

Northumbria Research Link

Citation: Shen, Yiran, Yang, Mingrui, Wei, Bo, Chou, Chun Tung and Hu, Wen (2017) Learn to Recognise: Exploring Priors of Sparse Face Recognition on Smartphones. IEEE Transactions on Mobile Computing, 16 (6). pp. 1705-1717. ISSN 1536-1233

Published by: IEEE

URL: <http://dx.doi.org/10.1109/TMC.2016.2593919> <<http://dx.doi.org/10.1109/TMC.2016.2593919>>

This version was downloaded from Northumbria Research Link: <http://nrl.northumbria.ac.uk/36434/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



Northumbria
University
NEWCASTLE



UniversityLibrary

Learn to Recognise: Exploring Priors of Sparse Face Recognition on Smartphones

Yiran Shen, *Member, IEEE*, Mingrui Yang, *Member, IEEE*, Bo Wei, *Member, IEEE*,
Chun Tung Chou, *Member, IEEE*, and Wen Hu, *Senior Member, IEEE*

Abstract—Face recognition is one of the important components of many smart devices apps, e.g., face unlocking, people tagging and games on smart phones, tablets, or smart glasses. Sparse Representation Classification (SRC) is a state-of-the-art face recognition algorithm, which has been shown to outperform many classical face recognition algorithms in OpenCV, e.g., Eigenface algorithm. The success of SRC is due to its use of ℓ_1 optimization, which makes SRC robust to noise and occlusions. Since ℓ_1 optimization is computationally intensive, SRC uses random projection matrices to reduce the dimension of the ℓ_1 problem. However, random projection matrices do not give consistent classification accuracy as they ignored the prior knowledge of the training set. In this paper, we propose to exploit the prior knowledge of the training set to improve the recognition accuracy. It first learns the optimized projection matrix from the training set to produce consistent recognition performance then applies ℓ_1 -based classification based on the group sparsity structure of SRC to further improve the recognition accuracy. Our evaluations, based on publicly available databases and real experiment, show that face recognition using optimized projection matrix is 8-17 percent more accurate than its random counterpart and Eigenface algorithm, and the recognition accuracy can be further improved by up to 5 percent by exploiting group sparsity structure. Furthermore, the optimized projection matrix does not have to be re-calculated even if new faces are added to the training set. We implement the SRC with optimized projection matrix on Android smartphones and find that the computation of residuals in SRC is a severe bottleneck, taking up 85-90 percent of the computation time. To address this problem, we propose a method to compute the residuals approximately, which is 50 times faster with little sacrificing recognition accuracy. Lastly, we demonstrate the feasibility of our new algorithm by the implementation and evaluation of a new face unlocking app and show its robustness to variation of poses, facial expressions, lighting changes, and occlusions.

Index Terms—Face Recognition, smartphones, random matrices, sparse representation, group sparsity, android, JavaCV/OpenCV, face unlocking

1 INTRODUCTION

FACE recognition is an important research problem in computer vision. With the availability of Open Source Computer Vision (OpenCV: opencv.org) on both Android and iOS, face recognition has also found its way into many smart devices apps such as people tagging and mobile gaming. In fact, we have observed more than 500 Android apps making use of face recognition. As an example, a recent update of Android includes a face unlocking app which replaces traditional password-access control with face recognition.

There are three face recognition algorithms in OpenCV: EigenFace [1], FisherFace [2] and LBPFace [3]. All these three algorithms use feature extractions (where different algorithms use different set of features) followed by the nearest neighbourhood classifier (NNC) to match the test image with

a best fit image in the training set. Recently, Wright, et al. [4] propose a new face recognition classifier based on sparse representation classification (SRC). SRC introduces a few new ideas into the face recognition algorithms. First, SRC uses image pixels directly which means feature selection is not required. Second, it uses ℓ_1 optimisation to make face recognition *robust* to lighting changes and occlusion. SRC has been shown in [4] to outperform NNC-based algorithms such as EigenFace and FisherFace. However, although ℓ_1 optimisation makes SRC robust, it introduces intensive computational cost [5]. In order to reduce the processing time while preserving the accuracy of recognition, SRC uses random projection matrices to reduce the dimensionality of ℓ_1 minimisation. In particular, SRC uses random Bernoulli or Gaussian matrices because of their information preserving properties, inspired by the recent theory of compressive sensing (CS) [6], [7]. We will refer to the SRC algorithm based on random projection matrices as rand-SRC.

Although random projection matrices can significantly reduce the computation time of SRC, the classification accuracy of rand-SRC is *not* consistent. The projection matrices are randomly generated irrespective to the training set. In fact, the accuracy of rand-SRC can change by up to 15 percent depending on the random projection matrix used. As the training set is known and collected, it is possible to learn an optimal projection matrix for the specific training set. The sparse representation of the face recognition tends to have group sparsity structure [8] which can be exploited in the

- Y. Shen is with the College of Computer Science and Technology, Harbin Engineering University, Harbin, HLJ 0086, China. E-mail: shenyiran@hrbeu.edu.cn.
- M. Yang is with the Department of Radiology, Case Western Reserve University, Cleveland, OH 44106. E-mail: mingrui.yang@case.edu.
- B. Wei is with the Department of Computer Science, University of Oxford, Oxford OX1 2JD, United Kingdom. E-mail: bo.wei@cs.ox.ac.uk.
- C.T. Chou and W. Hu are with the School of Computer Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia. E-mail: {ctchou, wenh}@cse.unsw.edu.au.

Manuscript received 20 Aug. 2015; revised 7 June 2016; accepted 19 July 2016. Date of publication 25 Aug. 2016; date of current version 3 May 2017. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TMC.2016.2593919

optimisation model to further improve the recognition accuracy. In this paper, we first propose a novel strategy to optimise the projection matrix to remove the variability in the performance of SRC. We will refer to SRC that uses an optimised projection matrix as opti-SRC. We then apply optimisation model with/ without considering the group sparsity structure to solve the sparse representation classification problem. We refer to opti-SRC using group sparsity model as opti-GSRC. As the results, in addition to providing consistent classification accuracy, opti-SRC and opti-GSRC, respectively, is up to 17 and 22 percent, more accurate than rand-SRC and the Eigenface algorithm. The contributions of this paper are:

- We propose a novel strategy to optimise the *row coherence* of a projection matrix while preserving low *mutual coherence*. The resulting combinatorial optimisation problem has a large search space and we propose an efficient off-line heuristic based on tabu search. The SRC algorithm based on the optimised projection matrix significantly outperforms its counterpart based on random projection matrices as well as face recognition algorithms in OpenCV. The optimised projection matrix is also robust in the sense that it does not have to be re-optimised even if new faces are added to the training set after optimisation. We also show that our optimised projection matrix outperforms other existing methods of optimising projection matrices [9], [10].
- We propose opti-GSRC to exploit group sparsity structure inherent from sparse representation classification problem to further improve the recognition accuracy. Although the optimisation solver for opti-GSRC is too computationally intensive for in-situ implementation on smartphones, the cloud based approach which shifts the computation burden to high performance server or cloud can be the solution.
- We implement the lightweight opti-SRC on Android platforms to evaluate its efficiency. We find that the computation of residuals in SRC is a severe bottleneck, taking up 85-90 percent of the computation time. To address this problem, we propose a method to compute the residuals approximately. The method reduces the residual computation time by 50-fold while maintains the classification accuracy.
- We evaluate the efficiency, in terms of computation time and energy consumption, of opti-SRC on three different smartphone models. We find the efficiency of opti-SRC is comparable to the OpenCV algorithms.

The organisation of this paper is as follows. We provide a brief introduction to SRC and matrix coherences in Section 2. In Section 3, three challenges and solutions on optimising performance of sparse representation classification are discussed. Section 4 evaluates the performance of opti-SRC and opti-GSRC using two publicly available databases. Section 5 evaluates opti-SRC on different smartphone platforms. Section 6 discusses related work and Section 7 concludes the paper.

2 TECHNICAL BACKGROUND

In this section, we introduce the rand-SRC face recognition algorithm in [4] and discuss the roles of matrix coherences in sparse representation.

