Privacy Preserving Face Retrieval in the Cloud for Mobile Users

Xin Jin[1], **Shiming Ge**[2,*], **Chenggen Song**[1]
[1]Department of Computer Science and Technology
Beijing Electronic Science and Technology Institute, Beijing, 100070, P.R. China
[2]Institute of Information Engineering
Chinese Academy of Sciences, Beijing, 100093, P.R. China
[*]Corresponding author: geshiming@iie.ac.cn

Abstract

Recently, cloud storage and processing have been widely adopted. Mobile users in one family or one team may automatically backup their photos to the same shared cloud storage space. The powerful face detector trained and provided by a 3rd party may be used to retrieve the photo collection which contains a specific group of persons from the cloud storage server. However, the privacy of the mobile users may be leaked to the cloud server providers. In the meanwhile, the copyright of the face detector should be protected. Thus, in this paper, we propose a protocol of privacy preserving face retrieval in the cloud for mobile users, which protects the user photos and the face detector simultaneously. The cloud server only provides the resources of storage and computing and can not learn anything of the user photos and the face detector. We test our protocol inside several families and classes. The experimental results reveal that our protocol can successfully retrieve the proper photos from the cloud server and protect the user photos and the face detector.

1 Introduction

In today's mobile Internet era, increasing mobile users backup their photos to the cloud storage servers. Some cloud servers provide face retrieval service, which allows one to retrieve photos that contain a specific person or a group of persons from all his/her photos in the own storage space of the cloud server.

Further more, as shown in Figure 1, people in one team or one family may share the same cloud storage space and upload their photos together. For example, people in one family take photos of each other using their mobile phones for a long time. In the traditional way, they may copy photos in each mobile phone using a cable and manage their photos manually. Nowadays, one can create a cloud storage space and share it to all the family members. All the photos shot by the family members can be stored to the shared cloud storage space automatically by the cloud Apps in their mobile phones. After that, each family member can browse photos using the cloud Apps in a friendly way. Besides, one can retrieve the photos contains one or multiply specific family members using the photo management module of the cloud Apps, as shown in Figure 1.

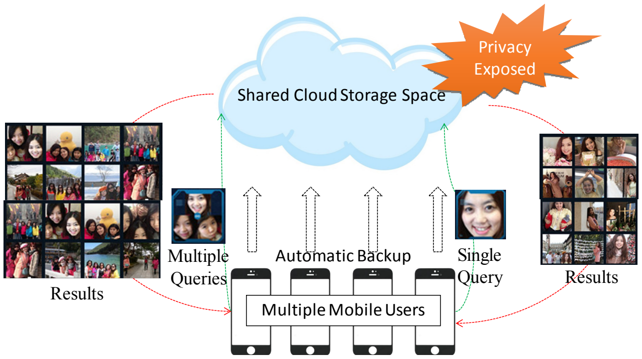

Figure 1: The typical scenario. Multiple users backup and share their photos in their mobile phones to a cloud storage. However, the privacy of photos is completely exposed to the cloud.

The above application models are typical scenarios in today's mobile Internet era. However, although the management of group photos is much more convenient, the privacy of the users' photos is completely exposed to the cloud server. The facial features of each member in one family, the relationship between members, and the school of their children, etc. can be learned from the family photos, which can threaten the personal and property security of the family.

In the meanwhile, the face detector used in the face retrieval task may be trained in a large scale of face images annotated by thousands of people. The copyright of the trained parameters of the face detector should also be preserved from the commercial provider's perspective.

Thus, in this paper, we propose a novel protocol to preserve the privacy of the cloud users' photos and the parameters of the commercial face detector simultaneously in such mobile cloud scenarios. The face retrieval problem can be decomposed into ***face detection***, ***face recognition*** and ***face label matching***. In the face detection stage, face regions are detected in users' photos with rectangles. In the recognition stage, each detected face is marked by a label of a member in a group. Then a label vector is generated according to the face recognition result for each photo so as to mark who is/are in each photo. The above is the off-line phase. In the on-line phase, a user queries a specific face of one person or faces of a group of person. Then, a label vector is generated for this query and compared to each label vector corresponding to each photo. Photos with the most similar label vector to the query label vector are selected as the retrieval result.

