Bilateral Random Projection based High-speed Face and Expression Recognition Method

Jieun Lee, Miran Heo and Yoonsik Choe
School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea

Keywords: Face Recognition, Greedy Algorithm, Random Projection, Sparse-based Representation Classification.

Abstract: Face and expression recognition problem can be converted into superposition of low-rank matrix and sparse error matrix, which have the merits of robustness to occlusion and disguise. Low-rank matrix manifests neutral facial image and sparse matrix captures emotional expression with respect to whole image. To separate these matrices, the problem is formulated to minimize the nuclear norm and L_1 norm, then can be solved by using a closed-form proximal operator which is called Singular Value Thresholding (SVD). However, this conventional approach has high computational complexity since it requires computation of singular value decomposition of large sized matrix at each iteration. In this paper, to reduce this computational burden, a fast approximation method for SVT is proposed, utilizing a suitable low-rank matrix approximation involving random projection. Basically, being associated with sampling, a low-rank matrix is modeled as bilateral factorized matrices, then update these matrices with greedy manner. Experiments are conducted on publicly available different dataset for face and expression recognition. Consequently, proposed algorithm results in the improved recognition accuracy and also further speeding up the process of approximating low-rank matrix, compared to the conventional SVT based approximation methods. The best recognition accuracy score of 98.1% in the JAFFE database is acquired with our method about 55 times faster than SVD based method.

1 INTRODUCTION

Over the past decades, face and expression recognition have been particularly influential in the field of computer vision and pattern recognition. The more basic trends have been based on Eigenfaces (Turk and Pentland, 1991), Fisherfaces (Belhumeur et al., 1997) and SVM (Support Vector Machine), which has arisen since past two decades (Yang et al., 2011). These common algorithms aim all to collect proper features from face images for recognition. Neither of these works treats the corrupted training data, and thus their recognition results are fragile to the presence of abrupt noise such as occlusion and disguise in face images (Wei et al., 2014). Despite a certain level of accurate performances of conventional face recognition algorithms, practical challenge remain regarding the dramatic variations of pose, expression and illumination. In addition, extreme illumination change such as shadows weakens the assumption of a low-dimensional linear model and then acts as occlusion for face appearance model. These points imply that face recognition should be robust to various occlusions for stable performance (Ou et al., 2014). Neutral face images of the same person lie on a low-rank subspace due to their high correlation properties, whereas facial expression can be regarded as sparse non-rigid deformation in the presence of arbitrary face regions. Namely we can employ the low-rank structure for finding the redundancy in the neutral face images since there exists the similarity between those images. Therefore researches have begun to investigate the link between low-rank or sparse structure and facial and expression recognition for high accuracy purpose (Georgakis et al., 2016). In the context of optimization skills, most works are conducted by an efficient Alternating Direction Method of Multipliers (ADMM) algorithm (Bertsekas, 2014) mostly at finding component structures. ADMM has the properties of strong optimality and practical convergence speed even for the case the objective function is non-smooth. However this method can not be extended to handle large scale dataset due to its limitation of iterative mechanism. Recently dictionary learning based study (Wei et al., 2014) promotes structural incoherence in order to improve discrimination ability. Also Augmented Lagrange multipliers (ALM) (Lin et al., 2010) as one of ADMM has been applied to solve this

Lee, J., Heo, M. and Choe, Y.
Bilateral Random Projection based High-speed Face and Expression Recognition Method.
DOI: 10.5220/0007346000990106
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 99-106
ISBN: 978-989-758-354-4

standard low-rank problem. ADMM is mainly associated with the calculation of Singular Value Thresholding (SVT) operator (Cai et al., 2010) involving SVD calculation which is the majority of the computational load. However even for this computational limit, SVT is widely used for low-rank approximation as following reasons.