2.1 Sparse Representation Classifier (SRC)

In [4], the authors formulate the face recognition as a sparse representation problem computed via ℓ_1 optimisation. The formulation uses a random projection matrix for dimensionality reduction. The steps of rand-SRC are:

(i) *Dictionary and Sparse Representation*. To model face recognition as a sparse representation problem, one needs to first build a dictionary \mathcal{D} . We assume there are K subject classes and T training images per class. All the images used should be scaled into the same size. Each training image consists of p pixels and is vectorised into a p -dimensional column vector. We then assemble the vectorised training images of the i th subject in a $p \times T$ sub-dictionary D_i . A $p \times KT$ dictionary $\mathcal{D} = [D_1, D_2, \dots, D_K]$ is then formed from the K classes. Let y denotes a vectorised test image, then its representation under the dictionary \mathcal{D} is obtained by solving the following linear equation with the knowledge of y and \mathcal{D} :

$$y = \mathcal{D}\theta, \quad (1)$$

where the unknown vector θ contains $n = KT$ unknowns which is equal to the number of columns in \mathcal{D} . If the vectorised test image y belongs to the k th class, then *ideally* y is within the space spanned by the T vectors in D_k class and independent of the other classes. If the ideal condition holds, then the representation vector θ for y has the form

$$\theta = [0, 0, \dots, \alpha_{k,1}, \alpha_{k,2}, \dots, \alpha_{k,T}, \dots, 0, 0, \dots, 0]^T, \quad (2)$$

where T denotes the matrix transpose, and the non-zero elements appear only in those positions related to the k th class in \mathcal{D} . If the number of classes K is large, then θ is a *sparse* vector if the ideal condition holds.

(ii) *Random Projections*. Because the dimension p of the image vectors is huge, solving Equation (1) can be computationally expensive. A random projection matrix can be applied to improve the computational efficiency while preserving recognition accuracy. The random projection matrix Φ in this application is generated from a Gaussian distribution with zero mean and unit variance [4] and does not consider the prior knowledge of the dictionary. Incorporating an $m \times p$ Gaussian matrix Φ in Equation (1), we have

$$\Phi y = \Phi \mathcal{D}\theta, \quad (3)$$

where $m \ll n$ makes the systems of linear equations under-determined. Since we are looking for a sparse representation θ , we aim to solve the following ℓ_0 optimisation problem

$$\hat{\theta} = \arg \min \|\theta\|_0 \quad \text{subject to } \Phi y = \Phi \mathcal{D}\theta, \quad (4)$$

where $\hat{\theta}$ is the sparse representation of y under dictionary \mathcal{D} and $\|\cdot\|_0$ represents the ℓ_0 norm, which counts the number of non-zero coefficients in $\hat{\theta}$. We remark that a basic requirement is that each test image must have a *unique sparsest representation* under training set \mathcal{D} ; this will be discussed further in Section 2.2.2. The optimisation problem (4) is *NP-hard* [11], which means no known algorithms can solve the problem within polynomial time.

(iii) ℓ_1 *Optimisation*. Inspired by the recent theory of CS, the solution of ℓ_0 optimisation in Equation (4) can be well

approximated by the following ℓ_1 optimisation problem,

$$\theta_{\text{opt}} = \arg \min \|\theta\|_1 \quad \text{subject to } \|\Phi y - \Phi \mathcal{D} \theta\|_2 < \epsilon, \quad (5)$$

where ϵ is a small positive value used to account for noise. The solution θ_{opt} from the ℓ_1 optimisation is used in the following classification procedure.

(iv) *Minimal Residual for Classification.* After obtaining the coefficient vector θ_{opt} , we can determine the class of the test vector y by using residuals. The residual for class i is:

$$r_i = \|y - \mathcal{D}_i \theta_{\text{opt}}^{(i)}\|_2, \quad (6)$$

where $\theta_{\text{opt}}^{(i)}$ is a T -dimensional vector containing the T elements in θ_{opt} related to class i . Then the final classification is determined by

$$\hat{i} = \arg \min_{i=1,2,\dots,K} r_i, \quad (7)$$

i.e., the class having the minimal residual among all classes.

2.2 Random Matrices and Coherence

Though random projection matrices can significantly reduce the dimension of the optimisation problem, they introduce substantial variation in classification accuracy too. Moreover, random projection matrices are not optimal. We will show later on that an optimised projection matrix can significantly improve the classification accuracy of SRC. We will optimise projection matrix using row coherence and this section provides background on matrix coherences.

2.2.1 Matrix Coherences

We consider the coherences of the sensing matrix A , where $A = \Phi \mathcal{D}$, i.e., the product of the projection matrix Φ and dictionary \mathcal{D} .

Mutual Coherence. Let a_i and a_j denote the i th and j th columns of A respectively. The mutual coherence $\mu(A)$, which is the coherence between columns, is defined as:

$$\mu(A) = \max_{i < j} \frac{|a_i^T a_j|}{\|a_i\|_2 \|a_j\|_2}. \quad (8)$$

In words, $\mu(A)$ is the maximum absolute value of cross-correlations between the columns of A which stays in between 0 and 1.

Row Coherence. Let \tilde{a}_i and \tilde{a}_j denote the i th and j th rows of A respectively. The row coherence is defined as

$$\nu(A) = \max_{i < j} \frac{|\tilde{a}_i \tilde{a}_j^T|}{\|\tilde{a}_i\|_2 \|\tilde{a}_j\|_2}, \quad (9)$$

which is the maximum absolute value of cross-correlations between the rows of the matrix A .

Mutual coherence has been well studied in literature [12], [13], while row coherence has not attracted as much attention. In fact, *both* mutual coherence and row coherence affect the classification accuracy of SRC as discussed in Section 2.2.2.

2.2.2 The Importance of Coherences

A basic requirement for SRC to work is that each test image must have a unique sparsest representation under training

set \mathcal{D} in (4). According to [14], the uniqueness of sparsest representation requires the following two conditions to hold

$$\mu(A) \leq \frac{c_0}{\log n}, \quad s \leq \frac{c_0 n}{\log n \cdot \|A\|_2^2}, \quad (10)$$

where $\mu(A)$ is the mutual coherence, c_0 is a constant and s is the sparsity of θ . $\|A\|_2$ is the 2-norm of the sensing matrix (also known as spectral norm) which equals to the largest singular value of A . These two conditions guarantee that a unique sparsest representation exists with a probability of $1 - O(n^{-1})$.

The second condition of (10) can be rewritten as

$$\frac{1}{\|A\|_2^2} \geq \frac{\log n \cdot s}{c_0 n}, \quad (11)$$

which means a smaller spectral norm for A makes it easier to satisfy this condition.

We now impose the standard condition that the columns of A has unit norm. By applying the theory of tight frames [15], the spectral norm of the $m \times n$ matrix A (which equals to the redundancy of A) has the property

$$\|A\|_2^2 \geq n/m, \quad (12)$$

where the equality holds if the rows of the matrix A are orthogonal to each other or in other words, the row coherence of A is zero. This inequality suggests that a way to reduce $\|A\|_2$ is to increase the number of projections m . Alternatively, we can reduce $\|A\|_2$ by making the matrix A more orthogonal by reducing its row coherence. This will be the approach taken by this paper. In particular, we propose a method to minimise the row coherence while preserving low mutual coherence. Intuitively, a sensing matrix with a lower row coherence means there is less overlap in the information provided by different projections. In fact, we will show a lower row coherence leads to better classification accuracy.

3 CHALLENGES AND PROPOSED SOLUTIONS

There are a number of challenges to produce robust and efficient face recognition via SRC on smartphones. We will point out the challenges and present the proposed solutions.

3.1 Challenge I: Random Matrices

As mentioned above, the recognition accuracy of rand-SRC varies substantially with different random projection matrices. To solve this challenge, we propose an approach to optimise the projection matrix by minimising the row coherence while preserving relatively low mutual coherence of the sensing matrix. Moreover, opti-SRC requires a smaller number of projections compared to rand-SRC to achieve the same recognition accuracy. This means opti-SRC requires less computational requirement than rand-SRC on smartphones.

3.1.1 Optimizing Row Coherence in Finite Space

From the discussion in Section 2.2.2, a good sensing matrix requires *both* low row coherence and mutual coherence. A naïve method is to minimise the row coherence alone. While

such a method can reduce row coherence, it can at the same time increase the mutual coherence significantly, which is undesirable. To address this problem, we propose an optimisation approach in a finite feasible space to minimise the row coherence of the sensing matrix while preserving a low mutual coherence.

It is well understood that random sampling strategies, such as projection matrices Φ generated through either i.i.d. Gaussian or Bernoulli distributions, produce sensing matrices with relatively low mutual coherence [16].

Let us assume we have a finite (but still very large) set Ω of projection matrices stemming from either a random Gaussian or Bernoulli distribution. We know that any projection matrix Φ stemming from the set Ω will result in a sensing matrix A which with high probability has relatively low mutual coherence. Furthermore, we know from Equation (9) that the row coherence of A can be minimised by choosing the sampling matrix Φ that minimises

$$\arg \min_{\Phi} v(\Phi \mathcal{D}) \quad \text{subject to } \Phi \subseteq \Omega. \quad (13)$$

The strength of this strategy lies in its ability to directly reduce the row coherence of A while with high probability maintaining a relatively low mutual coherence. The high probability guarantees that the sensing matrix A will have low mutual coherence stems from the fact that Ω is finite. For example, if Ω contained the infinite set of all possible projection matrices drawn from a Gaussian distribution then there clearly exists (but is extremely unlikely when drawn from a random distribution) a sensing matrix that has zero row coherence, but unity mutual coherence. Since the objective in Equation (13) is searching for the Φ with the lowest row coherence of the sensing matrix in the infinite set, it would inevitably choose this projection matrix which would result in a sensing matrix A that has the high mutual coherence.

