DICTIONARY-BASED VIDEO FACE RECOGNITION USING DENSE MULTI-SCALE FACIAL LANDMARK FEATURES

Jun-Cheng Chen, Vishal M. Patel, Huy Tho Ho and Rama Chellappa

Center for Automation Research, University of Maryland, College Park, MD 20742

{pullpull, pvishalm, huytho, rama}@umiacs.umd.edu

ABSTRACT

In video-based face recognition, different video sequences of the same subject contain variations in pose, illumination, and expression which contribute to the challenges in designing an effective video-based face-recognition system. In this paper, we propose a dictionary-based approach using dense and high-dimensional features extracted from multi-scale patches centered at detected facial landmarks for video-to-video face identification and verification. Experiments using unconstrained video sequences from Multiple Biometric Grand Challenge (MBGC) and Face and Ocular Challenge Series (FOCS) datasets show that our method performs significantly better than many state-of- the-art video-based face recognition algorithms.

Index Terms— Video-based face recognition, dense multi-scale features, facial landmark detection, dictionary learning.

1. INTRODUCTION

Face recognition is one of the fundamental problems in computer vision and has a wide range of applications [1], including surveillance, social networks and augmented reality. Although many face recognition algorithms have demonstrated promising results under controlled environments with cooperative subjects, face recognition in real-world scenarios is highly unconstrained and needs to handle large changes in pose, lighting, image quality (i.e., low-resolution and blur), expression and occlusion. These factors make the unconstrained face recognition extremely difficult. The unconstrained video face recognition is even more challenging than the still-face case because there is more intra-video and intra-class variations associated with pose and illumination conditions. Furthermore, a large amount of data in videos makes computing efficiency another important issue.

Motivated by the successes of high-dimensional facial features in still-face recognition [2], sparse representation [3] and dictionary learning for video-based face recognition [4][5][6], we propose a dictionary-based approach using dense high-dimensional feature for unconstrained video-to-video face identification and verification problems. We first segment the face videos into K partitions and extract multi-scale features from patches centered at detected dense facial landmarks. Then, we learn a compact and representative dictionary from dense features for each partition and form a video dictionary for each video by concatenating sub-dictionaries. Finally, the learned video dictionaries are used for face identification and verification. Moreover, because the dictionary for each training video is learned independently during the training phase, our approach can thus be easily parallelized not only for testing but also for training. This makes our approach attractive for large-scale video-based face recognition problems. Fig. 1 gives an overview of our method.

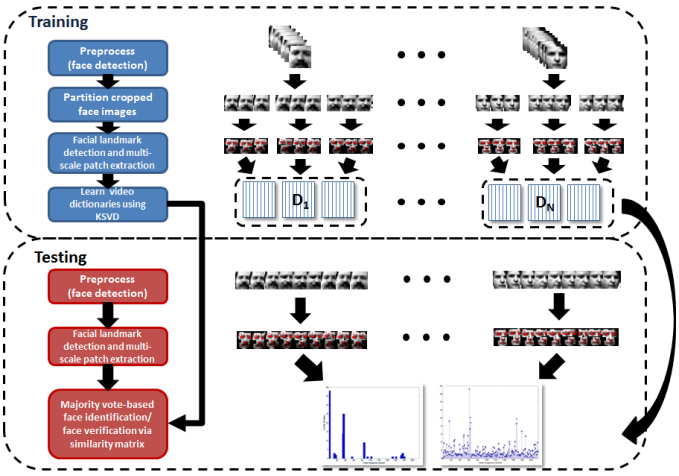


Fig. 1: An overview for our video-based face identification/verification system.

The rest of the paper is organized as follows: We briefly review related works in Section 2. In Section 3, we present our dictionary-based video face recognition algorithm and show experimental results on two challenging video datasets in Section 4. We conclude the paper in Section 5.

2. RELATED WORK

In this section, we briefly review several related works as follows.