Consider the singular value decomposition of a matrix $X \in \mathbb{R}^{m\times n}$ of rank r

$$X = U\Sigma V^*, \Sigma = diag\left(\{\sigma_i\}_{1\le i\le r}\right), \tag{1}$$

where U and V are respectively $m \times r$ and $n \times r$ matrices with orthonormal columns, and the singular values σ_i are positive. For each $\tau \geq 0$, the soft thresholding operator D_r is defined as follows:

$$D_r(X) := UD_r(\Sigma)V^*, D_r(\Sigma) = diag\left(t\left\{\sigma_i - \tau\right\}_+\right), \tag{2}$$

where t_+ is the positive part of t, namely, $t_+ = max(0,t)$. In other words, this operator simply applies a soft-thresholding rule to the singular values of X, effectively shrinking these towards zeros. In some sense, this shrinkage operator is a straightforward extension of the soft-thresholding rule for scalars and vectors. In particular, note that if many of the singular values of X are below the threshold τ, the rank of $D_r(X)$ may be considerably lower than that of X, just like the soft-thresholding rule applied to vectors leads to sparser outputs whenever some entries of the input are below threshold. The singular value thresholding operator is the proximity operator associated with the nuclear norm,

$$D_r(Y) = arg\min_X \left\{\frac{1}{2}\|X-Y\|_F^2 + \tau\|X\|_*\right\}. \tag{3}$$

Eq. 3 is proved in Theorem 2.1 by (Cai et al., 2010) in detail. Due to this powerful attributes, SVT is used frequently but it has a complexity equal to that of SVD, i.e., $O(mn \cdot min(m,n))$ at each iteration.

Therefore in order to solve the low-rank and sparse decomposition with a gross error term, the greedy bilateral scheme has been explored and exploited. This greedy scheme (Zhou and Tao, 2013) uses only QR decompositions, random projections, and matrix multiplications, thus, it reduces computational complexity very efficiently. In this paper, this greedy method is adopted to predict the low-rank matrix fast and a novel random projection based method is proposed in order to reduce the computational burden for low-rank approximation of classical facial recognition framework.

2 RELATED WORK

In this section, we firstly mention a brief formulation on face recognition based on (Georgakis et al., 2016). Then the optimization skill for reducing computational burden will be described in detail and the proposed overall method will be finally introduced in the next section.

2.0.1 Discriminant Incoherent Component Analysis

The goal of DICA (Discriminant Incoherent Component Analysis) (Georgakis et al., 2016) is to robustly learn components from training samples that 1) are discriminant and exhibit low-complexity structures (e.g., low-rank or sparsity) associated with facial attributes, 2) are mutually incoherent among different classes, and 3) facilitate the classification of test samples by means of sparse representation. This method learns the reconstruction matrices $\left\{U^{(i)}\right\}_{i=1}^{n_c}$ and projection matrices $\left\{V^{(i)}\right\}_{i=1}^{n_c}$ by employing the training matrix $X \in \mathbb{R}^{d\times N}$ which contains in its columns the vectorized training face images, with d being the dimensionality of each image and N the number of training observations. Also n_c denotes the total number of each class. The column of X, x represents a vectorized expressive face image. According to DICA algorithm, we can formulate the face and expression recognition problem as folowing:

$$\begin{aligned} &arg\min_{\mathcal{W}} \lambda^{(i)} \sum_{i=1}^{n_c} \left\|V^{(i)}\right\|_{(\cdot)} + \eta \sum_{i\neq j} \left\|V^{(i)}V^{(j)^T}\right\|_F^2 + \lambda_1 \|O\|_1, \\ &s.t.\ X = \sum_{i=1}^{n_c} U^{(i)}V^{(i)}X + O, \\ &\quad U^{(i)^T}U^{(i)} = I,\ i = 1,2,...,n_c. \end{aligned} \tag{4}$$