By constraining Ω to a finite set of randomly generated projection matrices one then guarantees with overwhelmingly high probability that all projection matrices in the set will produce a sensing matrix with relatively low mutual coherence. It is clear that this probabilistic guarantee on the mutual coherence is a function of the size of the finite set Ω . It is also clear that since the set Ω is finite and random the set is not convex making the objective in Equation (13) non-convex. Further, the search space for this combinatorial optimisation problem is enormous (even though Ω is finite) and is in fact substantially complex in computation. We remark that there is another reason why we restrict ourselves to find the optimal projection matrix within the set of Bernoulli or Gaussian matrices, rather than the set of all matrices. This will be discussed in Section 3.3.

3.1.2 Efficient Solutions

Because the optimisation problem (13) is NP-hard. We propose an efficient algorithm based on: Tabu search to minimise the objective although neither strategy is guaranteed of finding the global minima.

The algorithm explores the role of the set size Ω versus performance of recognition accuracy. We found that the larger the chosen set the better the recognition accuracy so in all our experiments we chose the largest finite set Ω possible for a tractable solution. This result is not entirely unexpected due

to the strong probabilistic guarantees of a random sampling forming sensing matrices with good mutual and row coherence.

Algorithm 1. Tabu Search Projection Matrix Optimisation

- 1: Input: Dictionary \mathcal{D} , search space \mathcal{S}_n and number of rows of the projection matrix m .
 - 2: Initialisation: allocate four empty lists: L_{in} , L_{out} , \bar{L} and \hat{L} , stop criteria $StabilityLimit = 500$, $BestCov = Inf$, $i = 0$
 - 3: **while** $i < StabilityLimit$ **do**
 - 4: $CurrentCov = Inf$, $j = 0$
 - 5: Randomly choose m vectors to form $\hat{\mathcal{P}}_m$, record the indices list of $\hat{\mathcal{P}}_m$ as \hat{L} ($\hat{L} \cap L_{out} = \emptyset$, $\hat{L} \cap L_{in} = L_{in}$). The lists are first-in-first-out;
 - 6: **while** $j < m$ **do**
 - 7: $\hat{\mathcal{P}}_{m-1} = \hat{\mathcal{P}}_m \setminus \{\hat{\mathcal{P}}_m\{j\}\}$
 - 8: $\mathcal{S}_{n-m+1} = \mathcal{S}_n \setminus \hat{\mathcal{P}}_{m-1}$
 - 9: $Cov_j = \min_{s_d \in \mathcal{S}_{n-m+1}} \max_{p_k \in \hat{\mathcal{P}}_{m-1}} |(p_k \mathcal{D})^T s_d \mathcal{D}|$
 - 10: // s_d and s_k are rows from \mathcal{S}_{n-m+1} and $\hat{\mathcal{P}}_{m-1}$ respectively.
 - 11: **if** $Cov_j < CurrentCov$ **then**
 - 12: $CurrentCov = Cov_j$, $PotentialObj = \hat{\mathcal{P}}_m\{j\}$
 - 13: Store indices of current solution $\hat{\mathcal{P}}_m$ in \bar{L}
 - 14: **end if**
 - 15: $j++$
 - 16: **end while**
 - 17: **if** $CurrentCov > BestCov$ **then**
 - 18: $BestCov = CurrentCov$; $BestList = \bar{L}$
 - 19: Clear L_{out} , push $PotentialObj$ in L_{in}
 - 20: $i = 0$
 - 21: **else**
 - 22: Push $\hat{\mathcal{P}}_m$ in L_{out}
 - 23: $i++$
 - 24: **end if**
 - 25: **end while**
 - 26: Outputs: The indices list of the optimised projection matrix in $BestList$ chosen from the search space \mathcal{S}_n
-

Tabu Search Algorithm. (Algorithm 1) Greedy search is one of the most popular used heuristics to search in large space. However, the solution of the greedy algorithm is known to converge to a local optimum which heavily depends on the starting point. This local optimum could be significantly different from the global optimum. To address this problem, we propose another algorithm based on tabu search [17] to find an improved and robust solution. Tabu search is metaheuristic algorithm for combinatorial optimisation problems. It enhances the performance of the local search algorithm by exploring the unreached area of the greedy algorithms. The computation complexity of tabu search is $O(n^2)$ [18] which means it can be solved in polynomial time.

Tabu search algorithm utilises memory structures by defining the neighborhood of the current solution and two tabu lists. The size of the neighbourhood can be as large as the size of the search space. It can be tuned according to the processing capability of the devices. The tabu lists store the specific atoms. In our case, one tabu list stores the row indices which can not be visited for a specific number of iterations and the other list stores the row indices that cannot be removed from the solution for another number of iterations. The number of iterations are determined by the size of the tabu lists. The tabu lists should be large enough to avoid

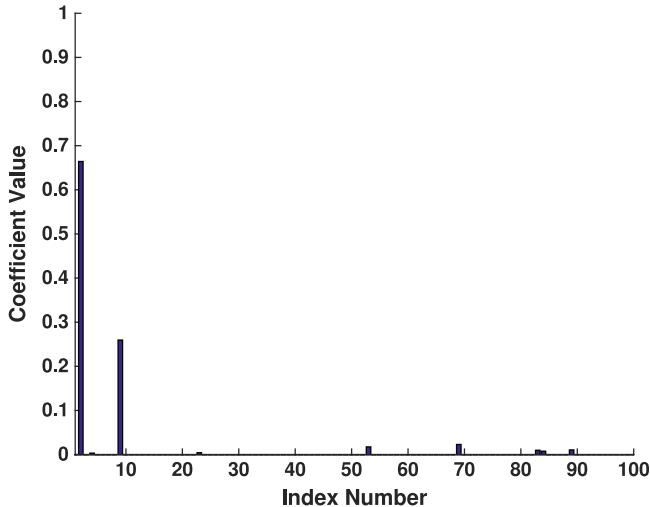


Fig. 1. Distribution of sparse coefficients.

cycles and we tuned them as half of the size of search space and the solution respectively. Details of the algorithm are shown in Algorithm 1.

This algorithm aims to pick m feasible rows (to form the projection matrix) in the search space to minimise the row coherence of the sensing matrix. Lines 1-2 represent the initial setup which specifies some parameters and creates two empty tabu lists. The proper algorithm starts in Line 3 by randomly choosing m rows in the search space satisfying the constraints of the tabu lists. This loop will continue searching for an optimal solution until the solution has not been updated by a number of *StabilityLimit* iterations. During the inner loop presented in Lines 6-16, it updates one of the rows in the chosen matrix by one row in the defined neighbourhood to find the local minimum (this is called *Intensification* stage). When the new minimum is larger than the one found previously, the algorithm explores the unreached area in the next iteration in *Diversification* stage (Lines 17-23); otherwise, when *Aspiration Level condition* (Line 17) is satisfied, it will reset the status of the algorithm (counter of the iteration, tabu list). The algorithm will terminate when the minimum has not been updated by a number of *StabilityLimit* (in Line 3) iterations. *StabilityLimit* should be large enough to guarantee the robustness of the algorithm. During the search progress, the algorithm will not search the recently reached area due to constraints of the tabu lists. By these approaches, tabu search can avoid converging to a local minimum by exhaustively searching the new area.

With the tabu search algorithm, we can obtain an optimised projection matrix. We replace the random matrix used in SRC approach with the optimised matrix to improve the performance of recognition accuracy. To distinguish SRC with different matrices, we call original SRC as rand-SRC and our approach as opti-SRC.

3.2 Challenge II: Group Sparse Structure

As discussed in Section 2.1, the sparse representation of the face images can be obtained by solving ℓ_1 optimisation problem. The ℓ_1 -norm is applied in Eq. (5) to boost the sparsity of the optimisation solution. Although ℓ_1 -norm is shown to be widely useful in CS and sparse representation, the performance can be even better with a more realistic

model by considering the prior knowledge of the specific application.