Generally, there are two major components of a face recognition system: (1) feature representation and (2) classification algorithm. For feature representation, Coates *et al.* [7] showed that over-complete representation is critical for achieving high recognition rates regardless of the encoding methods used. [8] also showed that densely sampling of overlapped images helps to improve the recognition performance. For still-face recognition, [2] demonstrates excellent results using the high-dimensional multi-scale features extracted from the patches centered at dense facial landmarks. These works show that over-complete and high-dimensional features are important for face recognition.

Most video-based face recognition algorithms can be classified into two main categories: (1) frame-based and (2) image set-based.

1. **Frame-based.** In this category, besides features (e.g., SIFT,

This work was supported by a DARPA grant FA8750-13-2-0279.

LBP) derived from the image intensity data, the temporal (e.g., motion) and spatial-temporal information between cropped faces in a video is usually utilized and encoded in the model to perform the recognition tasks. For example, Zhou *et al.* [9] proposed a tracking-and-recognition approach which lowers the uncertainties of tracking and recognition simultaneously in a unified probabilistic framework. Lee *et al.* [10] learned the nonlinear appearance manifold from face videos to handle both tracking and recognition in a unified framework. In addition, a Hidden Markov Model-based approach [11] has been proposed to exploit the temporal information. However, the performance for these approaches is greatly affected by the tracking accuracy. Poor tracking will introduce lots of background noise into the model and lower the recognition accuracy.

2. **Image set-based.** In this approach, each face video is transformed into an unordered set of images which implies no temporal information is used. The set of images for a subject is usually represented using a subspace model. Then, recognition task is done by measuring the distance between subspaces. Turaga *et al.* [12] presented a statistical method for video-based face recognition. The approach extracted the face subspaces by performing the standard Principle Component Analysis (PCA) for face videos and using tools from Riemannian geometry of the Grassmann manifold to measure distance. Cevikalp *et al.* [13] modeled face image sets using affine or convex hull, and Wang *et al.* [14] modeled them using covariance matrix to encode the underlying manifold structure. Hu *et al.* [15] improved the affine subspace model by enforcing sparsity constraint and used it to measure between-set dissimilarity which is the distance between sparse approximated nearest points of two image sets. Recently, Chen *et al.* [5] proposed a dictionary-based approach for face identification and verification tasks. They learned a compact and representative dictionary for each video and made use of the reconstruction errors of test videos using the learned video dictionaries. The approach is simple and efficient, especially suitable for large-scale video-based face recognition.

Our approach is mainly motivated by Chen *et al.* [5]. Nevertheless, the method in [5] did not take the face alignment into account, and it directly used the whole cropped face image as a feature which might contain irrelevant background and facial features due to inaccurate face tracking (i.e., the size of detected bounding box is much larger than the face.). In contrast, our algorithm (1) exploits the dense features extracted from multi-scale patches centered at facial landmarks [2] which not only mitigate the pose and noise problems due to alignment but also generate informative features, and (2) uses video dictionaries [5] which are efficient and effective representations for video-based face recognition.

3. PROPOSED APPROACH

In this section, we describe the construction of video dictionary using high-dimensional dense facial landmark features and their application to face identification and verification problems.

3.1. Constructing Video Dictionary Using Dense Multi-scale Facial Landmark Features

The training phase of our method consists of three main stages: video partitioning, multi-scale landmark feature extraction and video dictionary learning. In what follows, we describe them in detail.

Fig. 2: For illustration purpose, we visualize the single-scale patch image for the MBGC dataset through assembling all 5×5-pixel patches centered at 26 facial landmarks points together.

Video partitioning: Due to high variability of faces within a video and face tracking accuracy, we find that segmenting video into different partitions usually helps in improving recognition accuracy. A K-means clustering type of algorithm is used to segment the videos [5][16] which incrementally adds each cropped face into a partition with the minimum ratio of within-partition similarity over between-partition similarity.