In Eq. 4, the set $\mathcal{W}$ is comprised of three components $U^{(i)}$, $V^{(i)}$ and O. Furthermore the structure-inducing norm $\left\|V^{(i)}\right\|_{(\cdot)}$ is either the nuclear norm for face-specific projections or the l_1-norm for expession-specific projections. The term $\sum_{i\neq j}\left\|V^{(i)}V^{(j)^T}\right\|_F^2$ induces mutual incoherence among the projection spaces and $O \in \mathbb{R}^{d\times N}$ denotes the outlier matrix accounting for components that cannot be explained by the summand containing the class-specific reconstructions. The positive parameters $\lambda^{(i)}$, η, and λ_1 control the norm imposed on $\left\{V^{(i)}\right\}_{i=1}^{n_c}$, the mutual incoherence for all component pairs, and the sparsity of out-

liers O, respectively. The orthonormality is endowed wih $U^{(i)}$ in order to characterize each class properly.

2.0.2 Optimization

Previously, the Alternating-Directions Method of Multipliers (ADMM) (Bertsekas, 2014) is employed to solve Eq. 4. This utilizes the partial augmented Lagrangian function for Eq. 4. At each iteration, the Lagrangian function is minimized with respect to each variable in $\mathcal{W}$ in an alternating fashion. Subsequently the Lagrange multiplier and parameter are updated too. When the nuclear norm is enforced on $V^{(i)}$, the cost of each iteration is mainly associated with the calculation of the SVT. Hence computational complexity about $V^{(i)}$ update amounts to $O\left(max\left(d^2N, dN^2\right)\right)$.

Note that the optimization equation for V can be expressed as follows

$$\begin{aligned} &arg\min_{V^{(i)}} \mathcal{L}(V^{(i)}, Y[t], \mu[t]) \\ &\quad = arg\min_{V^{(i)}} \lambda^{(i)} \left\| V^{(i)} \right\|_* + \eta \sum_{i \neq j} \left\| V^{(i)} V^{(j)^T} \right\|_F^2 + \\ &\quad \frac{\mu[t]}{2} \left\| X - \sum_{i=1}^{n_c} U^{(i)} V^{(i)} X - O + \mu[t]^{-1} Y[t] \right\|_F^2 \\ &\quad = arg\min_{V^{(i)}} \lambda^{(i)} \left\| V^{(i)} \right\|_* + f(V^{(i)}) \end{aligned} \tag{5}$$

where μ is a positive parameter and $Y \in \mathbb{R}^{d \times N}$ is the Lagrange multiplier related to the linear constraint. Eq. 5 consists of a non-smooth term, induced by the nuclear norm, and a smooth, twice differentiable term described by the function f. It can easily be proved that the gradient ∇f is Lipschitz-continuous. By linearizing f in the vicinity of the current point $V^{(i)}[t]$, and by exploiting the Lipschitz-continuity of ∇f, we obtain the following equivalent problem

$$arg\min_{V^{(i)}} \lambda^{(i)} \left\| V^{(i)} \right\|_* + \frac{1}{2} \left\| V^{(i)} - (V^{(i)}[t] - \frac{1}{L} \nabla f(V^{(i)}[t])) \right\|_F^2 \tag{6}$$

where L is the Lipschitz constant. Instead of applying the SVT operator, we choose greedy bilateral method for Eq. 6.

The heart of our SVD-free method shares the idea of Zhou *et al.* (Zhou and Tao, 2013), in that the above equation can be found by applying greedy bilateral projection to matrix instead of the original SVT-based method as illustrated in (Georgakis et al., 2016).

3 PROPOSED METHOD

Greedy strategy has strength when it is used as *warm start* of the next higher rank optimization and speeds up convergence since it is robust to biased rank estimation. In addition, its mutually adaptive updates of two factors which comprise the target matrix yields a simple yet efficient SVD-free implementation. Generally under this technique, the overall time complexity of matrix completion is only dependent on rank of the target matrix. In real world application, finding the exact low-rank matrix is intractable. There exists trade-off between accuracy and time/space costs, even if singular values of the target matrix decay fast. Although the low-rank matrix approximation in Eq. 6 is provably optimal when constructed from SVD, the expensive time cost makes SVT prohibitive to large matrix. Therefore this type of problem can be resolved by designing a suitable random projection matrix with greedy manner.

