Face Recognition across Time Lapse Using Convolutional Neural Networks

Hachim El Khiyari, Harry Wechsler

Department of Computer Science, George Mason University, Fairfax, VA, USA
Email: helkhiya@gmu.edu, wechsler@gmu.edu

Abstract

Time lapse, characteristic of aging, is a complex process that affects the reliability and security of biometric face recognition systems. This paper reports the novel use and effectiveness of deep learning, in general, and convolutional neural networks (CNN), in particular, for automatic rather than hand-crafted feature extraction for robust face recognition across time lapse. A CNN architecture using the VGG-Face deep (neural network) learning is found to produce highly discriminative and interoperable features that are robust to aging variations even across a mix of biometric datasets. The features extracted show high inter-class and low intra-class variability leading to low generalization errors on aging datasets using ensembles of subspace discriminant classifiers. The classification results for the all-encompassing authentication methods proposed on the challenging FG-NET and MORPH datasets are competitive with state-of-the-art methods including commercial face recognition engines and are richer in functionality and interoperability than existing methods as it handles mixed biometric datasets, e.g., FG-NET and MORPH.

Keywords

1. Introduction

Time lapse characteristic of face aging is a complex process that has been studied in various disciplines including biology, human perception and more recently in biometrics. The effects of aging alter both the shape and texture of the face and vary according to age, time lapse and demographics such as gender and ethnicity. From birth to adulthood, the effects are encountered mostly in the shape of the face, while from adulthood through old age aging affects the face texture (e.g., wrinkles). Face aging is also affected by external factors such as envi-

ronment and lifestyle. Face recognition across time lapse belongs to the general topic of face recognition in uncontrolled or wild settings and affects security solutions that involve human biometrics. The challenge is substantial since the appearance of human subjects in images used for training or enrollment can vary significantly from their appearance during the ultimate recognition. To address these challenges, robust age-invariant methods must be developed. Age invariance can be implemented either at the feature extraction level or at the training and/or recognition level. At the feature extraction level, the goal is to derive image descriptors that are robust to intra-class aging variation. At the training and testing level, one thrives for low generalization errors in the presence of aging. This paper mostly copes with aging variation at the feature extraction level and leverages a deep learning approach for automatic feature extraction using cascaded convolutional neural networks (CNN). The approach advanced and described in this paper further copes with training and testing aspects using ensembles of classifiers driven by such features for authentication and ultimately is shown able to cope with mixed biometric datasets. This is characteristic of interoperability or equivalently of cross-training and testing across mixed datasets [1]. Interoperability is mostly concerned with domain adaptation, in general, and transfer learning, in particular, for cross-modal generalization over features and parameter settings.

Rather than deriving and using hand-crafted features, this paper advances the utility and feasibility of using convolutional neural networks for automatic feature detection for the purpose of face recognition subject to time lapse. Automatic feature extraction using convolutional neural networks (CNN) [2] can provide highly interoperable descriptors that are robust to aging variations and add much flexibility when deployed and used in biometric systems. We show later on that (a) encodings of facial images using CNN compare favorably to existing methods using generic but hand-crafted features such as Gabor-based descriptors, SIFT, or LBP, and that (b) coupling of CNN for feature extraction with ensemble of subspace discriminants (ESD) [3] yields robust and competitive methods for face recognition on the challenging FG-NET [4] and MORPH [5] aging datasets.

The outline of the paper is as follows. Section 2 reviews related work on the biometrics of face aging. Section 3 provides background on CNN. Section 4 presents the overall biometric architecture proposed for face aging subject to time lapse. This includes preprocessing, feature extraction using CNN and classification (for the purpose of identification). Section 5 reports on experimental design, datasets used, and results obtained. Section 6 provides a comparative performance evaluation of our CNN-inspired method and current face aging methods. Section 7 discusses the motivation behind CNN, impact of our novel approach, and directions for future R&D on face aging. Section 8 concludes the paper.

2. Related Work

Existing methods for face recognition across time lapse can be divided into two main groups, generative and discriminative. In the generative approach, the system leverages an aging model that can simulate and synthesize the face appearance at various ages. The better the simulation, the better the recognition. Learning such a model can be challenging due to the dependence of the aging process on several uncontrolled external factors. On the other hand, discriminative methods seek to match images directly for authentication without the intermediary step of synthesizing faces, as would be the case in generative methods.