Model-based CS [8] was introduced to reduce the degrees of freedom of the sparse coefficients. The prior knowledge of the specific signals is exploited and only certain configurations of the sparse representations are allowed. By imposing more constraints in the optimisation problem, the true sparse representation can be more distinguishable from the artifacts. A sparse representation problem is regarded as having certain *Group Sparsity* structure if the solution tends to have grouping property. Namely, the dominant non-zeros only exist in few number of groups. Fig. 1 presents an example of the sparse representation of a classification problem. The training set in the example contains 10 classes and each class has 10 atoms (face image vectors). The test face image belongs to class 1. The results show that the dominant coefficients group together in the first class (index number between 1 to 10), which coincides with the intuition. As the test face image belongs to the first class, it should be linearly represented by atoms from the first class. We do observe there are a few small coefficients located at the sixth, seventh and ninth classes which are caused by the noises. These small turbulences, however, does not affect the classification result.

In this paper, we apply the group lasso formulation [19] to model the group sparsity in face recognition problem where the solution space can be considered as a union of K T -dimensional subspaces. To enforce a solution with group sparsity structure, group lasso imposes the ℓ_1 norm on the ℓ_2 norms of the coefficient vectors from these K subspaces. By minimizing the ℓ_1 norm of the vector of ℓ_2 norms subject to certain constraints, a group sparsity solution residing in one of the subspaces could be found. Specifically, Eq. (5) can be rewritten as

$$\theta_{\text{opt}} = \arg \min \sum_{i=1}^K \|\theta_{\Lambda_i}\|_2 \quad \text{subject to } \|\Phi y - \Phi \mathcal{D} \theta\|_2 < \epsilon, \quad (14)$$

where Λ_i denotes θ restricted to the index set Λ_i for group i . The optimisation problem finds the solution which minimises the number of groups containing non-zeros. It can be solved by SPGL1 [20], [21] which is widely used solver for sparse representation problem with group sparsity structure. The group sparsity model exploits more prior knowledge of the sparse representation classification problem. Therefore, it is expected to provide more accurate face recognition results. We term the opti-SRC exploring group sparsity structure as *opti-GSRC*. As the evaluation with multiple datasets in Section 4, the recognition accuracy of opti-GSRC is up to 5 percent better than that of opti-SRC approach. However opti-GSRC is too computationally intensive for implementation on smartphones. As our evaluation in Section 4.4, it consumes 25 times more computation time for each classification compared with opti-SRC on the same platform, making it infeasible for in-situ implementation on smartphones. Fortunately, thanks to the availability of high speed networks (Wifi/4G) and cloud computing resources, it is possible to provide solutions for realtime implementation. For example, the smartphones detect and offload the face images to the cloud via high speed networks. The cloud performs the recognition tasks and sends the recognition results back to the smartphone.

The total computation time will be significantly saved by the power of cloud computing resources.

Note that we use for simplicity the standard group sparsity model to demonstrate the benefits of combining it with our optimised projection matrix. However, our optimised projection matrix is transparent to the choice of group sparsity model. Therefore, it is not difficult to apply our optimised projection matrix to other more complicated models, such as the locality-constrained group sparsity model proposed by Chao et al. [22], to obtain better recognition accuracy.

3.3 Challenge III: Residual Calculation

The SRC in [4] was designed to solve the face recognition problem assuming abundant computing power is available. However, if the algorithm is to be implemented on embedded systems (e.g., smartphones), we will need to optimise the computations in SRC to improve its efficiency. Our evaluation of SRC on various smartphone platforms shows that the computation of residuals in Equation (6) is a severe bottleneck, accounting for 85-90 percent of SRC’s running time. The computation of residual r_i in Equation (6) is of the order $\mathcal{O}(p + pT)$ where in particular p is the number of pixels in the training image, which is generally a large number. Instead of computing the exact r_i using (6), we propose to compute an approximate r_i by making use of the *Johnson-Lindenstrauss (J-L) Lemma* [23]. Let Φ be the $m \times p$ (with $m \ll p$) projection matrix used in the ℓ_1 optimisation, which is either Gaussian or Bernoulli distributed, the J-L Lemma provides an approximation of the residual by:

$$\|y - \mathcal{D}\theta_{\text{opt}}^{(i)}\|_2 \approx \|\Phi y - \Phi \mathcal{D}\theta_{\text{opt}}^{(i)}\|_2. \quad (15)$$

Since both Φy and $\Phi \mathcal{D}$ are used in the ℓ_1 optimisation problem (5), they have already been computed. Also, both of them have m rows, so the computation on the right-hand side of (15) has a complexity of $\mathcal{O}(m + mT)$, which is much lower because $m \ll p$. We will refer to the computation of residuals using the right-hand side of (15) as *compressed residuals*.

Note that the validity of the J-L Lemma requires that the projection matrix is either Gaussian or Bernoulli distributed. This is another reason why we have restricted the optimisation problem in Section 3.1.1 to search for a Gaussian or Bernoulli projection matrix.

The classification is then carried out using the compressed residuals after they have been computed. Our experiments show that the recognition accuracy of using compressed residuals is almost the same as that of using full residual calculations when sufficient number of projections are used.

4 PERFORMANCE EVALUATION ON PUBLICLY AVAILABLE DATABASES

4.1 Goals, Metrics, and Methodology

The aim of this section is to evaluate the classification accuracy of opti-GSRC and opti-SRC then compare them with rand-SRC as well as Eigenface recognition algorithm in OpenCV. The comparison makes use of one publicly available database: the Extended Yale database B (Ext-YaleB) [24] and one private face database. Yale database is a popular choice [4], [25] to evaluate the recognition accuracy of face

recognition algorithms and our private database is collected by off-the-shelf smartphones because we aim at face recognition system on smartphones in this paper.

For fair comparison between different algorithms, a common parameter m is used as the number of projections in the three SRC-based algorithms as well as the number of features in Eigenface algorithm in OpenCV. For the SRC-based algorithms, projections play the same role as features.

We compare our strategy against two existing methods for optimising projection matrix. We also investigate the tolerance of the optimised projection matrix to training set changes. The evaluation in this section is carried out on a computer. Experimental results on smartphones will be presented in the next section.

In this paper, we use the percentage of correct recognition as the performance metric (same as [4]) for the recognition evaluation, which is simply the number of true recognitions over the total number of tests. We express the recognition accuracy of different methods under different number of projections/features and plot the recognition accuracy curves of different face recognition methods.

4.2 Extended Yale Database B

The Extended Yale Database B contains $K = 38$ subjects under nine poses and 64 illumination conditions. We choose the first $T = 32$ images from each subject as the training images and the following 10 images forms the test set. We remark that our choice of training images from each subject is different from the the work applying original SRC [4] and the work with locality and group sparsity [22], where training images were chosen randomly from each subject. Because the images are sorted by the time that they are captured, our choice is more consistent with the face recognition practice on smartphones where training images are added well before the test images. We have reproduced the results in [4] and [22] with same settings and observed that their choice tends to produce a better recognition accuracy because it distributes more similar face images into the training and test sets (empirically, images of one subject taken on temporal contiguity share more information. They are more likely to be classified into the same class).

The evaluation in [26] shows that face recognition can tolerate some degree of resolution reduction. We scale the images from the original size of 192×168 to 96×84 (i.e., 4 times reduction), which does not affect the recognition accuracy. (We also try to reduce the resolution by 16 times, but the accuracy decreases significantly). This setting gives an dictionary \mathcal{D} of size $8,064 \times 1,216$. We choose to reduce the image resolution because of two considerations: first, the high resolution of the image will introduce intensive computation especially for solving the sparse representation; second, the resolution of the embedded cameras varies among the smartphones, a face recognition system using low resolution images can be readily used on different smartphones.

4.2.1 Coherence Analysis

We compare the row coherence of the sensing matrix $\Phi \mathcal{D}$ where Φ comes from minimising (13) or is randomly generated. We vary the number of projections from 10 to 100. For

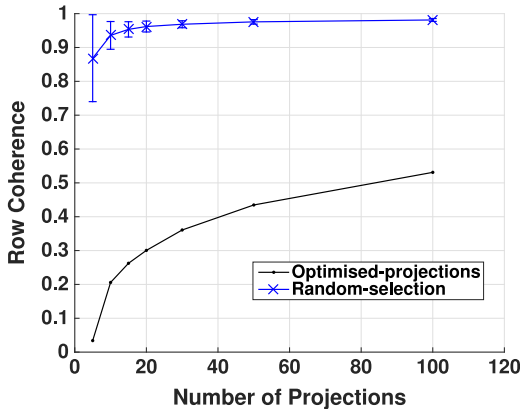


Fig. 2. YaleB: Comparison of row coherence of the sensing matrices.

each number of projection, we find an optimised Φ using tabu search algorithm with the search domain Ω consisting of Gaussian distributed vectors and compute the row coherence of the sensing matrix using the optimised Φ . We also generate 30 random Gaussian matrices, and calculate the mean and standard deviation of their row coherence for the corresponding sensing matrices. The results are plotted in Fig. 2. The x -axis is the number of projections from 10 to 100 and the y -axis is the row coherence of the sensing matrix. Fig. 2 shows that tabu search provides sensing matrices with lower row coherence. It can reduce the row coherence of the sensing matrix by up to four times when the number of projections is 10. Moreover, our algorithm can provide more consistent row coherence while the random approach shows a large variance (error bars in the figure). We have also calculated the mutual coherence of the sensing matrix with the optimised projection matrix. The result in Fig. 3 shows that our optimisation approach will not increase the mutual coherence (actually it reduces the mutual coherence a bit). Note that for face recognition, \mathcal{D} is a dictionary consisting of face image vectors as its columns. The faces in the same class can be very similar to each other. So the mutual coherence can be high (over 0.99 in random selection approach). The large row coherence in Fig. 2 arises for a similar reason: images have high spatial coherence. The adjacent pixels tend to have similar pixel values which lead to high row coherence.