Dense landmarks and multiple-scale features: It was shown in [2] that multi-scale features centered around facial landmarks contain strong discriminative information and the recognition performance improves as the dimensionality of the feature vector is increased. We extract multi-scale patches centered at facial landmarks of inner faces (i.e., landmarks at eye brows, eyes, nose, and mouth corners. 26 landmarks in total are used in our work) and concatenate them together to form a high-dimensional feature vector. With recent progress in face alignment, there are numerous approaches providing accurate and dense facial landmark detection [17][18]. We adopt [19] because of its excellent performance on low-resolution and lower-quality face images[1]. Detected landmarks and extracted features are shown in Fig.2. However, unlike still-face recognition, directly applying the approach in [2] to video-to-video face recognition is infeasible because the concatenation of feature vectors extracted from each frames in a video yields extremely high-dimensional feature vector (i.e., imagine a video with 100 frames can result in a 100 times long feature vector). A compact and representative model has to be learned to remove noisy and irrelevant features.

Video dictionary: Various algorithms have been proposed in the literature for learning compact and representative dictionaries. One of the well-known algorithm is the K-SVD algorithm [20]. For each partition, we apply the K-SVD algorithm to construct a dictionary which not only captures variations caused by changes in pose and illumination but also reduces temporal redundancy. Let $\mathbf{D}^i_{j,k}$ be the dictionary and $\mathbf{G}^i_{j,k} = [\mathbf{g}^i_{j,k,1}\ \mathbf{g}^i_{j,k,2}\ \ldots]$ be the feature matrix for the kth partition of the jth face video for the ith subject where each column $\mathbf{g}^i_{j,k,l}$ is the extracted dense multi-scale feature for lth face in the kth partition of the jth video. In the K-SVD formulation, the dictionary and sparse coefficient are learned through iteratively minimizing the following reconstruction errors by fixing $\mathbf{D}^i_{j,k}$ and

[1] https://sites.google.com/site/akshayasthana/clm-wild-code.

$\mathbf{X}^i_{j,k}$ in turn.

$$(\hat{\mathbf{D}}^i_{j,k}, \hat{\mathbf{X}}^i_{j,k}) = \underset{\mathbf{D}^i_{j,k}, \mathbf{X}^i_{j,k}}{\operatorname{argmin}} ||\mathbf{G}^i_{j,k} - \mathbf{D}^i_{j,k}\mathbf{X}^i_{j,k}||^2_F \; s.t. \; \forall l, ||\mathbf{x}_l||_0 \leq T_0, \tag{1}$$

where $T_0 \in \mathbb{N}$ is the sparsity constraint and $\mathbf{x}_l$ is the lth column of sparse coefficient matrix $\mathbf{X}^i_{j,k}$. $||\cdot||_0$ is the zero-norm which counts the number of nonzero entries, and $||\cdot||_F$ is the Frobenius norm. Finally, the video dictionary $\mathbf{D}^i_j$ for the jth video of ith subject can be obtained via concatenating all sub-dictionaries learned from the corresponding K partitions

$$\mathbf{D}^i_j = [\mathbf{D}^i_{j,1} \; \mathbf{D}^i_{j,2} \; \ldots \; \mathbf{D}^i_{j,K}]. \tag{2}$$

After video dictionaries are learned, for testing phase we first do the same image preprocessing as in training and extract the multi-scale features for each cropped face image. Then, we perform face identification and verification which are shown in the following subsections.