2.1. Generative Methods

Early research on face recognition across time lapse has been reported by Cootes [6]-[8] and involved developing a generative statistical model to simulate aging effects. Cootes *et al.* have found that the aging process is best modeled by a second order polynomial and that a single statistical model cannot be applied to all faces since the aging process is influenced by many factors such as health, gender, and lifestyle. Consequently, they refined their model and developed aging functions that take into consideration the lifestyle of the subject. Wang *et al.* [9] proposed a method similar to Cootes *et al.* where they learn an aging function augmented with aging way classification that characterizes how each subject ages. Once the age is estimated, the facial image of the subject can be synthesized at a different target age. Their method is based on Principal Component Analysis (PCA) and the results reported show that the use of age simulation models yields significant improvements in recognition accuracy compared to recognition without age simulation. Park *et al.* [8] use 3D morphable models to learn the aging patterns of shape and its corresponding texture. Since acquiring 3D scans is not practical in an operational scenario, they adapted their 3D models directly from the 2D FG-NET image dataset. While their 3D model provided better representation of the craniofacial features, such as better localization of muscle fibers and genera-

tion of wrinkles, the results obtained were similar to those achieved using only 2D models.

2.2. Discriminative Methods

Biswas *et al.* [10] have proposed a discriminative approach based on the observation that facial appearance changes in a coherent manner and therefore matching can be done by analyzing the coherency of the drift of feature vectors across age progression. They proposed a simple metric to measure the drift between pairs of images. If two images are of the same subject, then the drift will be small and coherent; if they are of different subjects, then the drift will be extreme and incoherent. They measured drift coherency in both children and adults and presented promising preliminary results. Ling *et al.* [11] used discriminant image descriptors based on Gradient Orientation Pyramids (GOP). They found that GOP representation combined with SVM classification is more invariant to aging effects compared to other methods, and that the difficulty involved in face verification saturates for age gaps between 4 years and 10 years. Klare *et al.* [12] used a discriminative approach to show that training to improve performance in the presence of aging has a negative effect on performance in a non-aging scenario. They suggested that in order to improve performance on a given time lapse, the system should be trained using samples from that specific time lapse. They adopted a technique based on Random LDA subspaces (RS-LDA) [13].

3. Convolutional Neural Networks

Convolutional neural networks (CNN) [2] expand on traditional neural networks by including both fully-connected hidden layers and locally-connected convolutional layers. In the traditional neural network, each hidden layer node is fully-connected to all nodes in the preceding layer. For large raw input images, full-connectivity becomes computationally intensive and does not take advantage of the local correlations in natural images. Inspired by the biology of human vision, convolutional neural networks take advantage of the spatial correlation in natural images and restrict the connectivity of each node to a local area known as its receptive field. In addition, convolutional neural networks use parameter sharing, pooling, and dropout to greatly reduce the number of parameters ("features") learned by the CNN.

In spite of the efficient architecture of convolutional neural networks, they still require a large number of images for training. Using a small dataset for training leads to over-fitting. This requirement restricts the use of convolutional neural networks for classification purposes to large datasets only. However, recent research has shown that the output of the locally-connected convolutional layers produces highly discriminative descriptors. Consequently, a convolutional neural network pre-trained on a large image dataset can be used as a feature extractor for other closely related datasets. Reusing a learned model in another domain or task is known as transfer learning.

Several CNN architectures have been proposed in the literature and some have been shown to produce better results than the most advanced state-of-the-art recognition methods. In our work, we use face descriptors based on the VGG-Face [2] deep architecture described in the next section.

VGG-Face

The VGG-Face descriptors are based on the VGG-Very-Deep-16 CNN architecture described in [2]. The network is composed of a sequence of convolutional, pool, and fully-connected (FC) layers. The convolutional layers use filters of dimension 3 while the pool layers perform subsampling with a factor of 2. The architecture of the VGG-Face network is shown in **Figure 1**.

Table 1 provides additional details on the CNN layers. The volume column represents the width, height, and depth or each layer, respectively. The parameters column shows the number of parameters learned in each layer.