Related Work. The secure face detection method is proposed as Blind Vision [Avidan and Butman, 2006] for securely evaluating a Viola-Jones type face detector. After that Jin et al. accelerate secure face detector by introducing a random base image representation [Jin *et al.*, 2017]. A system called Secure Computation of Face Identification (SCiFI) [Osadchy *et al.*, 2010] is developed for secure face recognition. This system use two cryptography tools (homomorphic encryption and oblivious transfer) to implement a privacy preserving computation of the Hamming distance between two binary vectors. Recently, A lot of researchers have addressed the privacy preserving computer vision problems [Shashank *et al.*, 2008; Upmanyu *et al.*, 2009; Osadchy *et al.*, 2010; Sohn *et al.*, 2010; Fanti *et al.*, 2013; Chu *et al.*, 2014; Bost *et al.*, 2015; Jin *et al.*, 2016a; 2016b]. Most of them leverage the cryptography tools which are not efficient. The main mechanism in our protocols is to security compute the inner product. In 2009, Wong et.al. [Wong *et al.*, 2009] proposed a secure kNN (k-nearest neighbor) scheme on encrypted database, which developed a new asymmetric encryption that preserves inner product. We tailor the encryption scheme to meet our scenario, and construct our privacy preserving face retrieval application.

Our Approach. In this paper, we leverage a simple but efficient secure inner production protocol to protect the contents of user photos and the parameters of the face detector. The cloud server only provides the resources of storage and computing and can not learn anything of the user photos and the face detector. The face detection stage is protected by the secure face detection protocol using our secure inner production protocol. The face recognition stage is running locally in the users' mobile phone. The photos is encrypted and uploaded to the shared cloud storage space together with the corresponding label vector. The face label matching stage is running in the on-line face retrieval phase. The query label vector and the label vector in the cloud are compared using our secure inner production protocol.

2 PROBLEM FORMULATION

2.1 Overview

Our proposed methods are shown in Figure 2. In the face detection stage, face regions are detected in users' photos with a rectangle. In the recognition stage, each detected face are marked by a label of a member in a group. Then a label vector is generated according to the face recognition result for each photo so as to mark who is/are in each photo. The above is the off-line phase. In the on-line phase, a user queries a specific face of one person or faces of a group of person. Then, a label vector is generated for this query and compared to each label vector corresponding to each photo. Photos with the most similar label vector to the query label vector are selected as the retrieval result.

2.2 Security Model

We adopt the "honest-but-curious" model for the cloud server. It assumes that the cloud server would honestly follow the designated protocols and procedures to fulfill its service provider's role, while it may analyze the information stored and processed on the server in order to learn additional information about its customers.

The objective of our scheme is to preserve the 3^{rd} party and users' data privacy, which includes: 1). face detector privacy; 2). detected windows privacy; 3). photos content privacy; 4). label vectors privacy; 5). query privacy. While photos content privacy can be achieved by encryption-before-outsourcing schemes, this paper focuses on preserving the data privacy due to the face detection and matching, as follows:

Detection Privacy Besides the detection result, the cloud server should not deduce any face classifier information from the secure face detector, and face information from the secure detected windows.

Matching Privacy Besides the matching result, the cloud server should not deduce any face information from the secure label vectors and secure query.

3 Secure Face Retrieval

3.1 Secure Face Detection

Denote some finite field F that is large enough to represent all the intermediate results. Denote by X the image that Alice owns. A particular detection window within the image X will be denoted by $x \in F^L$ and x will be treated in vector form. Bob owns a strong classifier of the form

$$H(x) = \text{sign}(\sum_{n=0}^{N-1} h_n(x)), \tag{1}$$

where $h_n(x)$ is a threshold function of the form

$$h_n(x) = \begin{cases} \alpha_n & x^T y_n > \theta_n \\ \beta_n & \text{otherwise}, \end{cases} \tag{2}$$

and $y_n \in F^L$ is the hyperplane of the threshold function $h_n(x)$. The parameters $\alpha_n \in F, \beta_n \in F$ and $\theta_n \in F$ of $h_n(x)$ are determined during training; N is the number of weak classifiers used.

