

-We implement a new algorithm to face recognition that is robust to occlusion.

-Our new algorithm can be used to detect disguised faces.

RELATED WORK

Compressed Sensing

Compressive sensing proposes that if a signal is compressible in the sense that it has a sparse representation in certain basis, then the signal can be reconstructed from a small (as compared to its dimension) set of measurements. More precisely, for a sparse signal x, if we write the measurement as $y = Ax$, $y \in \mathbb{R}^d$, and A satisfying the restricted isometry property (RIP), then the signal recovery can be done by convex l_1 optimization [1].

Wright et al. Approach

One application of sparse representation to the face recognition problem with extensive experiments and comparative results appears in [5] and [4]. It provides a comparison of the accuracy of several algorithms against their own algorithm that uses the theory of compressed sensing. This central algorithm, in which the paper is mainly based, exploits the scheme of the linear programming method of l_1-minimization and outperforms the accuracy of basic principal component analysis (PCA) and various other wellknown methods of face recognition (for the Extended Yale B image database). The algorithm is a modification of conventional SRC (sparse representation-based classification) that uses a random Gaussian matrix to compute the feature matrices in a reduced dimension. This convention makes it more suitable for applying the ideas of compressed sensing. The premise is that the new system contains the same information as the original system. Then, once the image is transported to the sparse domain, a residual is calculated for each class. The identity of the class (face) is determined by the smallest residual in the Euclidean norm. The authors in [4] investigate various ways to decrease processing time and increase accuracy in identification through techniques of feature extraction such as down-sampling images, selection of a quantity of pixels, and special matrix multiplications that reduce dimension but retain information. Also, much experimentation is carried out under certain types of occlusion or corruption. Efforts are made to simulate and test practical conditions that partially include the covering of the eyes to mimic the effect of wearing sunglasses and random noise. They were also able to conclude that, in such circumstances where certain parts of the face are hidden, partitioning an image provides exceptionally improved results. The paper demonstrates the algorithm's disadvantages as well as its superiority and provides good discussion about all types of theories and methods concerning compressed sensing and face recognition.

Recognition via Sparse Representation

For convenience, images are represented as column vectors contained in $\mathbb{R}^m$.

Algorithm (SRC):

1. Input: A matrix of n training images $A \in \mathbb{R}^{m \times n}$ for k subjects, a test image $y \in \mathbb{R}^m$, a linear feature transform $R \in \mathbb{R}^{d \times m}$, and an error tolerance of ε.

2. Compute features $\tilde{y} = Ry$ and $\tilde{A} = RA$.

3. Solve the convex optimization problem.

$$\min_x \| x \| l_1 \text{ subject to } \| \tilde{y} - \tilde{A}x \| l_2 \leq \epsilon.$$

4. Compute the residuals $r_i(y) = \| \tilde{y} - \tilde{A}\delta_i(x) \| l_2$ for $i = 1, 2, \ldots, k$, where $\delta_i(x)$ sets the components of x to zeros if these components do not correspond to the i-th subject.

5. Identify $(y) = arg \min_i r_i(y)$.

ITERATIVELY REWEIGHTED LEAST SQUARES (IRLS)

The Speed Issue of l_1 Minimization

While the SRC algorithm is a breakthrough in face recognition, its speed is still not optimal. The testing takes a few seconds per testing image in a regular personal computer, and that will become an issue when the database is huge. As a result, it is not efficient to use SRC (i.e. l_1 minimization) to do face recognition on realistic database. While l_1 minimization has complexity $O(m^2 N^{3/2})$, traditional least square minimization has complexity $O(mN^2)$. Although least square optimization has the convenient closed-form solution, researchers dropped the use of least square because it will almost never find a sparse solution.

Connection Between l_1 and Weighted l_2 Minimization

We are trying to find $arg \min \|x\| l_0$ or $arg \min \|x\| l_1$ subject to $y = Ax$. Now, let's look at how a weighted l_2 minimization can mimic l_1 minimization:

$$\begin{aligned} arg \min \|Wx\|_{\ell_2} &= arg \min \sum_{j=1}^{n} w_j^2 |x_j|^2 \\ &= arg \min \sum_{j=1}^{n} \left(w_j^2 |x_j|\right) |x_j| \\ &\text{if } w_j^2 |x_j| \approx 1 \\ &\approx arg \min \sum_{j=1}^{n} |x_j| = arg \min \|x\|_{\ell_1}. \end{aligned} \quad (1)$$

This shows that with appropriate weights, l_2 minimization can produce the sparse solution that is found by l_1 minimization.