In our experiments, we use a pre-trained implementation of the VGG-Face CNN. The pre-trained CNN was learned from a large face dataset containing 982,803 web images of 2622 celebrities and public figures. While the pre-trained VGG-Face CNN can only identify the subjects in its training dataset, it can however be used as a feature extractor for any arbitrary face image by running the image through the entire network, then extracting the output of the first fully-connected layer, FC-1 shown in **Table 1**. The extracted feature is a highly discriminative, compact, and interoperable encoding of the input image. Once the features are extracted from the FC-1 layer of the VGG-Face CNN, they can be used for training and testing arbitrary face classifiers as will be shown in the next section. We use the MatConvNet toolbox [14], which consists of a library of MATLAB functions
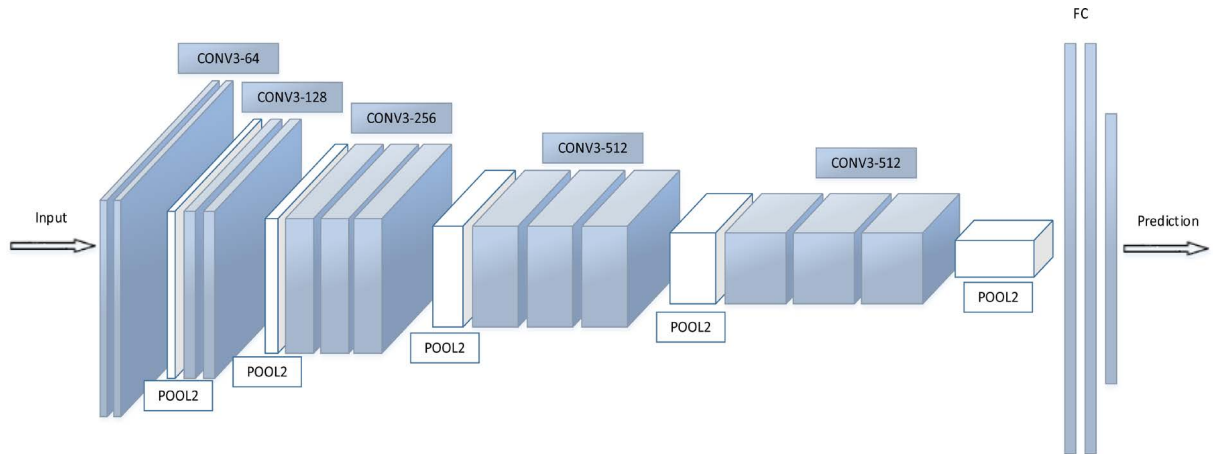
Figure 1. VGG-Face CNN architecture.

Table 1. VGG-Face CNN layers.

Layer (1-11)	Volume	Parameters	Layer (12-22)	Volume	Parameters
INPUT	$224 \times 224 \times 3$	0	CONV3-512	$28 \times 28 \times 512$	$(3 * 3 * 256) * 512 = 1,179,648$
CONV3-64	$224 \times 224 \times 64$	$(3 * 3 * 3) * 64 = 1,728$	CONV3-512	$28 \times 28 \times 512$	$(3 * 3 * 512) * 512 = 2,359,296$
CONV3-64	$224 \times 224 \times 64$	$(3 * 3 * 64) * 64 = 36,864$	CONV3-512	$28 \times 28 \times 512$	$(3 * 3 * 512) * 512 = 2,359,296$
POOL2	$112 \times 112 \times 64$	0	POOL2	$14 \times 14 \times 512$	0
CONV3-128	$112 \times 112 \times 128$	$(3 * 3 * 64) * 128 = 73,728$	CONV3-512	$14 \times 14 \times 512$	$(3 * 3 * 512) * 512 = 2,359,296$
CONV3-128	$112 \times 112 \times 128$	$(3 * 3 * 128) * 128 = 147,456$	CONV3-512	$14 \times 14 \times 512$	$(3 * 3 * 512) * 512 = 2,359,296$
POOL2	$56 \times 56 \times 128$	0	CONV3-512	$14 \times 14 \times 512$	$(3 * 3 * 512) * 512 = 2,359,296$
CONV3-256	$56 \times 56 \times 256$	$(3 * 3 * 128) * 256 = 294,912$	POOL2	$7 \times 7 \times 512$	0
CONV3-256	$56 \times 56 \times 256$	$(3 * 3 * 256) * 256 = 589,824$	FC-1	$1 \times 1 \times 4096$	$7 * 7 * 512 * 4096 = 102,760,448$
CONV3-256	$56 \times 56 \times 256$	$(3 * 3 * 256) * 256 = 589,824$	FC-2	$1 \times 1 \times 4096$	$4096 * 4096 = 16,777,216$
POOL2	$28 \times 28 \times 256$	0	FC-3	$1 \times 1 \times 2622$	$4096 * 2622 = 10,739,712$

implementing CNN architectures for computer vision applications. MatConvNet provides the pre-trained implementation of the VGG-Face CNN that we use for feature extraction.