3.2. Face Identification

Let $\mathbf{P}$ represent the set of the entire gallery videos (i.e., training videos) and $\mathbf{Q}$ represent the set of the entire query videos (i.e., test videos) where $\mathbf{Q}^m$ is the mth query video with $m = 1, 2, \ldots, |\mathbf{Q}|$. In addition, the feature vector for lth frame in mth query video is denoted as q^m_l where $l = 1, 2, \ldots, |\mathbf{Q}^m|$. The learned dictionary for the pth gallery videos is denoted as $\mathbf{D}_p$ where $p = 1, 2, \ldots, |\mathbf{P}|$. The original identification problem can be converted as finding the gallery video dictionary which produces the minimum reconstruction error for $\mathbf{q}^m_l$:

$$\hat{p} = \underset{p}{\operatorname{argmin}} ||\mathbf{q}^m_l - \mathbf{D}_p\mathbf{D}^\dagger_p\mathbf{q}^m_l||_2, \tag{3}$$

where $\mathbf{D}^\dagger_p = (\mathbf{D}^T_p\mathbf{D}_p)^{-1}\mathbf{D}^T_p$ is the pseudo inverse of $\mathbf{D}_p$ and $\mathbf{D}_p\mathbf{D}^\dagger_p\mathbf{q}^m_l$ is the projection of $\mathbf{q}^m_l$ onto the subspace spanned by the atoms of $\mathbf{D}_p$.

Then, the final decision is made for $\mathbf{Q}^m$ through aggregating the voting results from its frames as

$$p^* = \underset{p}{\operatorname{argmax}} \; \mathbf{C}_p, \tag{4}$$

where $\mathbf{C}_p$ is the total number of the frames in $\mathbf{Q}^m$ voting to the pth gallery video. The subject identity can be decided through the video-to-subject mapping as $i = m(p^*)$.

3.3. Face Verification

Different from the identification problem, the goal of face verification is to determine whether a pair of faces is from the same subject, and its performance is usually measured using the receiver operating characteristic (ROC) curve which shows the relations between false acceptance rates (FAR) and true acceptance rates (TAR). In addition, ROC curves are ploted based on a similarity matrix which is computed according to the distance between gallery and query videos. In our framework, we can directly use the minimum reconstruction error between $\mathbf{Q}^m$ and $\mathbf{D}_p$ as the distance. Thus, the (m,p)th entry of the similarity matrix $\mathbf{S}$ can be computed as

$$\mathbf{S}_{m,p} = \min_{l \in \{1,2,\ldots,|\mathbf{Q}^m|\}} ||\mathbf{q}^m_l - \mathbf{D}_p\mathbf{D}^\dagger_p\mathbf{q}^m_l||_2. \tag{5}$$

Fig. 3: The upper row shows the example frames from the MBGC walking sequences in four different scenarios. Similarly, the bottom row presents the example frames from the FOCS UT-Dallas walking videos.

4. EXPERIMENTAL RESULTS

To evaluate our approach, we present the face identification and verification results on two well-known public datasets for unconstrained video-based face recognition: (1) Multiple Biometric Grand Challenge (MBGC) [21], and (2) Face and Ocular Challenge Series (FOCS) [22]. We perform our experiments following the experimental design described in [5][23].

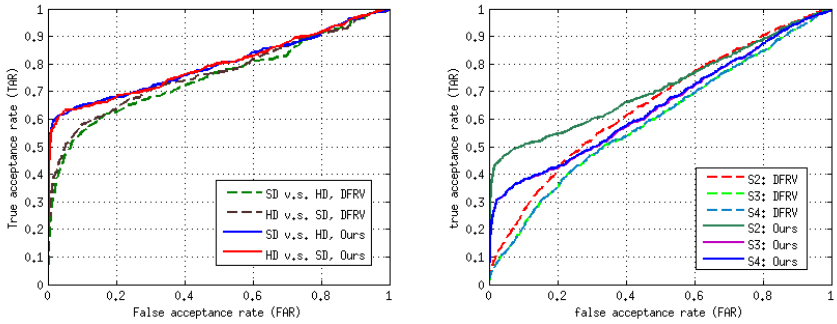
4.1. Implementation Details

We used the face detector in OpenCV [24] and IVT [25] for face detection and face tracking respectively to crop the faces from each video. All cropped faces are downsampled and normalized to 20×20 pixels, and two patch sizes are used for multi-scale feature extraction: (1) 5×5 and (2) 7×7 pixels. In addition, we segment $K = 3$ partitions for each video in the MBGC dataset and the FOCS dataset in all of our experiments. Prior to dictionary learning, we augment the feature matrix for each partition by adding more multi-scale patch features which are extracted via shifting the original bounding boxes of patches by one or two pixels to all directions or rotating them with a small angle. This helps the partition step in assigning video frames to learn an improved dictionary and helps in reducing the noise caused by tracking and landmark detection. The same augmentation is also applied to query videos before recognition.