As in Figure 2, step (1), the 3^{rd} party $3P$ first generate the product key according to the users US's purchase as:

D-KeyGen(m): Given a security parameter m as the most length of the classifiers in the face detector, output the product key $SK(M_1, M_2, S)$, where $M_1, M_2 \in \mathcal{R}^{m \times m}$ are randomly invertible matrices and $S \in \{0,1\}^m$ is a randomly vector.

the next, $3P$ send this product key to US via secure channel.

The second, $3P$ encrypt his classifiers in the face detector and upload to the cloud server with the detect parameter $\{\alpha_i, \beta_i, \theta_i\}_{i=1,\cdots,n}$.

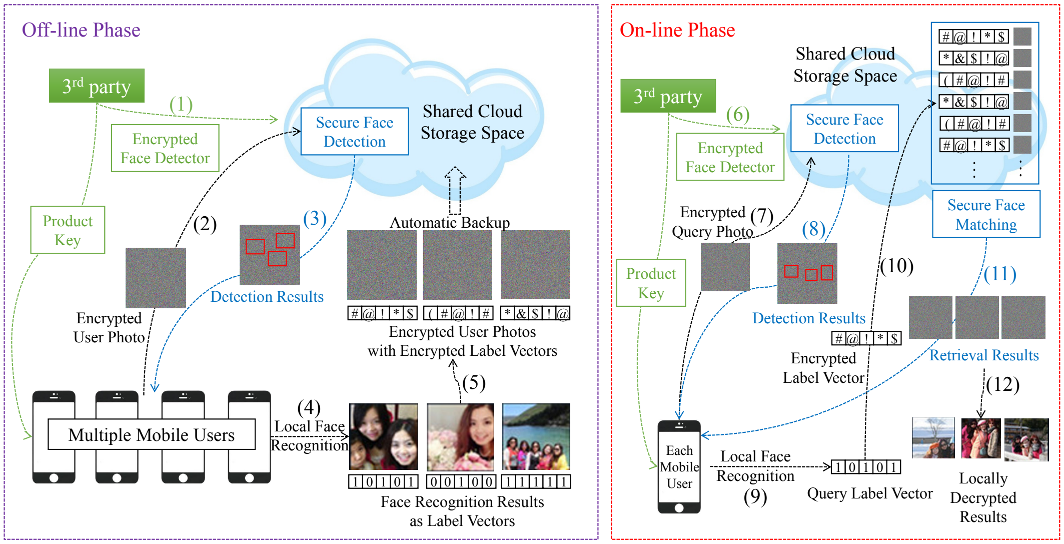

Figure 2: The overall system architecture. (1) The face detector is provided by a 3rd party with a product key. (2) A user encrypts photos with the same key as used in (1), before sending them to the cloud. (3) Our secure face detector protocol is running in the cloud with the encrypted face detector and photos. The detected face windows are sent back to the user. (4) The local face recognition algorithm is called to mark each photo with a label vector, which reveals who is/are in this photo using 1 for exist. (5) The photos in all the shared users are encrypted and uploaded to the cloud storage together with the encrypted label vectors. The off-line phase is end. In the on-line phase, a user want to query photos from the cloud storage with all the faces in the query photo. (6)(7)(8)(9) is the same as (1)(2)(3)(4). Then, the label vector of the query photo is computed. (10) The query label vector is encrypted and uploaded to the cloud and compared to all the label vectors in the cloud using our secure face matching protocol. (11) The corresponding encrypted corresponding photos with the top N matching label vectors are sent back to the user. (12) The user decrypts the matching photos and obtain the final retrieval results.