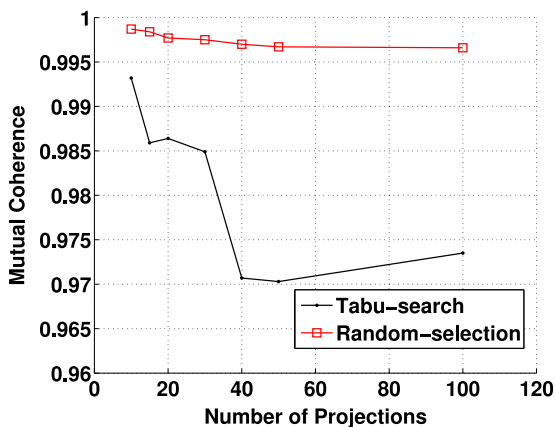


Fig. 3. YaleB: Comparison of mutual coherence of the sensing matrices.

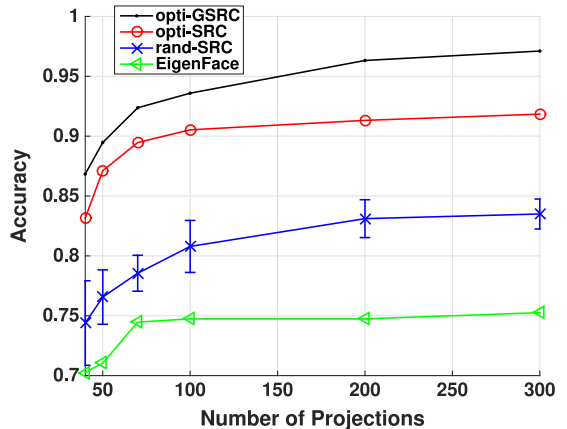


Fig. 4. YaleB: Comparison of face recognition accuracy of different methods.

4.2.2 Recognition Accuracy

To show that our strategy gives good face recognition accuracy, we compare opti-SRC (with optimised projection matrix tabu search) opti-GSRC (with optimised projection matrix and exploiting group sparsity structure), rand-SRC and Eigenface algorithm from OpenCV. The results from the rand-SRC are represented by the average and standard deviation over 30 independent trials where a different random projection matrix is used in each trial. The solver for opti-SRC and rand-SRC is the ℓ_1 -Homotopy implemented in [27] as it is computationally efficient while the solver for opti-GSRC is SPGL1 implemented in [21]. The reconstruction tolerances of the optimisation problems (e.g., ϵ in Equations (5) and (14)) for the three SRC based approaches are all set to be 0.001 considering the trade-off between the recognition accuracy and computational efficiency. In Fig. 4, the x -axis represents the number of features/projections ranging from 30 to 300 and the y -axis is the recognition accuracy. The results show that opti-SRC with projection matrix from tabu search algorithm improves the recognition accuracy by up to 12 percent compared with other existing algorithm, and opti-GSRC using group sparsity model performs up to 5 percent better than opti-SRC. The results give strong evidence that our strategy of minimising the row coherence of the sensing matrix delivers a face recognition algorithm with consistently high accuracy. In other words, for a given accuracy, opti-SRC needs less projections than rand-SRC; this translates to less computation resource requirement for opti-SRC. Moreover, exploiting the group sparsity structure can further improve the accuracy of SRC based classification.

4.2.3 Stability to Training Set Evolvment

For many face recognition applications, new classes are constantly added to the training set. This means the dictionary \mathcal{D} evolves constantly over time. According to our evaluation, the projection matrix optimisation is not suitable to be implemented on smartphones because it is too computationally intensive. Therefore, it is not convenient to recalculate the optimised projection matrix frequently. Since the projection matrix Φ is optimised for a given dictionary \mathcal{D} (see Equation (13)), we investigate whether it is necessary to re-optimize Φ if new classes are added to the dictionary. For

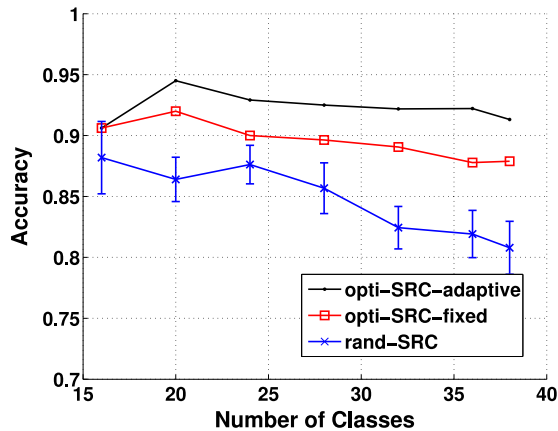


Fig. 5. YaleB: Training set evolution.

this evaluation, we do not consider opti-GSRC. Since as discussed earlier, opti-GSRC is designed for a cloud-based approach. The high performance computing resources can be used to re-optimize the projection matrix.

We begin with an initial training set containing 16 classes. We use tabu search to obtain an optimized projection matrix based on the initial dictionary. We then add training classes to the dictionary: four classes at a time until 36 classes and then two additional classes are added to reach 38 classes. We consider two different strategies. The first strategy, which we call opti-SRC-fixed, uses the optimized projection matrix obtained from the initial dictionary with 16 classes even when the dictionary evolves. The second strategy, called opti-SRC-adaptive, re-optimizes the projection matrix each time new classes are added to the dictionary. The number of projections is kept at 100.

Fig. 5 compares the classification accuracy of opti-SRC-adaptive, opti-SRC-fixed and rand-SRC. It is not surprising that opti-SRC-adaptive gives the best performance while rand-SRC is the worst. The interesting observation is that opti-SRC-fixed gives fairly stable performance in spite of the increase in number of classes. In fact, the performance of opti-SRC-fixed and opti-SRC-adaptive differs by at most 3 percent. Moreover, part of the degradation in the performance of opti-SRC-fixed is due to the use of the same number of projections as the number of classes increases. This is a very encouraging result which shows that re-optimization of projection matrix is only required occasionally. Note that there is performance improvement when the number of classes increases from 16 to 20. One explanation is the newly added test images bear some similarity to those in the dictionary.

4.2.4 Comparison with Other Projection Matrix Optimization Approaches

The idea of computing a projection matrix that is optimally designed for a certain signal class is not new. Elad [10], and Duarte-Carvajalino and Sapiro [9] recently proposed strategies for learning projection matrix directly from a dictionary to improve the performance of signal recovery in CS. The algorithm in [10] aims to find a projection matrix Φ to reduce the mutual coherence of the sensing matrix $A = \Phi D$ while that in [9] finds a Φ so that $A^T A$ is close to the identity matrix. There is another work on designing projection matrix with orthogonal rows. However, the “orthoprojectors” proposed

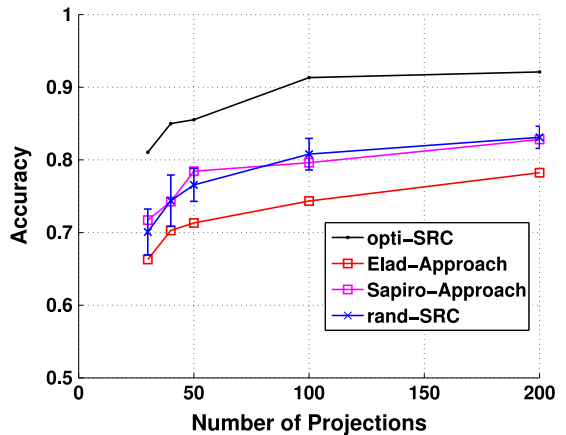


Fig. 6. YaleB: Comparison of different optimized matrices.

in [28] aim to make the rows of projection matrix orthogonal. They do not take the dictionary into account. Therefore, we only compare with the approaches in [10] and [9]. For this comparison, we use the optimized projection matrices from these two algorithms for face recognition; note that residual calculations use (6) because it is not known whether these optimized projection matrices satisfy the J-L Lemma. The comparison is against opti-SRC and rand-SRC. The results are shown in Fig. 6, which shows that the optimized projection matrix from [10] and [9] cannot improve the recognition accuracy. It is probably because the requirements on Φ for signal recovery and classification are different.