In this paper, we adopt two main algorithms by mitigating computational complexity compared to the SVT-based algorithm like DICA (Georgakis et al., 2016).

3.0.1 Greedy Bilateral Method (GBM)

To apply this method, we slightly change the problem of Eq. 6 in terms of UV factors based on the assumption of low-rank constraint on F

$$\min_{U,V} \|F - UV\|_F^2, \; s.t. \; rank(U) = rank(V) \leq r, \tag{7}$$

where $F = V^{(i)} - \left(V^{(i)}[t] - \frac{1}{L} \nabla f(V^{(i)}[t])\right)$. In U, V in Eq. 7 are different from the aboved mentioned reconstruction and projection matrices. Alternatively optimizing U and V in Eq. 7 immediately yields the following updating rules, note subscript in $\cdot_k$ denotes the variable in the k^{th} iterate and $(\cdot)^\dagger$ stands for the Moore-Penrose pseudo-inverse:

$$\begin{cases} U_k = F V_{k-1}^T \left(V_{k-1} V_{k-1}^T\right)^\dagger \\ V_k = \left(U_k^T U_k\right)^\dagger U_k^T F. \end{cases} \tag{8}$$

It can be observed that the object value in Eq. 7 is merely determined by the matrix product UV rather than individual U or V, and different (U,V) pair can produce the same UV. It is then of interest to find a pair of (U,V) that have the same product as (U_k, V_k) in Eq. 8 but can be computed in less time than U_k and V_k. This observation is represented by investigating the product $U_k V_k$,

$$U_k V_k = U_k \left(U_k^T U_k\right)^\dagger U_k^T F = \mathcal{P}_{U_k} F. \tag{9}$$

This implies that the product U_kV_k equals to the orthogonal projection of F onto the column space of U_k. According to Eq. 8, the column space of U_k can be represented by arbitrary orthonormal basis for the columns of FV_{k-1}^T.

It is worth noting that we can compute it as Q via fast QR decomposition $FV_{k-1}^T = QR$. In this case, the product U_kV_k can be equivalently computed as $U_kV_k = \mathcal{P}_Q F = QQ^TF$. Therefore U_k and V_k in Eq. 8 can be replaced by Q and Q^TF repsectively, while the product U_kV_k and the corresponding object value are kept the same. This gives a faster updating procedure

$$\begin{cases} U_k = Q,\ QR = \mathbf{qr}(FV_{k-1}^T) \\ V_k = Q^TF. \end{cases} \tag{10}$$

This relation can then be derived, from which the alternating update can be viewed as mutually adaptive optimization of right sketch FV^T and left sketch Q^TF for F. Since the right and left sketches respectively describe the column and row spaces, which largely decide the approximation precision, we can temporarily ignore the QR decomposition in order to see how the column/row space is tracked within this scheme. Actually Eq. 8 has the same accuracy as power scheme randomized SVD method (Halko et al., 2011). Different from power scheme, we updates Eq. 8 with a greedy incremental rank for both U and V. The computatioin of GBM is dominated by the two matrix multiplications that take $2dNr_i$ flops. It can be further speeded up if assigning sparsity on U and V, which will be described in the next subsection. The overall greedy bilateral solver is wrapped up in Algorithm 1.

3.0.2 Random Row-Space Projection (RRSP)

Beyond greedy bilateral method, in this section we outline a scheme that is based on the approximate SVD algorithm of Sarlos (Sarlos, 2006), (Fazel et al., 2008). This method casts SVD-free algorithm as a direct sensing the row and column space of the target matrix.

Suppose $rank(F) = r$, we again perform two sets of measurements (arbitrary random matrix) of F. Here, the output of the first set is used as the *sensing matrix* for the second set. Thus this method needs to access F and F^T to obtain the two sets of measurements sequentially. The second set of measurements are in fact quadratic in F.