E-FD(SK,Y): To encrypt the classifiers $Y = y_1, \cdots, y_n$ in the face detector, $3P$ split each vector y_i into two vectors $\{y'_i, y''_i\}$ following the rule: for each $y_{i,j} \in y_i$, set $y'_{i,j} = y''_{i,j} = y_{i,j}$ if $s_j \in S$ is 1; otherwise $y'_{i,j} = \frac{1}{2}y_{i,j} - r$ and $y''_{i,j} = \frac{1}{2}y_{i,j} + r$ where $r \in \mathcal{R}$ is a random number. Then encrypt $\{y'_i, y''_i\}$ with (M_1, M_2) into $\{M_1^T y'_i, M_2^T y''_i\}$. Output $EY = \{M_1^T y'_i, M_2^T y''_i\}_{i=1,\cdots,n}$

As in Figure 2, step (2), to detect whether a detection window is a face, US encrypt the window and upload to the cloud server.

E-DW(SK,w): To encrypt the window w, US split vector w into two vectors $\{w', w''\}$ following the rule: for each $w_j \in w$, set $w'_j = w'_j = w_j$ if $s_j \in S$ is 0; otherwise $w'_j = \frac{1}{2}w_j - r'$ and $w''_j = \frac{1}{2}w_j + r'$ where $r' \in \mathcal{R}$ is another random number. Then encrypt $\{w', w''\}$ with (M_1, M_2) into $\{M_1^{-1}w', M_2^{-1}w''\}$. Output $EW = \{M_1^{-1}w', M_2^{-1}w''\}$

After receiving the secure classifiers and secure detected window, the cloud server output the detection results as in Figure 2, step (3).

DC(EY,EW): For each secure classifier $\{M_1^T y'_i, M_2^T y''_i\}$, the cloud server first compute

$$t_i = (M_1^T y'_i)^T \cdot M_1^{-1}w' + (M_2^T y''_i)^T \cdot M_2^{-1}w'' = y_i^T \cdot w$$

and set $h_i = \alpha_i$ if $t_i \geq \theta_i$ or $h_i = \beta_i$ otherwise. At last, the could server output $H = sign(\sum_{i=1}^n h_i)$ as the detection result.

3.2 Face Recognition and Label Vector

After detect all the face in the photos, as in Figure 2, step (4), the users US run the face recognition algorithm, i.e. SPR [Wright *et al.*, 2009], and form the face label vector $L_i \in \{0,1\}^t$ for each photo, which describes who is/are in each photo. US set the label set $\{L_i\}$ as the index of the photo.

3.3 Secure Face Label Matching

In order to build the secure label vectors of the photo set, the users US first generate the private key as follows:

M-KeyGen(t): Given a security parameter t as the totally face number of the photo set, output the private key $PrK(N_1, N_2, T)$, where $S_1, S_2 \in \mathcal{R}^{t\times t}$ are randomly invertible matrices and $T \in \{0,1\}^t$ is a randomly vector.

As in Figure 2, step (5), US encrypt his label vectors and upload to the cloud server.

E-LV(PrK,L): To encrypt the label vector L_i in the vector set, US split each vector L_i into two vectors $\{L'_i, L''_i\}$ following the rule: for each $L_{i,j} \in L_i$, set $L'_{i,j} = L''_{i,j} = L_{i,j}$ if $t_j \in T$ is 1; otherwise $L'_{i,j} = \frac{1}{2}L_{i,j} - u$ and $L''_{i,j} = \frac{1}{2}L_{i,j} + u$ where $u \in \mathcal{R}$ is a random number. Then encrypt $\{L'_i, L''_i\}$

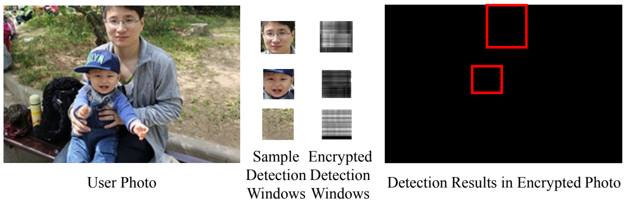
User Photo
Sample Detection Windows
Encrypted Detection Windows
Detection Results in Encrypted Photo

Figure 3: The secure face detection. The user photo are divided into detection windows, which are sent to the cloud one by one. The detection results are shown in rectangle.

with (N_1, N_2) into $\{N_1^T L'_i, N_2^T L''_i\}$. Output $EL = \{N_1^T L'_i, N_2^T L''_i\}_{i=1,\cdots,t}$

Next, The users US choose the standard encrypted algorithm such as AES [Daemen and Rijmen, 2002], or other photo encryption scheme such as [Jin *et al.*, 2015], with their own secret key, to encryption the photos, and upload to the cloud server with the secure label vectors.