4.3 Private Mobile Dataset

To make the results more convincing, we collected another dataset using mobile phones to evaluate the recognition accuracy of our proposed methods. Ten subjects, seven males and three females, with different skin tones are recruited for the experiment. We take 40 photos of each subject in random poses, lighting conditions and expressions. We then use the Android OS *face-detection* API to automatically crop the face region of the images to a size of 92×56 and store them on the phones’ memory. The resolution of the face images is much lower than that of the embedded cameras on off-the-shell smartphones. We will refer to this database as *Private database*.

We use the first half of the images of each subject as training and the rest as test. Therefore, the training set \mathcal{D} is a matrix of size $5,152 \times 200$. Then we evaluate the performance of different face recognition methods as in Section 4.2.

Fig. 7 compares the recognition accuracy of opti-GSRC, opti-SRC, rand-SRC and Eigenface algorithm from OpenCV. Opti-SRC achieves consistently better recognition accuracy than the existing methods. The improvement is up to 7 percent compared with the rand-SRC. Opti-GSRC further improves the recognition accuracy by up to 5 percent compared with opti-SRC.

4.4 Computation Cost

Although opti-GSRC is more accurate than opti-SRC, it is too computationally complex for implementation on smartphones. We use matlab on a laptop to evaluate and compare the computation cost of opti-GSRC and opti-SRC. The laptop

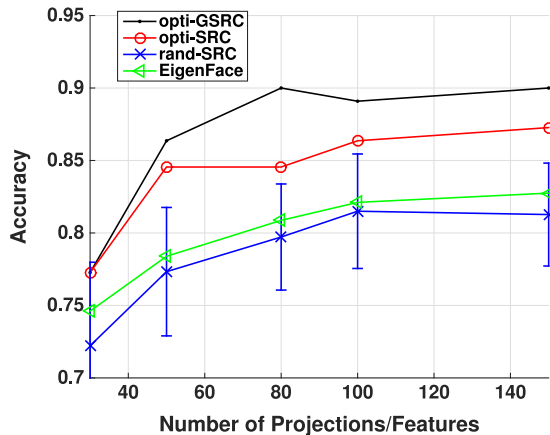


Fig. 7. Face recognition accuracy comparison under different number of projections/features with private database.

we use has 2.5 GHz Core i7 and 16 GB memory and the operating system is Mac OS X Yosemite. The evaluation is based on YaleB database. The number of projections are set as 100 and the trainingset contains 1,216 face image vectors. We run 760 classification trials and compute the average computation time for each classification task. The average computation time of opti-GSRC is about 25 times more than that of the opti-SRC. Considering the fact that the computation time of opti-SRC is about 300-400 ms on off-the-shelf smartphones according to our implementation in Section 5.4, opti-GSRC is not applicable for implementation on smartphones. However, opti-GSRC can be used in cloud-based approach which shifts the computation burden to the high performance local server or cloud through high speed network. In this work, we only implement in-situ approach, namely, opti-SRC face recognition system on smartphones as demonstration. While the real-time implementation of opti-GSRC is possible when the cloud computing resources are available and they are becoming pervasive.

4.5 Conclusions of This Section

From the results of this section, we can conclude that 1) opti-SRC has significantly better recognition accuracy compared with the state-of-the-art face recognition algorithms; 2) exploiting group sparse structure (i.e., opti-GSRC) further improves the recognition accuracy 3) the optimised projection matrix can tolerate the evolution of the training set which makes it more convenient to use on smartphones.



Fig. 8. Android smartphones used in the experiment; from left to right: Samsung Galaxy S3, Google Nexus 4, and HTC OneX.

TABLE 1
Smartphones Specifications Relating to the Face Recognition Application

	Galaxy S3	Nexus 4	HTC OneX
CPU	Quad 1.4 GHz	Quad 1.5 GHz	Quad 1.5 GHz
Storage	16 GB	8 GB	16 GB
RAM	1 GB	2 GB	1 GB
Screen	4.7 inches	4.7 inches	4.7 inches
Battery	2,100 mAh	2,100 mAh	1,800 mAh
Android OS	v4.0.4	v4.2.2	v4.1.1

5 EXPERIMENT ON SMARTPHONES

5.1 Experiment Description

The aim of this section is to evaluate the performance of opti-SRC, and compare it with rand-SRC as well as the OpenCV algorithms on the smartphones.

We implement opti-SRC on the Android platform. The ℓ_1 minimisation uses the ℓ_1 -homotopy solver [27] because it is much more efficient than linear programming: complexity of $\mathcal{O}(s^3 + smn)$ (where s is the sparsity of the solution, with $s \ll n$) versus $\mathcal{O}(n^3)$.

In this section, we present experimental results on computation time and resource consumption (Section 5.4), and recognition accuracy (Section 5.5) of opti-SRC and OpenCV algorithms on the chosen smartphone platforms. We have also designed a group face unlocking app (Section 5.6) with the opti-SRC and compared it against the Android built-in face unlocking app on under various different conditions to test its robustness.

5.2 Platforms

In order to demonstrate the feasibility of opti-SRC, we implement it on three different popular smartphones: Samsung Galaxy S3, Google Nexus 4, and HTC OneX (shown in Fig. 8). The specifications of these smartphones relating to the application are shown in Table 1.

5.3 Preliminary Experiments

5.3.1 Number of Projections

The training set used for the experiment is built from our private database as the same group of subjects are recruited for the experiment on smartphones. From the results shown in Fig. 7, the recognition accuracy becomes level when the number of projections is over 100. We therefore choose to use 100 projections for further experiments on the smartphones. Then we randomly pick 20 images from each subject to form the training set. The training set will be a matrix of size $5,152 \times 200$. The size of the projection matrix is $100 \times 5,152$.

5.3.2 Compressed Residual

According to challenge III in Section 3.3, the computation of residuals using the full-size image vectors as in Equation (6) is computationally intensive (complexity of $\mathcal{O}(p + pT)$ where $p = 5,152$ is the size of the image) and we propose to approximate the residual calculations by Equation (15) (complexity of $\mathcal{O}(m + mT)$ where $m = 100$ is the number of projections). We measure the computation time (per test image classified) of both methods of residual calculations on the smartphones

TABLE 2
Time for Original/Compressed Residual Calculations
Per Face Classification

	Galaxy S3	Nexus 4	HTC OneX
Compressed Residual	42 ms	33 ms	43 ms
Original Residual	1,923 ms	1,926 ms	2,105 ms

and show the results in Table 2. We see that, by using compressed residuals, the computation time is reduced by a factor of 45-50 times. This is consistent with the complexity expressions which predict a $\frac{p}{m} \approx 52$ times reduction.

Table 3 shows that the computation time of opti-SRC without the residual calculation component (i.e., face detection and ℓ_1 minimisation components only) is 200-350 ms for the three platforms. If the original method of residual calculations is used, then it will take up 85-90 percent of the computation time, which is a severe bottleneck. By using compressed residuals, the total classification time per image is reduced by about 6 times.

5.4 Resource Consumption on Smartphones

Table 3 shows the resource consumption (running time and energy consumption) of opti-SRC on different smartphones. The running time is obtained from the console of Eclipse development environment and the energy consumption is evaluated using the PowerTutor (powertutor.org) app on Android.

In terms of processing speed, Table 3 shows that the SRC component (ℓ_1 minimisation + compressed residual calculation) is the dominant factor of time consumption which is about 3 times of the face detection component. Therefore, the use of compressed residual can drastically reduce the total processing time. Another noticeable result is that most of the energy is consumed by the display which is inevitable for visualisation of the applications. The CPU only takes 10-15 percent of the total energy consumption. Considering the fact that the smartphones mostly target to a lifespan of one day usage which indicates an average energy consumption of 1.25 kJ per hour, with only 1 percent of the hourly budget, our proposed system can perform face authentication about 63 times. It is significantly more than the number of authentications required in common usage per hour. Therefore the proposed face recognition system only has minimal impact on the lifespan of the smartphones.