We again have several choices for the sensing matrix $P \in \mathbb{R}^{r\times m}$, for example we can pick P with i.i.d. Gaussian entries. It is also possible to use structured matrices that are faster to apply, for example the SRFT (The Subsampled Randomized Fourier Transform) matrix which consists of randomly selected rows of the product of a discrete Fourier transform matrix and a random diagonal matrix (Woolfe et al., 2008). From the viewpoint of sparsity, SRFT matrix is encouraged to be the sensing matrix P. We can consider the following scheme:

- Sensing : Make linear measurement
$$Y_1 = PF,\ followed\ by\ Y_2 = Y_1F^T. \tag{11}$$
- Recovery : Given measurements $Y_{1,2}$, construct
$$\hat{F}^T = Y_1^\dagger Y_2 \tag{12}$$

The recovery step can be implemented eifficiently using a QR decomposition of Y_1.

A geometric interpretation is as follows: using $\hat{F}^T = Y_1^\dagger Y_1 F^T = (PF)^\dagger (PF) F^T$ and noting that $Y_1^\dagger Y_1$ is the orthogonal projection matrix onto the range of Y_1, we see that the estimate $\hat{F}$ is given by the projection of each row of F onto the row-space of PF, which is spanned by random linear combinations of the rows of F. That is, each row of F is approximated by its closest vector in the row-space of PF. Employing random projection matrix P is of crucial importance in extracting informative spaces from the target matrix and determining the effective rank. The methodology presented in this work is verified by the following Lemma 1.

Lemma 3.1. *(Exact Recovery) Suppose entries of P are i.i.d. Gaussian.*
If $rank(F) = r$, the scheme described in Eq. 11 and 12 yields $\hat{F} = F$ with probability one.

Proof. Let p_i denote the ith row or P (for example, from a Gaussian or SRFT ensemble). If $rank(F) = r$ the set of random vectors $F^T p_i$, $i = 1,...,r$ are linearly independent with probability one, which implies that row-space of PF is equal to row-space of F with probability one, and projecting F onto its own row-space gives F. □

To cope with relative error of SRFT matrix used, the proof is in section 5.2 (Woolfe et al., 2008).

Lemma 3.2. *Suppose P is an SRFT matrix and there are $\alpha, \beta > 1$ such that*

$$\frac{\alpha^2\beta}{(\alpha-1)^2}\left(2r^2\right) \le l < m \tag{13}$$

Then,

$$\left\|\hat{F} - F\right\| = C\sqrt{m}\cdot\sigma_{r+1}(F) \tag{14}$$

holds with probability at least $1 - \frac{1}{\beta}$. Constant C depends on α.

However when F is not low-rank structure, the truncated r-term SVD of F is approximated by $\left(Y_1^\dagger Y_2\right)_r$.

Table 1: Recognition Rates (%) For Subject-Independent Face & Expression Recognition (SIR) on **CK+ Dataset** (Lucey et al., 2010).

Method	Image Size (pixel)	SIR		Time (Sec)
		Face	Expression	
DICA	32 × 32	5.5556	61.2903	$4.3052 \cdot 10^3$ / $3.4130 \cdot 10^3$
	48 × 48	5.5556	**79.6296**	$1.7219 \cdot 10^3$ / $1.8398 \cdot 10^3$
	56 × 56	5.5556	**78.7037**	$1.6105 \cdot 10^4$ / $1.4448 \cdot 10^4$
BGM	32 × 32	27.7778	61.2903	7.5450 / 6.1080
	48 × 48	41.6667	68.5185	23.1306 / 25.1301
	56 × 56	42.5926	69.4444	141.3831 / 125.3712
RRSP [Gaussian]	32 × 32	45.3704	**72.2222**	7.3872 / 8.8441
	48 × 48	47.2222	73.1481	21.3828 / 25.57
	56 × 56	52.7778	74.0741	134.2379 / 123.5661
RRSP [SRFT]	32 × 32	**78.7037**	35.1852	9.5474 / 6.7210
	48 × 48	**81.4815**	42.5926	21.4391 / 21.5486
	56 × 56	**78.7037**	49.0741	132.4659 / 119.2481

Table 2: Recognition Rates (%) For Subject-Independent Face & Expression Recognition (SIR) on **The Japanese Female Facial Expression (JAFFE) Database** (Lyons et al., 1998).