To search the photos with target faces, US first generate the query as $Q = \{0, 1\}^t$, as in Figure 2, step (9), where $Q_i = 1$ if the i-th face is one of the target faces, then US encrypt the query and upload it to the cloud server with the amount of target faces λ, as in Figure 2, step (10).

E-Q(PrK,Q): To encrypt the query Q, US split vector Q into two vectors $\{Q', Q''\}$ following the rule: for each $Q_j \in Q$, set $Q'_j = Q'_j = Q_j$ if $t_j \in T$ is 0; otherwise $Q'_j = \frac{1}{2}Q_j - v$ and $Q''_j = \frac{1}{2}w_j + v$ where $v \in \mathcal{R}$ is another random number. Then encrypt $\{Q', Q''\}$ with (N_1, N_2) into $\{N_1^{-1}Q', N_2^{-1}Q''\}$. Output $EQ = \{N_1^{-1}Q', N_2^{-1}Q''\}$

After receiving the secure index and secure query, the cloud server output the matching result.

MAT(EL,EQ): For each secure label $\{N_1^T L'_i, N_2^T L''_i\}$, the cloud server first compute

$$ret_i = (N_1^T L'_i)^T \cdot N_1^{-1}Q' + (N_2^T L''_i)^T \cdot N_2^{-1}Q'' = L_i^T \cdot Q$$

if $ret_i = \lambda$ the cloud server then set the $i - th$ photo with the label L_i as one of the matching photos.

As in Figure 2, step (11), the cloud server return the encryption retrieval photos to US, then US decryption them to get the matching photos.

4 Experiments

We convert the Viola-Jones type face detector [Viola and Jones, 2001; 2004] to our secure face detector, which is implemented by OpenCV 2.4.3. [1] package. The face detector consists of a cascade of 22 rejectors. The first rejector consists of 3 weak classifiers. The most complicated rejector consists of 213 weak classifiers. There is a total of 2135 weak classifiers.

In this section, we show an experiment on photos from an authorized family, which consists of 5 family members with 4 mobile phones. We use 100 photos (20 photos for each member) to build the dictionary. The number of total family photos in the simulated cloud is 1000. The secure face detection results are shown in Figure 3. The local face recognition results are shown in Figure 4. The secure face matching results are shown in the supplementary material.

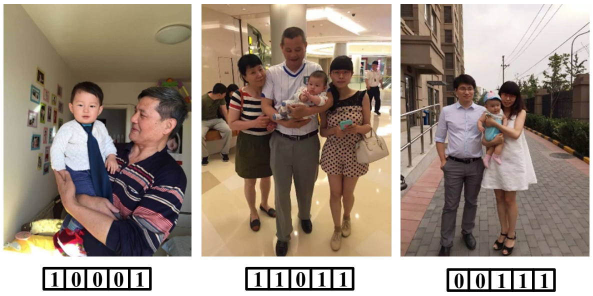

Figure 4: The local face recognition. We label the recognition results in label vector, where 1 means that the corresponding person is in that photo. Then, the label vector is encrypted and sent to the cloud server.

5 Conclusion

In this paper, we propose a novel protocol to preserve the privacy of the users' photos and the parameters of the commercial face detector simultaneously in mobile cloud scenarios. The experimental results reveal that our protocol can successfully retrieve the proper photos from the cloud server and protect the user photos and the face detector. One of the core of the convolutional neural network (CNN) is inner product. Thus, in the future work, we will extent our approach to privacy preserving deep learning framework for face retrieval.

References

[Avidan and Butman, 2006] Shai Avidan and Moshe Butman. Blind vision. In *Computer Vision - ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part III*, pages 1–13, 2006.

[Bost *et al.*, 2015] Raphael Bost, Raluca Ada Popa, Stephen Tu, and Shafi Goldwasser. Machine learning classification over encrypted data. In *22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2014*, 2015.