4.2. Multiple Biometric Grand Challenge

In the MBGC video version 1 dataset (Notre Dame dataset), there are 146 subjects in total, and videos for each subject are available in two formats: standard definition (SD, 720×480 pixels) and high definition (HD, 1440×1080 pixels). It consists of 399 walking sequences where 201 of them are in SD format and 198 in HD, and 371 activity sequences where 185 in SD and 186 in HD. For the walking sequences as illustrated in Fig. 3, subjects usually walk toward the camera and keep their faces frontal with respective to it for most of the time and turn their face to the left or right at the end. On the contrary, the activity sequences contains most non-frontal views for each subject. The challenge for the dataset includes blurred faces caused by motion, frontal and non-frontal faces in shadow which also induce face tracking difficulty to crop faces from the video.

We conduct leave-one-out identification experiments on three subsets of the cropped face images acquired from the walking videos and present the identification accuracy in Table 1. Our proposed method outperforms the other approaches. The three subsets are S2, S3, and S4, respectively where S2 is the set of subjects who have at least two face videos available, S3 at least three available, and S4



(a) (b)

Fig. 4: (a) shows the leave-one-out verification results for MBGC walking videos, and (b) is the verification results for SD v.s. HD and HD v.s. SD where DFRV is the dictionary-based method proposed in [5]. From the figures, we see that our method with dense multi-scale feature achieves better verification results.

at least four available (S2: 144 subjects, 397 videos in total, S3: 55 subjects, 219 videos in total, and S4: 54 subjects, 216 videos).

MBGC walking videos	WGCP [12]	SANP [15]	DFRV [5]	KSRV [23]	Ours
S2	63.79	83.88	85.64	86.65	**89.17**
S3	74.88	84.02	88.13	88.58	**89.04**
S4	75	84.26	88.43	88.89	**89.35**
Average	71.22	84.05	87.40	88.04	**89.19**

Table 1: Identification rate for leave-one-out face identification experiments for the MBGC walking videos. Our method achieves the best results.

Furthermore, we divide videos from S2 into two groups according to their recording resolution: SD and HD. Then, we use both of them as gallery and probe in turn and perform the face identification to study the performance change under different image quality setting. The results are presented in Table 2. As can be seen from this table, our algorithm outperforms other approaches even when the quality of videos are different in the gallery and probe videos.

MBGC walking videos	WGCP [12]	SANP [15]	DFRV [5]	KSRV [23]	Ours
SD v.s. HD	30.15	41.71	86.93	91.46	**92.46**
HD v.s. SD	30.30	45.96	91.41	91.41	**93.94**
Average	30.23	43.84	89.17	91.44	**93.2**

Table 2: Identification rate for SD v.s. HD (SD as probe; HD as gallery) and HD v.s. SD (HD as probe; SD as gallery) face identification experiments for the MBGC walking videos. Our method achieves the best results.

We also present the verification results in Fig. 4 which compares our approach with DFRV [5] which is also a dictionary-based algorithm. The proposed approach achieves better ROC than DFRV which essentially demonstrates the effectiveness of dense multi-scale facial landmark features.

4.3. Face and Ocular Challenge Series

The FOCS UT-Dallas dataset contains 510 walking (frontal-face) and 506 activity (non-frontal face) video sequences for 295 subjects in the resolution, 720 $\times$ 480 pixels. The sequences were acquired on different days. For the walking sequences, subjects stand far away from the camera originally, and then walk toward the camera keeping their face in frontal pose and turn away at the end. For the dataset, we conducted the same leave-one-out tests on 3 subsets: S2 (189 subjects, 404 videos), S3 (19 subjects, 64 videos), and S4 (6 subjects, 25 videos) for UT-Dallas walking videos.