Method	Image Size (pixel)	SIR		Time (Sec)
		Face	Expression	
DICA	32 × 32	8.3333	40.6250	$1.3333 \cdot 10^3$ / $1.3565 \cdot 10^3$
	48 × 48	4.1667	42.1875	$1.418 \cdot 10^3$ / $1.7989 \cdot 10^3$
	56 × 56	4.1667	42.1875	$6.3770 \cdot 10^3$ / $1.9214 \cdot 10^4$
BRM	32 × 32	75	**48.4375**	5.3755 / 11.8196
	48 × 48	75	**50**	10.6103 / 10.4784
	56 × 56	50	48.4375	136.8656 / 114.6812
RRSP [Gaussian]	32 × 32	**75**	**48.4375**	5.5414 / 7.2722
	48 × 48	62.5	48.4375	10.5549 / 12.2985
	56 × 56	79.1667	**50**	107.6488 / 124.2132
RRSP [SRFT]	32 × 32	**98.1**	32.8125	5.5350 / 6.1037
	48 × 48	**98.1**	28.1250	11.86 / 10.9866
	56 × 56	**98.1**	35.9375	114.9448 / 127.9194

Table 3: Recognition Rates (%) For Subject-Independent Face & Expression Recognition (SIR) on **The Yale Face Database B** (Belhumeur et al., 1997).

Method	Image Size (pixel)	SIR		Time (Sec)
		Face	Expression	
DICA	32 × 32	34.2857	33.3333	337.7671 / 194.3378
	48 × 48	31.4286	33.3333	700.5026 / 350.008
	56 × 56	31.4286	40	$7.1175 \cdot 10^3$ / $3.8683 \cdot 10^3$
BGM	32 × 32	60	**46.6667**	9.5883 / 5.7042
	48 × 48	42.8571	**43.3333**	21.2742 / 9.8354
	56 × 56	60	36.6667	201.1026 / 101.0096
RRSP [Gaussian]	32 × 32	74.2857	36.6667	10.5794 / 5.7042
	48 × 48	**82.8571**	**43.3333**	21.3447 / 9.8354
	56 × 56	**82.8571**	36.6667	196.1838 / 105.8419
RRSP [SRFT]	32 × 32	**80**	33.3333	10.1339 / 6.5179
	48 × 48	**82.8571**	**43.3333**	24.1405 / 10.3150
	56 × 56	71.4286	**43.3333**	205.8391 / 103.4070

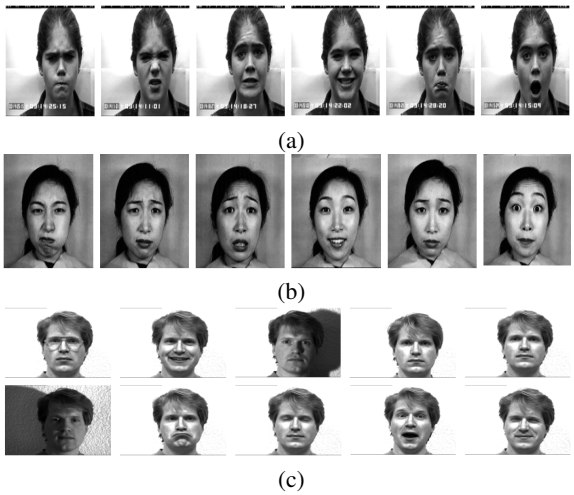

Figure 1: Example images from each of the datasets used. (a) CK+ (Lucey et al., 2010). (b) The Japanese Female Facial Expression (JAFFE) Database (Lyons et al., 1998). (c) The Yale Face Database B (Belhumeur et al., 1997).

