2012 IEEE Second International Conference on Consumer Electronics - Berlin (ICCE-Berlin)

A Real-time Multi-User Face Unlock System via Fast Sparse Coding Approximation

I-Kuei Chen, Yi-Min Tsai, Jyh-Jing Hwang, and Liang-Gee Chen
DSP/IC Design Lab, Graduate Institute of Electronics Engineering, National Taiwan University, Taiwan
{*ikchen,ymtsai,jjhwang,lgchen*}*@video.ee.ntu.edu.tw*

***Abstract*—We propose a multi-user face unlock system based on fast sparse coding approximation. Different approximation techniques of sparse coding are compared for real-time processing and recognition rate. The system is capable of online new user registration without specific training, and it is designed for continuously processing video sequences as on-the-fly face classification. The system is tested with a public face dataset for real-world applications, which is able to recognize 100 different face identities in 0.0028s on a regular PC, with 93.8% recognition rate.**

I. Introduction

For human, face recognition is the most natural and convenient way to identify people compared to other biometric information, such as finger prints, palm prints or iris patterns. In recent years, face recognition becomes robust and realistic in consumer electronics for unlocking systems. For mobile devices such as smart phones or tablets, typing passwords is sometimes inconvenient. In addition, facilities like banks and hospitals, which involves security and privacy, are all suitable places adopting a face unlock system. Therefore, face unlock mechanism becomes a potential and promising solution in these situations.

Traditional face recognition technologies often require huge amounts of training samples, computation time for training and memory usage for tuned parameters (such as SVM, Random forests or Adaboost). In recent years, sparse coding technologies are developed in various areas, including object recognition, superresolution, inpainting and denoising. For classification problems, sparse coding requires only few samples from a user class to achieve competing performance. It does not need explicit training when a new user class is added, which is also beneficial to the ability in separating backgrounds and target patterns [2].

Yang et al. [1] propose a SIFT-based approach with sparse coding, and Wang et al. [3] further reduce the computation complexity by locality-constrained coding (LLC). Sparse representation classification (SRC) [2] achieves state-of-art performance in face recognition database with only solving the $l1$ minimization problem. Zhang et al. [4] further analyze the power of sparsity in detail, and they discover $l2$ sparsity is good enough for face recognition in SRC scenario.

In this paper, we focus on providing a face unlock system which can be used in real time on input video. The system is capable of adding multiple new user registrations without specific training. Our face unlock system is based on SRC, but we further compare different suitable approximation techniques to reduce the overall computation time while preserving or even improving the recognition rate.

II. Proposed System

The proposed system (Fig. 1) consists of two phases, on-line registration (training) and on-the-fly activation (testing). The registration phase enrolls data of a new user into the existing training results which is applied to the activation phase for verifying the current input frame. One registered user maps to one user class when classified in the activation phase. In our system, we provides a GUI-aided visual preprocessing tool to guide users actively to crop their frontal face images. Users are requested to align frontal face for registration or activation (unlocking). This prevents variations including rotation-in-plane, rotation-off-plane, scaling, occlusion and misalignment, which may reduce overall accuracy.