Improvement on l_1 Minimization

As you can see from the visualizations of Figure 1, l_T norm minimization is a way to further improve l_1 minimization. However, l_T minimization is a nonlinear programming problem while l_1 minimization is a linear programming problem. We want to find a practical way to simplify the demand of convex programming. Daubechies et al. in [3] provided an idea to connect l_T where $(0 < T \leq 1)$ and l_2 norm. The idea is to minimize the following functional:

$$\mathcal{L}_\tau(\vec{x}, \vec{w}, \varepsilon) = \sum_{i=1}^{N} \left[(x_i^2 + \varepsilon^2) w_i + w_i^{\delta(\tau)} \right]. \quad (2)$$

URJ THE UNIVERSITY OF CENTRAL FLORIDA UNDERGRADUATE RESEARCH JOURNAL 6: 26–32

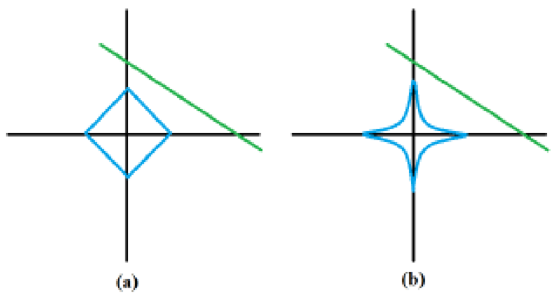

Figure 1. (a) Visualization of the l_1 ball that finds the sparse point-of-contact. (b) Visualization of l_T (0 < t < 1) ball that finds the sparse solution, and it is more desirable because it is more pointy.

Using the inequality $(\sqrt{a} - \sqrt{b})^2 \geq 0$ we have,

$$(x_i^2+\varepsilon^2)w_i + w_i^{\delta(\tau)} \geq 2\sqrt{(x_i^2+\varepsilon^2)w_i \cdot w_i^{\delta(\tau)}} \tag{3}$$

and the equality holds if $(x_i^2 + \varepsilon^2)w_i = w_i^{\delta(T)}$, i.e.

$$w_i = (x_i^2+\varepsilon^2)^{\frac{1+\delta(\tau)}{2}} \tag{4}$$

When this holds, the minimum value of the function is:

$$\begin{aligned}\mathcal{L}_\tau(\vec{x},\vec{w},\varepsilon) &= 2(x_i^2+\varepsilon^2)^{\frac{1}{2}}(x_i^2+\varepsilon^2)^{\frac{1+\delta(\tau)}{2(\delta(\tau)-1)}}\\ &= 2(x_i^2+\varepsilon^2)^{\frac{\delta(\tau)}{\delta(\tau)-1}}.\end{aligned} \tag{5}$$

If ε is sufficiently small and $\frac{\delta(\tau)}{\delta(\tau)-1} = \frac{\tau}{2}$, the minimization of the functional $\mathcal{L}_T$ is equivalent to l_T minimization (0 < t ≤ 1) and the functional becomes

$$\mathcal{L}_\tau(\vec{x},\vec{w},\varepsilon) = \sum_{i=1}^{N}\left[(x_i^2+\varepsilon^2)w_i + w_i^{\frac{\tau}{\tau-2}}\right] \tag{6}$$

Closed-Form Solution

Daubechies et al. in [3] argued that the unique solution to min || Wx || l_2 subject to $Ax = y$ is

$$\hat{x} = D^2A^T(AD^2A^T)^{-1}y, \tag{7}$$

where $D = diag(1/w_1, 1/w_2, \ldots)$. The derivation of the formula can be done by using Lagrange multipliers. Below is the verification of the fact that $\hat{x}$ is indeed the solution to the constrained minimization problem:

Proof:
1. $A\hat{x} = AD^2A^T(AD^2A^T)^{-1}y = y$.
2. If $Ax = y$ and $x \neq \hat{x}$, note $DW = I$ and $D^T = D$:

$$\begin{aligned}&(Wx - W\hat{x})^TW\hat{x}\\ &= (Wx-W\hat{x})^TWD^2A^T(AD^2A^T)^{-1}y\\ &= (ADWx - ADW\hat{x})^T(AD^2A^T)^{-1}y\\ &= 0.\end{aligned} \tag{8}$$

$$\begin{aligned}&\|Wx\|_{\ell_2}^2\\ &= \|Wx - W\hat{x} + W\hat{x}\|_{\ell_2}^2\\ &= \|Wx-W\hat{x}\|_{\ell_2}^2 + 2(Wx-W\hat{x})^TW\hat{x} + \|W\hat{x}\|_{\ell_2}^2\\ &= \|Wx-W\hat{x}\|_{\ell_2}^2 + \|W\hat{x}\|_{\ell_2}^2\\ &> \|W\hat{x}\|_{\ell_2}^2.\end{aligned} \tag{9}$$

Now we are ready to describe the iteratively reweighted least squares algorithm, which use an alternative method for choosing minimizers and weights based on the functional $\mathcal{L}_T$. Let ε be the sparsity parameter of the solution vector. We define $r(x)_i$ as the i-th largest element of the solution vector x. The smaller $r(x)_i$ is, the sparser the solution vector is.

Algorithm (IRLS)

1. Set $w^0 = (1, 1, \ldots)$ and $\varepsilon_0 = 1$.

2. Compute $\hat{x} = D^2A^T(AD^2A^T)^{-1}y$.

6: 26–32

3. Update $w_j^{n+1} = [(x_j^{n+1})^2 + \varepsilon_{n+1}^2]^{\frac{\tau}{\tau-2}}$ and

$$\varepsilon_{n+1} = \min\left(\varepsilon_n, \frac{r(x^{n+1})_{K+1}}{N}\right).$$

4. Repeat step 2 to 3. Stop the algorithm if $\varepsilon_n = 0$ or small enough.

Speed up with Gauss-Seidel method

Since in each iteration, the major change occurs at the diagonal, we want to speed up the IRLS algorithm in step 2 by not taking the inverse each time. If we rewrite the formula as

$$\begin{cases} x^{n+1} = D_n A^T z, \\ \text{where } A D_n A^T z = y, \end{cases}$$

the question will reduce to how to find z fast. Consider the Gauss-Seidel algorithm for solving the linear system $AD_nA^Tz = y$. Note that the i-th diagonal entry of AD_nA^T is for $k, l = 1, 2, ..., N$,

$$\begin{aligned}(a_{ik})D_n(a_{il})^T &= \frac{(a_{ik})}{w_k^{(n)}}(a_{il})^T \\ &= \sum_{k=1}^{N} \frac{a_{ik}^2}{w_k^{(n)}} \\ &= \sum_{k=1}^{N} a_{ik}^2[(x_k^{(n)})^2 + \epsilon_n^2]^{1/2}. \end{aligned} \quad (10)$$

Similarly, the ijth entry of AD_nA^T is for $k, l = 1, 2, ..., N$,

$$(a_{ik})D_n(a_{jl})^T = \sum_{k,l=1}^{N} a_{ik}a_{jl}[(x_k^{(n)})^2 + \epsilon_n^2]^{1/2}. \quad (11)$$

Assuming that $a_{ii} \neq 0$, the Gauss-Seidel method is

$$x_i^{(k)} = \sum_{j=1,j<i}^{n} \left(-\frac{a_{ij}}{a_{ii}}\right) x_j^{(k)} + \sum_{j=1,j>i}^{n} \left(-\frac{a_{ij}}{a_{ii}}\right) x_j^{(k-1)} - \frac{b_i}{a_{ii}}. \quad (12)$$

EXPERIMENTS

In this section, we quantitatively compare [1] the performance of SRC and IRLS using public face databases: the Extended Yale B database.

Extended Yale B Database

The Extended Yale B data consists of 2,414 frontal face imges from 38 individuals captured under various lighting conditions.

Figure 2. A sample of Extended Yale B Database. Each row shows the images of one subject under various lighting conditions.