4 EXPERIMENTAL RESULTS

In this section, the effectiveness of our random projection based approach is verified through a number of experiments. Our method explores concrete application, face and expression recognition on CK+, JAFFE and The Yale Face Database B dataset. The experimental setting is identical to that of DICA algorithm (Georgakis et al., 2016). Specifically, training and test dataset for expression recognition is comprised of the six universal emotions (*Anger, Disgust, Fear, Happiness, Sadness and Surprise*).

CK+ has been widely used for the task of face and expression recognition. It contains 123 subjects in a total of 593 sequences, 327 out of which are annotated with respect to the emotion portrayed. We does not consider the temporal dimension, only the last 4 frames are used as expressive images, as those are close to the apex phase of the expression. JAFFE database contains 213 images of 6 basic facial expressions posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. The Yale Face Database B database contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions. For every subject in a particular pose, an image with ambient background illumination was captured. We randomly extract the number of training subjects 25, 10 and 15 for CK+, JAFFE and Yale Face Database B respectively.

Addtionally, to examine how dimensionality of the image affects accuracy, the following experiments Tab. 1, 2, and 3 are conducted. Only the choice of $48 \times 48, 56 \times 56$ pixels for the image size of the DICA algorithm (Georgakis et al., 2016) of CK+ Dataset leads to the best performance. Except in that event, the proposed methods achieve the best results and are simultaneously conducted with fast time. In face recognition works, RRSP based on SRFT projection matrix performs the best, primarily due to test images being associated with sparse linear combinations of similar faces rather than similar expressions in the dictionary. We achieved the best performances on the JAFFE dataset because it consist of no dramatic illumination or pose changes.

On average the computation time of DICA (Georgakis et al., 2016) is about 130.7098, 113.3263 and 35.6343 times higher than the random projection based methods (*i.e.*, Bilateral Greedy method, RRSPs) in CK+, JAFFE and The Yale Face Database B, respectively. This is due to the iterative steps of the cost function in ADMM algorithm of DICA because the shrinkage operator about L_1 norm becomes the most time-consuming calculation, thus entailing linear complexity $O(dN)$.

We remark that these experimental results are feasible because the face and expression recognition system does not need to restore the whole pixels of the structured image with perfect accuracy. It clearly shows that our method can handle data in large scale.

5 CONCLUSION

We propose a novel method for face and expression recognition that utilizes two simple and extensible random projection based optimization algorithms. The proposed method updates two factors of target matrix in ways of bilateral and direct projection methods, maintaining the accuracy of extracting low-rank matrix. In experiments, we compare our approach with a SVD-based method through real-world examples. Experimental results depict that excellent face and expression recognition results up to 98.1% can be obtained with a surprisingly small amount of time by 55 times smaller than SVD based method, such as DICA.

ACKNOWLEDGEMENTS

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program(10073229, Development of 4K high-resolution image based LSTM network deep learning process pattern recognition algorithm

Algorithm 1: Greedy bilateral solver.

Input : Data $X \in \mathbb{R}^{d\times N}$. Parameters : $\lambda^{(i)}, \eta, \lambda_1$, and $\left\{m^{(i)}\right\}_{i=1}^{n_c}$, Objective function f, rank step size Δr, power K.

Output: $\left\{U^{(i)} \in \mathbb{R}^{d\times m^{(i)}}, V^{(i)} \in \mathbb{R}^{m^{(i)}\times d}\right\}_{i=1}^{n_c}, O \in \mathbb{R}^{d\times N}$.