UT-Dalas walking videos	PM [26][12]	WGCP [12]	SANP [15]	DFRV [5]	Ours
S2	38.12	53.22	48.27	59.90	**61.39**
S3	60.94	70.31	60.94	78.13	**79.69**
S4	64	76	68.00	80.00	**84.00**
Average	54.35	66.51	59.07	72.68	**75.02**

Table 3: Identification rate for leave-one-out face identification experiments for the FOCS UT-Dallas walking videos. Our method achieves the best results.

The results are shown in Table 3. Our approach performs the best when compared to other competitive methods.

5. CONCLUSION AND FUTURE WORK

In this paper, we demonstrated that the proposed dictionary approach with dense facial landmark features is effective for unconstrained video-based face identification and verification. Experiments using the MBGC and FOCS datasets have shown that high-dimensional features extracted from multi-scale patches centered around the detected dense facial landmarks provide strong discriminative information upon different pose and illumination conditions among subjects, and video dictionaries provide an efficient and feasible way to utilize the high-dimensional features for large-scale unconstrained video-based face recognition. For future work, we will study different approaches for learning dictionaries [27][28] to robustly handle situations when faces have severe occlusion and extreme illumination changes.

6. REFERENCES

[1] W. Y. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] D. Chen, X. D. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[3] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[4] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 954–965, 2012.

[5] Y.-C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition from video," in *European Conference on Computer Vision*, pp. 766–779. 2012.

[6] V. M. Patel, Y.-C. Chen, R. Chellappa, and P. J. Phillips, "Dictionaries for image and video-based face recognition," *Journal of the Optical Society of America A*, vol. 31, no. 5, pp. 1090–1103, 2014.

[7] A. Coates, A. Y. Ng, and H. L. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 215–223.

[8] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2559–2566.

[9] S. H. Zhou, V. Krueger, and R. Chellappa, "Probabilistic recognition of human faces from video," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 214–245, 2003.

[10] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 303–331, 2005.

[11] X. M. Liu and T. H. Chen, "Video-based face recognition using adaptive hidden markov models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003, vol. 1, pp. I–340.

[12] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2273–2286, 2011.

[13] H. Cevikalp and B. Triggs, "Face recognition based on image sets," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2567–2573.

[14] R. P. Wang, H. M. Guo, L. S. Davis, and Q. H. Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2496–2503.

[15] Y. Q. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 121–128.

[16] N. Shroff, P. Turaga, and R. Chellappa, "Video precis: Highlighting diverse aspects of videos," *IEEE Transactions on Multimedia*, vol. 12, no. 8, pp. 853–868, 2010.

[17] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 545–552.

[18] X. D. Cao, Y. C. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2887–2894.

[19] A. Asthana, S. Zafeiriou, S. Y. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3444–3451.

[20] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[21] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. OToole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, and Y. M. Lui, "Overview of the multiple biometrics grand challenge," in *Advances in Biometrics*, pp. 705–714. Springer, 2009.

[22] "National institute of standards and technology: Face and ocular challenge series, http://www.nist.gov/itl/iad/ig/focs.cfm," .

[23] Y.-C. Chen, V. M. Patel, S. Shekhar, R. Chellappa, and P. J. Phillips, "Video-based face recognition via joint sparse representation," in *IEEE conference on Automatic Face and Gesture Recognition*, 2013.

[24] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[25] D. A. Ross, J. W. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125–141, 2008.

[26] P. Turaga, A. Veeraraghavan, and R. Chellappa, "Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[27] H. V. Nguyen, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Design of non-linear kernel dictionaries for object recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5123–5135, 2012.

[28] E. J. Candes, X. D. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11, 2011.