[Chu *et al.*, 2014] Chun-Te Chu, Jaeyeon Jung, Zhicheng Liu, and Ratul Mahajan. strack: Secure tracking in community surveillance. In *Proceedings of the ACM International Conference on Multimedia, MM'14, Orlando, FL, USA, November 03 - 07, 2014*, pages 837–840, 2014.

[Daemen and Rijmen, 2002] Joan Daemen and Vincent Rijmen. The design of rijndael: Aes - the advanced encryption standard. *Springer-Verlag*, 2002.

[Fanti *et al.*, 2013] Giulia C. Fanti, Matthieu Finiasz, and Kannan Ramchandran. One-way private media search on

[1] http://opencv.org/

public databases: The role of signal processing. *IEEE Signal Process. Mag.*, 30(2):53–61, 2013.

[Jin *et al.*, 2015] Xin Jin, Yingya Chen, Shiming Ge, Kejun Zhang, Xiaodong Li, Yuzhen Li, Yan Liu, Kui Guo, Yulu Tian, Geng Zhao, Xiaokun Zhang, and Ziyi Wang. *Applications and Techniques in Information Security: 6th International Conference, ATIS 2015, Beijing, China, November 4-6, 2015, Proceedings*, chapter Color Image Encryption in CIE L*a*b* Space, pages 74–85. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[Jin *et al.*, 2016a] Xin Jin, Kui Guo, Chenggen Song, Xiaodong Li, Geng Zhao, Jing Luo, Yuzhen Li, Yingya Chen, Yan Liu, and Huaichao Wang. Private video foreground extraction through chaotic mapping based encryption in the cloud. In *MultiMedia Modeling - 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part I*, pages 562–573, 2016.

[Jin *et al.*, 2016b] Xin Jin, Yaming Wu, Xiaodong Li, Yuzhen Li, Geng Zhao, and Kui Guo. Ppvibe: Privacy preserving background extractor via secret sharing in multiple cloud servers. In *8th International Conference on Wireless Communications & Signal Processing, WCSP 2016, Yangzhou, China, October 13-15, 2016*, pages 1–5, 2016.

[Jin *et al.*, 2017] Xin Jin, Peng Yuan, Xiaodong Li, Chenggen Song, Shiming Ge, Geng Zhao, and Yingya Chen. Efficient privacy preserving viola-jones type object detection via random base image representation. In *in Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10-14 July, 2017*, 2017.

[Osadchy *et al.*, 2010] Margarita Osadchy, Benny Pinkas, Ayman Jarrous, and Boaz Moskovich. Scifi - A system for secure face identification. In *31st IEEE Symposium on Security and Privacy, S&P 2010, 16-19 May 2010, Berleley/Oakland, California, USA*, pages 239–254, 2010.

[Shashank *et al.*, 2008] Jagarlamudi Shashank, Palivela Kowshik, Kannan Srinathan, and C. V. Jawahar. Private content based image retrieval. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA*, 2008.

[Sohn *et al.*, 2010] Hosik Sohn, Konstantinos N. Plataniotis, and Yong Man Ro. Privacy-preserving watch list screening in video surveillance system. In *Advances in Multimedia Information Processing - PCM 2010 - 11th Pacific Rim Conference on Multimedia, Shanghai, China, September 21-24, 2010, Proceedings, Part I*, pages 622–632, 2010.

[Upmanyu *et al.*, 2009] Maneesh Upmanyu, Anoop M. Namboodiri, Kannan Srinathan, and C. V. Jawahar. Efficient privacy preserving video surveillance. In *IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 - October 4, 2009*, pages 1639–1646, 2009.

[Viola and Jones, 2001] Paul A. Viola and Michael J. Jones. Robust real-time face detection. In *IEEE 8th International Conference On Computer Vision ICCV 2011, Vancouver, British Columbia, Canada, July 7-14, 2001*, page 747, 2001.

[Viola and Jones, 2004] Paul A. Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[Wong *et al.*, 2009] Wai Kit Wong, David Wai-lok Cheung, Ben Kao, and Nikos Mamoulis. Secure knn computation on encrypted databases. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, page 139152, New York, NY, USA, 2009. ACM.

[Wright *et al.*, 2009] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S. Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009.