Normalize each column of X to unit l_2-norm

Initialize : Set $\left\{\left\{U^{(i)}[0]\right\}, \left\{V^{(i)}[0]\right\}\right\}_{i=1}^{n_c}$, $O[0], Y[0]$ to zero matrices. Set $\mu[0] = 1/\|X\|, \rho = 1.1, \mu_{max} = 10^{10}$.

while *not converged* **do**

  for $i = 1 : n_c$ **do**

    Calculate $L = 1.02\lambda_{max}\left[\mu[t]XX^T + 2\eta\sum_{i\neq j} V^{(j)}[t]^T V^{(j)}[t]\right]$;

    if $V^{(i)}$ *is associated with nuclear norm* **then**

      for $k \leftarrow K$ **do**

        $U_k = Q$, $QR = \mathbf{qr}(FV_{k-1}^T)$;

        $V_k = Q^T F$;

      end

      Calculate the top Δr right singular vectors v (or Δr-dimensional random projections) of $\frac{\partial f}{\partial V}$;

      Set $V := [V; v]$;

    else if $V^{(i)}$ *is associated with* l_1*-norm* **then**

      $V^{(i)}[t+1] \leftarrow \mathcal{S}_{\frac{\lambda^{(i)}}{L}}\left[V^{(i)}[t] - L^{-1}\nabla f(V^{(i)}[t])\right]$;

    else

      $U^{(i)}[t+1] \leftarrow \mathcal{P}\left[\left(X - \sum_{i\neq j} U^{(j)}[t]V^{(j)}[t+1]X - O[t] + \mu[t]^{-1}Y[t]\right)\left(V^{(i)}[t+1]X^T\right)\right]$;

    end

  end

  $O[t+1] \leftarrow \mathcal{S}_{\frac{\lambda_1}{\mu[t]}}\left[X - \sum_{i=1}^{n_c} U^{(i)}[t+1]V^{(i)}[t+1]X + \mu[t]^{-1}Y[t]\right]$;

  Update the Lagrange multiplier by $Y[t+1] \leftarrow Y[t] + \mu[t]\left(X - \sum_{i=1}^{n_c} U^{(i)}[t+1]V^{(i)}[t+1]X - O[t+1]\right)$;

  Update μ by $\mu[t+1] = \min(\rho\cdot\mu[t],\ \mu_{max})$;

end

for real-time parts assembling of industrial robot for manufacturing) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea)

REFERENCES

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Technical report, Yale University New Haven United States.

Bertsekas, D. P. (2014). *Constrained optimization and Lagrange multiplier methods*. Academic press.

Cai, J.-F., Candès, E. J., and Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982.

Fazel, M., Candes, E., Recht, B., and Parrilo, P. (2008). Compressed sensing and robust recovery of low rank matrices. In *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pages 1043–1047. IEEE.

Georgakis, C., Panagakis, Y., and Pantic, M. (2016). Discriminant incoherent component analysis. *IEEE transactions on image processing*, 25(5):2021–2034.

Halko, N., Martinsson, P.-G., and Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288.

Lin, Z., Chen, M., and Ma, Y. (2010). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 94–101. IEEE.

Lyons, M., Akamatsu, S., Kamachi, M., and Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 200–205. IEEE.

Ou, W., You, X., Tao, D., Zhang, P., Tang, Y., and Zhu, Z. (2014). Robust face recognition via occlusion dictionary learning. *Pattern Recognition*, 47(4):1559–1572.

Sarlos, T. (2006). Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 143–152. IEEE.

Turk, M. A. and Pentland, A. P. (1991). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE.

Wei, C.-P., Chen, C.-F., and Wang, Y.-C. F. (2014). Robust face recognition with structurally incoherent low-rank matrix decomposition. *IEEE Transactions on Image Processing*, 23(8):3294–3307.

Woolfe, F., Liberty, E., Rokhlin, V., and Tygert, M. (2008). A fast randomized algorithm for the approximation of matrices. *Applied and Computational Harmonic Analysis*, 25(3):335–366.

Yang, M., Zhang, L., Yang, J., and Zhang, D. (2011). Robust sparse coding for face recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 625–632. IEEE.

Zhou, T. and Tao, D. (2013). Greedy bilateral sketch, completion & smoothing. In *International Conference on Artificial Intelligence and Statistics*. JMLR. org.