To provide a benchmark for comparison, we also implement the OpenCV based face recognition applications on the smartphones. Because the original OpenCV library was coded in C++, we use the Java Native Interface

TABLE 4
Speed of Face Detection and Recognition with Feature
Based Recognition Methods in OpenCV

	Galaxy S3	Nexus 4	HTC OneX
EigenFace (total)	177 ms	143 ms	161 ms
FisherFace (total)	221 ms	159 ms	186 ms
LBPFace (total)	268 ms	220 ms	257 ms

(JavaCV-v0.6) which is based on OpenCV-v2.4.6. The average processing time for detecting and recognising one face is shown in Table 4. ℓ_1 minimisation is known to be computational intensive. However, with the power of dimensionality reduction and compressed residual, opti-SRC achieves the same order of speed as the algorithms in OpenCV, which are known to be fast. It means that our algorithm is quite efficient for face recognition. Note that our ℓ_1 -homotopy implementation uses Java which is known to be less efficient than C++. The speed of classification can be further improved with C++ implementation but it is out of the scope of this paper.

5.5 Recognition Accuracy on Smartphones

In this section, we evaluate the recognition accuracy of opti-SRC, rand-SRC and OpenCV algorithms using a *new* test set. This new test set has 400 photos, 40 from each subject. (Note the 400 photos from this new test set are different from those in the Private database. These photos are taken at two different locations with lighting conditions: bright (above 100 lux) and dark (below 50 lux) respectively. The luminance can be measured by using the light sensor embedded on the smartphones. Twenty photos are taken for each lighting condition per subject. These 20 photos come from 20 different (*expression, pose*) pairs where expression is neutral, happy, sad or neutral with glasses; and pose is front, up, down, left or right as shown in Fig. 9. All the photos are manually labelled by their lighting condition, expression and pose. This new test set presents a rich variety of situations to test the face recognition algorithms. Also, there is mobility when the experiments were conducted. During the experiments the subjects walk around in the lab and take pictures freely.

For this evaluation, the training set is Private-Training used in Section 5.3.1. Note that there are only 20 images per subject in the training set, but there are 40 images per subject in the test set. Therefore, the test set presents a much richer variations than the training set. The number of projections/features remains as 100.

Table 5 shows the recognition accuracy of opti-SRC using different subsets of the test set. For example, the accuracy

TABLE 3
Resource Consumption of opti-SRC on Different Smartphones

	Galaxy S3	Nexus 4	HTC OneX
Face Detection (ms)	100	70	97
ℓ_1 (ms)	240	136	247
Residual (ms)	42	33	43
Total (ms)	382	239	387
Display Energy (mJ)	1,500	1,040	1,300
CPU Energy (mJ)	175	139	246



Fig. 9. Face samples of different poses.

TABLE 5
The Recognition Accuracy of opti-SRC Under
Different Subsets of the Test Set

Neutral	Happy	Sad	Glasses	Dark	Bright	Total
92%	89%	89%	90%	85%	95%	90%

under *Happy* means we take the 100 photos (= 10 subjects \times 2 lighting conditions \times 5 poses) in the test set with the label *Happy* and test whether opti-SRC can recognise the subject in the photo correctly. For the accuracies under *Dark* and *Bright*, 200 test photos are used in each subset. From the results we can see that, opti-SRC is more sensitive to the change of expressions than the small occlusion (wearing glasses). This is reasonable because the change of expressions is related to much larger area of face than the small occlusion. Another observation is that the well-lighted environment can improve the recognition accuracy (about 10 percent improvement).

For the methods in OpenCV and rand-SRC, the classification accuracy of EigenFace, FisherFace, LBPFace and rand-SRC is only 81-82 percent which is significantly lower than those of opti-SRC (90 percent).

5.6 Group Face Unlocking

Face unlocking is a built-in security feature in the recent Android OS release. It uses face, rather than password, for unlocking the smartphone. Here we design a new face unlocking app for smartphones based on opti-SRC. Different from the built-in face unlocking function, our app considers the situation that the smartphones can be shared by a small group of users. We call our app group face unlocking.

5.6.1 Resistance to Intruders

Because a security system should have “zero” tolerance to intruders, we include a criterion to validate the recognition results so that intruders are rejected. In our experiments, we choose three subjects as the “true” persons in the group unlocking apps and consider the other seven subjects as intruders. There are $K = 3$ training classes where each class corresponds to a subject in the group.

We apply opti-SRC as before and compute the compressed residual for each class. Let r_j^c be the compressed residual of class j . If a person, say person 1, in the group is presented for recognition, we expect the residual r_1^c is much lower than r_2^c and r_3^c . We can define the confidence of recognition as

$$\text{confidence} = \left(\frac{1}{K} \sum_{j=1}^K r_j^c - \min_{j=1, \dots, K} r_j^c \right) / \frac{1}{K} \sum_{j=1}^K r_j^c. \quad (16)$$

The confidence is in the range $[0, 1]$ and should be close to 1 if a subject in the group is presented for recognition. On the other hand, if it is an intruder, the confidence should be close to 0.

The unlocking algorithm will first use opti-SRC to perform a recognition. After that, a confidence level is calculated and a person will be considered to be an intruder if the confidence level is lower than a pre-defined threshold. The threshold can be tuned according to the security level of the application. There is a trade-off between the security level and the recognition of true users: a higher threshold

makes the system more resistant to intruders but true users are also more likely to be rejected. We use a threshold of 0.16 in the following experiment so that no false positives (an intruder is recognised as user) are shown with the test of the seven intruders.

5.6.2 Robustness to Variations

In this part, we evaluate the recognition accuracy of our group unlocking app under many different conditions to test its robustness. This is important because a good face unlocking app needs to recognise the right user under different facial expressions and environment variations. We compare our group unlocking based on opti-SRC against the built-in face unlocking app under different variations, including lighting, occlusion, expression and poses. During the evaluation, we carefully control the factors so that only one type of variation is present at a time. Unlike most of the Android OS, the built-in face unlocking is *closed source*, therefore we cannot use exactly the same face image to compare these two apps. Instead, we have to capture two images for each test and make these two images as similar as possible. We again use 20 face images for each class as the training set for opti-SRC and we train the face unlocking with the default conditions mentioned below.

Lighting Variation: The capability of recognition under different lighting conditions is the key for the face unlocking app because the users will use the app under different environment. The robustness to the lighting changes is evaluated in five locations with different illumination intensity which are around 5, 20, 100, 200 and 800 lux respectively during the experiments. The default lighting illuminance under which the face unlocking is trained is around 100 lux.

Occlusion: We test the face unlocking apps for the different ratio of occlusions. The users wear either normal prescription glasses, headphones or large black sunglasses during the test. The faces in the training set do not contain any occlusions.

Expression Variation: Expression is another common variation when the face images are taken. The expressions of neutral, happy and sad are used during the test. The default expression is neutral.

Pose Variation: Pose changes should be addressed because direction of the faces to the camera cannot be always strictly frontal. To compare the robustness of the face unlocking apps to the variation of the pose, the users gradually change their pose of face from front to two directions (up, down) until the false recognition appears (the angles to the gravity direction can be measured with built-in sensor in smartphones). The front pose is used as default pose.

The evaluation results are shown in Table 6. The pose change limit is demonstrated by angles to the gravity direction. It is the largest angles that the face unlocking apps can recognise the users correctly. (We are not able to do experiments for the left and right poses because they cannot be directly measured. However, our experience shows that opti-SRC can tolerate larger pose angles in right and left directions.) The check/cross markers (\checkmark/\times) represent the face unlocking app recognises the face of users correctly/wrongly under the corresponding variance (true positive/false negative).

TABLE 6
The Comparison of Our Approach and Built-in Face
Unlocking Apps on Different Variance Conditions

lighting	5 lux	20 lux	100 lux	200 lux	800 lux
built-in	x	✓	✓	✓	x
opti-SRC	✓	✓	✓	✓	x
occlusions	none	headphone	glasses	sunglasses	
built-in	✓	✓	✓	x	
opti-SRC	✓	✓	✓	x	
expressions	neutral	happy	sad		
built-in	✓	✓	✓		
opti-SRC	✓	✓	✓		
poses	up	down			
built-in	17°	40°			
opti-SRC	20°	36°			

The results shown in Table 6 demonstrate that our group unlocking app is more robust than the built-in app for lighting and pose changes and the two face unlocking apps achieve the same results on the expression and occlusion changes.

6 RELATED WORK

In sparse representation, random matrices are often used to reduce the dimensionality of the problem while preserving the accuracy of the applications. They have been applied to speed up background subtraction on embedded system [29] and cross-correlation computation in sensor networks [30]. In [5], SRC is used for acoustic classification and a *column reduction* procedure is proposed to reduce the dimension of ℓ_1 minimisation. Note that column reduction in [5] is complementary to the techniques of projection matrix optimisation and compressed residuals proposed in this paper; all three can be applied to improve the performance of SRC. The most recent work by Xu et al. [31] proposed a sensor-assisted face recognition system for smart glasses based on SRC which exploited face images from multi-view angles to improve the recognition accuracy. Other application of random projection matrix is to enable efficient moisture data collection in sensor networks [32], privacy preservation of voice data [33], device free activities recognition [34] and efficient GPS signals acquisition [35]

In this paper, we apply the standard group sparsity formulation, i.e., group lasso norm, which was proposed by Yuan et al. [19]. The group lasso norm was applied to deal with visual classification problem in [36]. To further improve the recognition accuracy, Chao et al. [22] jointly considered the data locality and group sparsity to formulate the sparse representation classifier. As our major contribution in this paper is the optimised projection matrix which is transparent to the usage of different group sparsity models, we apply the standard group lasso norm to formulate the face recognition problem.