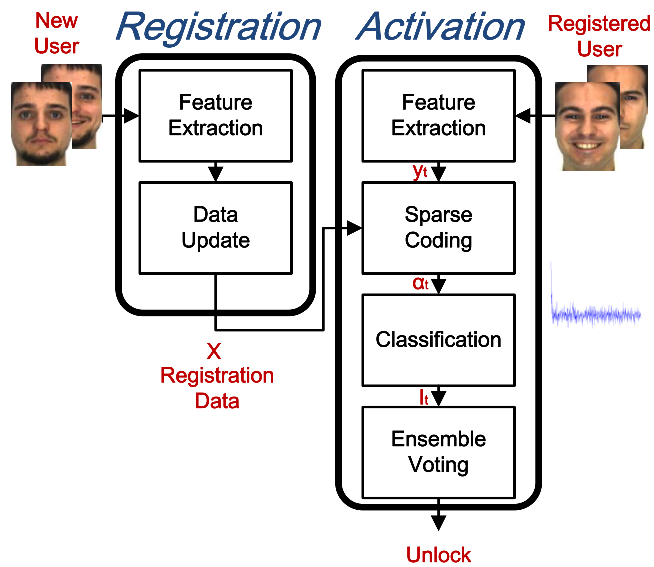


Fig. 1. Details of System Flow

The activation phase treats input frames as continually testing features for identifying. After feature extraction, sparse coding, classification and ensemble voting stages, the user can be identified as one of registered users or not. The system simultaneously filters non-facial images using sparse coefficients analysis [2], and thus input frames can be processed continuously in real time whether the user is settled or not, preventing heavy face detection resources.

A. Registration

1) Feature Extraction: Feature extraction performs dimension reduction for input faces. After reading video frames, images are cropped and resized into 60×43 frontal face images, which are then transformed into low-dimensional features through eigen-face extraction [7]. Each extracted feature is 300-dimension.

2) Data Update: We adopt lazy learning for previously registered users' data X_{prev} and current new one. New user features of multiple frames around confirmation point are set as new identity class X_{new}, and they are directly appended (cascaded) to the existing registration data. The total registered set X is thus formed for the subsequent activation phase.

$$X = [X_{prev},\ \ X_{new}] \tag{1}$$

B. Activation

1) Feature Extraction: We apply the same feature extraction as that in Registration phase. The feature output of the current frame is denoted as y_t.

2) Sparse Coding: After reading registration data X, we solve the feature y_t of the current input face into the sparse coefficients α_t of all possible candidate samples from X. Note that the registration data X is composed of cascaded features of registered class, and each registered user class includes features of 7 input frames. The sparse coding algorithms we tested are as below:

•*l*1-minimization sparse coding (*l*1):

We use Sparse Representation Classification (SRC) [2] to solve α_t for later error term calculation, and we adopt GPSR_BB [6] as the *l*1 solver,

$$\hat{\alpha}_t = \arg\min_{\alpha_t} \|\alpha_t\|_1 \; s.t. \; \|y_t - X\alpha_t\|_2 < \varepsilon \tag{2}$$

•*l*2-minimization sparse coding (*l*2):

In [4], Collaborative Representation Classification (CRC) was proposed as a *l*2 solver, which outperforms other *l*1 solvers in recognition rate and processing time. We test this solver as fast sparse approximation for solving α_t, and the analytical is as (4).

$$\hat{\alpha}_t = \arg\min_{\alpha_t} \{\|y_t - X\alpha_t\|_2^2 + \lambda \|\alpha_t\|_2^2\} \tag{3}$$

$$\hat{\alpha}_t = (X^T X + \lambda \cdot I)^{-1} X^T y_t \tag{4}$$

•K-nearest Neighbor *l*2 sparse coding (KNN-*l*2):

In [3], Locality-constraints Coding (LLC) was proposed for fast SIFT code words solving, which fortifies sparsity in *l*2 solvers to improve performance. We apply a similar LLC concept for SRC scenario. We use *l*2 solver to find K-nearest-neighbor of y_t in X first, and then use the KNN results to generate a new dataset X_{KNN} to solve (3) for α_{KNN}. Later restore it with 0 into full α_t. By doing so, we improve the sparsity of *l*2 solution. (K is 200 in our optimal setting. The reason why we choose this K is that we want to follow and compare with the l1 optimal sparsity, as shown in Table. I.)

3) Classification: The system performs classification based on the solved sparse coefficients α_t. The candidate user classes from database with the smallest reconstruction error is chosen as the classification result $Identity_t$ (I_t). i denotes the specific user class used in X and α.

$$Identity_t = \arg\min_i \{\|y - X_i \hat{\alpha}_i\|_2\} \tag{5}$$

4) Ensemble Voting: Ensemble voting is executed on successive input frames. The system is unlocked only if certain continuous frames are all indicates the same user class I. After unlocking, private theme, personal setting and related files can be restored for the identified user.