To fairly compare the performance, we randomly select about half of the images from each person for training, and the other half for testing, as they did in [4]. The reason for randomly choosing the training set is to make sure that our results and conclusions will not depend on any special choice of the training data.

We desire to experiment with recent advancements in compressive sensing in order to accurately identify images of faces to their respective classes based on a training set of images. Based on the database (Extended Yale B) we are using for testing, we must deal with obstacles such as radical lighting conditions that change brightness levels and result in concealed areas of the face due to shadows. We also would like to examine how the algorithm can handle potential occlusion for the use of detecting disguised faces. Thus, we randomly select half of the images from each person for training. For the other half, we randomly apply a mask to the left eye, right eye, nose, left cheek, right cheek, or mouth area of the image. Figure 3 is a sample of the images we used for testing.

---

[1]We used Matlab in all experiments on a typical 3G Hz PC.

Figure 3. A sample of Extended Yale B Database with random occlusions over (left eye, right eye, nose, left cheek, right chek, and mouth).

Experimental Results

We are able to reproduce the recognition rate claimed by Yang et al in [5]. For all the experiments, the error distortion ϵ for SRC and in ours implementation is set to be 0.05. We are using L1-Magic as the optimization solver as they did. It takes a few seconds to classify one test image on a typical 3G Hz PC comparable to the speed in paper [5] where 5G Hz PC is used.

YaleB	Accuracy (%)	Training (ms/img)	Speed (ms/img)
PCA	87.3	0.4	0.6
SRC (GPSR)	95.4	0	5850
SRC(L1-Magic)	96.9	0	15083
IRLS	96.2	0	388

Table 1. Testing results for applying PCA, SRC (GPSR and L1-Magic), and IRLS minimization for Extended Yale B database. Since PCA is also a popular method, we include the result for comparison purposes.

As we claimed, our method IRLS performs comparably to SRC with faster speed. Furthermore, our algorithm is more robust to occlusion as the following table demonstrates its performance.

Occluded YaleB	Accuracy (%)
SRC(L1-Magic)	89.32
IRLS	91.24

Table 2. Testing results for applying SRC (L1-Magic) and IRLS minimization on Occluded Extended Yale B database.

CONCLUSION

-Run-time depends exponentially on the dimension of the face space.

-If sufficient measurements are extracted (≥3%) and the sparse representation is correctly found, high recognition rates can be achieved.

-Reweighted l_2 algorithm is very comparable to traditional l_1 algorithm.

-Even if there are random occlusions in the testing images, high recoginiton rates can be achieved using IRLS, provided sufficient measurements are extracted and the sparse representation is correctly found.

ACKNOWLEDGEMENTS

Very special thanks to Dr. Neils Lobo (Computer Science, UCF) and Dr. Mubarak Shah (Mathematics and Computer Science, UCF) who are also the mentors for this project. Thanks to Enrique G. Ortiz (Ph.D. Student in Computer Science, UCF) and my fellow students from the GAUSS 2010 cohort James M. Wilson and Talon Ward for the discussion on this project. Funding for this project was provided by the National Science Foundation grant number: DMS 0803059.

REFERENCES

[1] E. Candes, J. Romberg, and T. Tao. Stable Signal Recovery from Incomplete and Inaccurate Measurements. *Communications on Pure and Applied Mathematics*, Vol. 59, No. 8. (August 2006), pp. 1207-23.

[2] E. Candes, M. Wakin, and S. Boyd. Enhancing Sparsity by Reweighted L1 Minimization. *TJournal of Fourier Analysis and Applications*, 14.5-6 (2007): 877-905.

[3] I. Daubechies, R. DeVore, M. Fornasier, and C. Gunturk. Iteratively Re-weighted Least Squares Minimization for Sparse Recovery. *Communications on Pure and Applied Mathematics*, Vol. 63, No. 1. (2010), pp. 1-38.

[4] John Wright, Allen Yang, and Arvind Ganesh. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence archive*, Vol. 31, Issue 2, (2009).

[5] A. Yang, J. Wright, Y. Ma, and S. Sastry. Feature Selection in Face Recognition: A Sparse Representation Perspective. *UC Berkeley Tech Report,* No. YCB/EECS-2007-99. (2007)