7 CONCLUSION

In this paper, we address the challenges of performing face recognition *accurately* and *efficiently* on smartphones by designing new face recognition algorithms called opti-SRC and opti-GSRC. A key idea behind opti-SRC is a novel strategy to optimise the projection matrix so that the resulting sensing matrix has both low row and mutual coherence. Opti-GSRC takes a step forward by exploiting the group

sparsity structure in the sparse representation classification problem to further improve the recognition accuracy. We show that opti-SRC is 8-17 percent more accurate than the state-of-the-art face recognition algorithms and opti-GSRC further improves the recognition accuracy by up to 5 percent. The use of optimised projection matrix means that opti-SRC can achieve the same recognition accuracy using a lower dimension projection matrix. This translates to a lower computation requirement on smartphones. In addition, we propose the use of compressed residuals in order to significantly reduce the computational time of opti-SRC on resource constrained smartphones. We compare the running time of opti-SRC against OpenCV algorithms on three smartphone platforms and find they have comparable running time.

REFERENCES

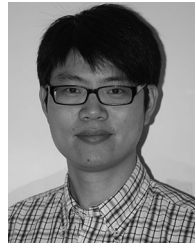
- [1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Pattern Recognition*, 1991, pp. 586–591.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces versus fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [3] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
- [4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [5] B. Wei, M. Yang, Y. Shen, R. Rana, C. T. Chou, and H. Wen, "Real-time classification via sparse representation in acoustic sensor networks," in *Proc. 11th ACM Conf. Embedded Netw. Sensor Syst.*, 2013, Art. no. 21.
- [6] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [7] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [8] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [9] J. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.
- [10] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec. 2007.
- [11] A. Krause and C. Guestrin, "Optimizing sensing: From water to the web," *Computer*, vol. 42, pp. 38–45, 2009.
- [12] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, 2007, Art. no. 969.
- [13] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [14] J. A. Tropp, "The sparsity gap: Uncertainty principles proportional to dimension," in *Proc. 44th IEEE Conf. Inf. Sci. Syst.*, 2010, pp. 1–6.
- [15] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 358–374, Apr. 2010.
- [16] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [17] F. Glover and M. Laguna, *Tabu Search*. Norwell, MA, USA: Kluwer, 1997.
- [18] G. Paul, "An efficient implementation of the robust tabu search heuristic for sparse quadratic assignment problems," *Eur. J. Oper. Res.*, vol. 209, no. 3, pp. 215–218, 2011.

- [19] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statistical Soc.: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [20] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Scientific Comput.*, vol. 31, no. 2, pp. 890–912, 2008.
- [21] E. van den Berg and M. P. Friedlander, "Sparse optimization with least-squares constraints," *SIAM J. Optimization*, vol. 21, no. 4, pp. 1201–1229, 2011.
- [22] Y.-W. Chao, Y.-R. Yeh, Y.-W. Chen, Y.-J. Lee, and Y.-C. F. Wang, "Locality-constrained group sparse representation for robust face recognition," in *Proc. 18th IEEE Int. Conf. Image Process.*, 2011, pp. 761–764.
- [23] W. B. Johnson and J. Lindenstrauss, "Extensions of lipschitz mappings into a hilbert space," in *Proc. Conf. Modern Anal. Probability*, 1984, pp. 189–206.
- [24] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
- [25] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, 2007, pp. 1–7.
- [26] B. Boom, G. Beumer, L. J. Spreeuwiers, and R. N. Veldhuis, "The effect of image resolution on the performance of a face recognition system," in *Proc. 9th Int. Conf. Control Autom. Robot. Vision*, 2006, pp. 1–6.
- [27] M. Salman Asif and J. Romberg, "Dynamic updating for l1 minimization," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 421–434, Apr. 2010.
- [28] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "The compressive matched filter," Rice Univ., Houston, USA, Tech. Rep. TREE 0610, 2006.
- [29] Y. Shen, W. Hu, J. Liu, M. Yang, B. Wei, and C. T. Chou, "Efficient background subtraction for real-time tracking in embedded camera networks," in *Proc. 10th ACM Conf. Embedded Netw. Sensor Syst.*, 2012, pp. 295–308.
- [30] P. Misra, W. Hu, M. Yang, and S. Jha, "Efficient cross-correlation via sparse representation in sensor networks," in *Proc. ACM/IEEE 11th Int. Conf. Inf. Process. Sensor Netw.*, 2012, pp. 13–24.
- [31] W. Xu, Y. Shen, N. Bergmann, and W. Hu, "Sensor-assisted face recognition system on smart glass via multi-view sparse representation classification," in *Proc. 15th ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, 2016, pp. 1–12.
- [32] X. Wu and M. Liu, "In-situ soil moisture sensing: Measurement scheduling and estimation using compressive sensing," in *Proc. 11th Int. Conf. Inf. Process. Sensor Netw.*, 2012, pp. 1–12.
- [33] B. Liu, Y. Jiang, F. Sha, and R. Govindan, "Cloud-enabled privacy-preserving collaborative learning for mobile sensing," in *Proc. 10th ACM Conf. Embedded Netw. Sensor Syst.*, 2012, pp. 57–70.
- [34] B. Wei, W. Hu, M. Yang, and C. T. Chou, "Radio-based device-free activity recognition with radio frequency interference," in *Proc. 14th Int. Conf. Inf. Process. Sensor Netw.*, 2015, pp. 154–165.
- [35] P. K. Misra, et al., "Energy efficient GPS acquisition with sparse-GPS," in *Proc. 13th Int. Symp. Inf. Process. Sensor Netw.*, 2014, pp. 155–166.
- [36] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multi-task joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.



Yiran Shen received the BE degree in communication engineering from Shandong University, China, and the PhD degree in computer science and engineering from the University of New South Wales. He is an associate professor in the College of Computer Science and Technology, Harbin Engineering University, China. Before joining Harbin Engineering University, he was a SMART scholar fellow of the Future Urban Mobility IRG, Singapore-MIT Alliance for Research and Technology. He publishes regularly at top-tier

conferences and journals like ACM SenSys, IEEE/ACM IPSN, and the *IEEE Transactions on Mobile Computing*. His current research interests are wearable/ mobile computing, wireless sensor networks, and applications of compressive sensing. He is a member of the IEEE.



Mingrui Yang (M'11) received the PhD degree in mathematics from the University of South Carolina, in 2011. He is currently a senior research associate at Case Western Reserve University. He is interested in interdisciplinary researches in compressive sensing, sparse approximation, signal/image processing, greedy algorithms and nonlinear approximation, and their applications. He has published in high quality journals and conferences in mathematics and computer science, such as the *Advances in Computational Mathematics*, ACM/IEEE IPSN, and ACM SenSys. He is a member of the IEEE.



Bo Wei received the bachelor's degree from the College of Information Science and Engineering and the master's degree from Northeastern University, Shenyang, China, both in engineering, in 2009 and 2011, respectively. He received the PhD degree from the University of New South Wales, Australia. He is currently a postdoctoral research associate in the Department of Computer Science, University of Oxford. He was also a visiting student in the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia, and Swedish Institute of Computer Science (SICS). He is a member of the IEEE.



Chun Tung Chou received the BA degree in engineering science from the University of Oxford, United Kingdom, and the PhD degree in control engineering from the University of Cambridge, United Kingdom. He is an associate professor in the School of Computer Science and Engineering, The University of New South Wales, Australia. His current research interests are wireless sensor networks, compressive sensing, and nano-communication. He is a member of the IEEE.



Wen Hu is a senior lecturer in the School of Computer Science and Engineering, the University of New South Wales (UNSW). Much of his research career has focused on the novel applications, low-power communications, security and compressive sensing in sensor network systems, and Internet of Things (IoT). He publishes regularly in the top rated sensor network and mobile computing venues such as ACM/IEEE IPSN, ACM SenSys, the ACM and IEEE Transactions, the *Proceedings of the IEEE*, and the *Ad-hoc Networks*. He serves on the organizing and program committees of networking conferences including ACM/IEEE IPSN, ACM SenSys, ACM MobiSys, and IEEE ICDCS. He is a senior member of the ACM and IEEE.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.