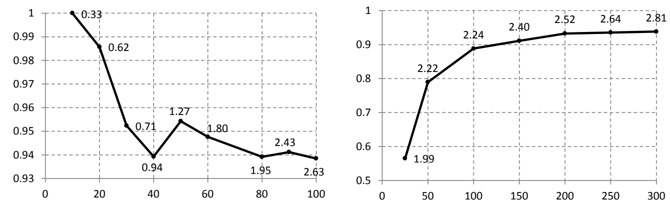
III. Experimental Results

Our multi-user face unlock system is tested with AR face dataset [5]. The testing environment is on a regular PC with 3.4GHz CPU and 4Gb RAM. Experiments are conducted with Matlab. All faces of a registered user map to one user class only. 7 registered faces from each 100 candidate classes are included in training data, and the system tests other 7 faces in these 100 classes separately. Table I shows the benchmarking of different sparse coding methods, including recognition rate, processing time and sparsity. Sparsity is defined as the ratio of zeros in the α_t

As Table. I shows, simply fortifying sparsity does not help recognition rate, and *l*2 is still the best solution so far. The reason why sparsity is not so important in SRC scenario is that the dataset X is composed of direct instances from different classes, unlike some dictionaries which are designed for sparsity. Thus in those sparsity-related dictionaries, fortifying sparsity may be helpful.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT SPARSE CODINGS.

	*l*1	*l*2	KNN-*l*2
Recognition rate (%)	0.823	**0.938**	0.901
Processing time (s)	0.1040	**0.0028**	0.0034
Frame per second (FPS)	10	**356**	290
Sparsity	0.861	0	0.857



(a) (b)

Fig. 2. Comparison of recognition rate (%, Y axis) in different settings (X axis). The number beside points denotes processing time (ms). (a) the number of user classes. (The 300-dim feature is used.) (b) the dimension of features. (100 classes are used.)

We also conduct experiments on *l*2 with different settings of class numbers and feature dimensions (Fig. 2). As the number of user classes decreases as shown in Fig. 2(a), recognition rate increases, and processing time decreases. It indicates that the system runs better when total registered users are not so much. It can be observed that users can be recognized well if only 10 classes (10 users) of the AR face dataset are needed to be classified. Reasonable trade-offs between the recognition rate and the processing time are shown in Fig. 2(b). It shows merely 100-dimension features can achieve acceptable recognition rate.

IV. Conclusion

In this paper we propose a multi-user face unlock system which can be used in real-time video input by solving fast sparse approximation. The proposed system includes functionality of fast registration for new users and unlocking activation in real time. It is capable of classifying 100 user identities with 356 fps and 93.8% recognition rate, which demonstrates a practical solution for real-world applications.

References

[1] Jianchao Yang, Kai Yu, Yihong Gong, T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, vol., no., pp.1794-1801, 20-25 June 2009

[2] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Yi Ma, "Robust Face Recognition via Sparse Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.31, no.2, pp.210-227, Feb. 2009

[3] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, T. Huang, Yihong Gong, "Locality-constrained Linear Coding for image classification," *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, vol., no., pp.3360-3367, 13-18 June 2010

[4] Lei Zhang, Meng Yang, Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," *Computer Vision (ICCV), 2011 IEEE International Conference on*, vol., no., pp.471-478, 6-13 Nov. 2011

[5] A.M. Martinez, A.C. Kak, "PCA versus LDA," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.23, no.2, pp.228-233, Feb 2001

[6] M.A.T. Figueiredo, R.D. Nowak, S.J. Wright,"Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," *Selected Topics in Signal Processing, IEEE Journal of* , vol.1, no.4, pp.586-597, Dec. 2007

[7] M. Kirby, L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.12, no.1, pp.103-108, Jan 1990