4. Face Recognition Using Convolutional Neural Networks

We describe here the modular face aging recognition approach in terms of image preprocessing, feature extraction using CNN (see **Figure 1** for the VGG-Face CNN architecture), and classification functional blocks (see **Figure 2** for the architecture of our face aging recognition system). Image preprocessing is about face detection and subsequent image normalization in terms of pose and image size. Feature extraction is about capturing face descriptors using CNN rather than hand-crafted features as current face aging methods do. The classification methods used include nearest neighbor (NN), linear discriminant analysis (LDA), and ensemble of subspace discriminants (ESD). Dimensionality reduction for efficiency reasons using principal component analysis (PCA) is done prior to classification using NN and LDA, but not for ESD as PCA was found to degrade its performance.

4.1. Image Preprocessing

All images are normalized using in-plane rotation to horizontally align the left and right eyes. For FG-NET, the eye coordinates are available from the metadata provided with the dataset. For MORPH, the eyes are localized using an automatic eye detector based on the Viola-Jones object detection algorithm [15]. Face images are cropped and rescaled to a standard 224×224 dimension. The datasets contain a mix of grayscale and RGB im-
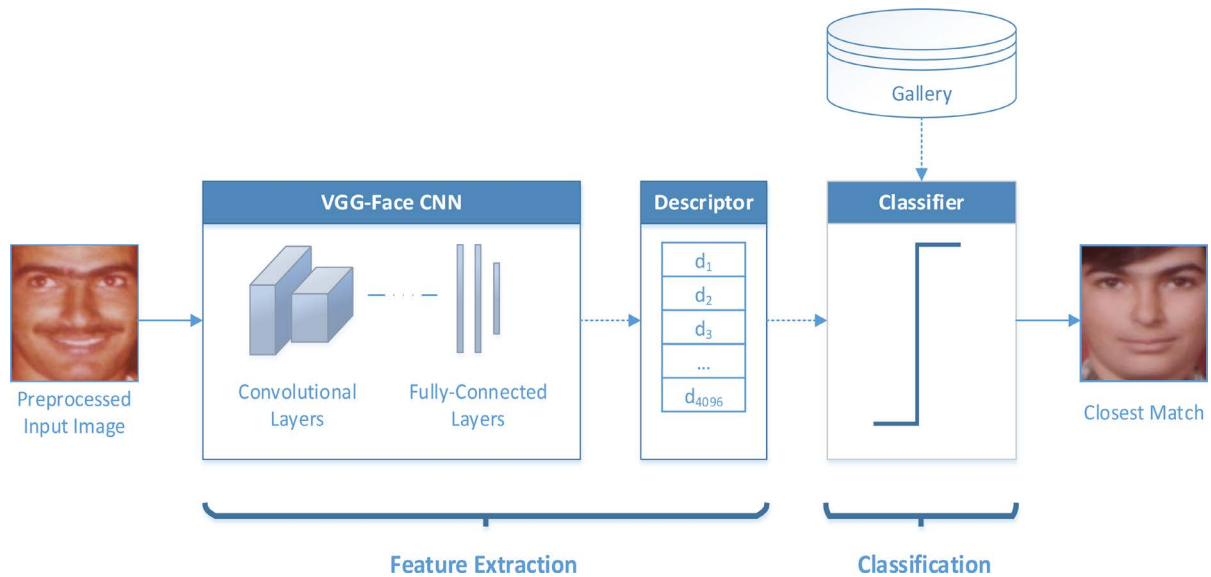
Figure 2. Authentication architecture using VGG-Face CNN feature descriptors.

ages. No additional gray scaling is performed and images are input to the convolutional neural network in their original color channels. A sample of preprocessed images for FG-NET and MORPH is shown in **Figure 3** and **Figure 4**.

4.2. Feature Extraction

We use the VGG-Face network architecture provided by the MatConvNet toolbox [14] for feature extraction. The VGG-Face network described in section 3.1 has a deep architecture composed of 3×3 convolution layers, 2 $\times 2$ pooling layers, and 3 fully-connected layers. While the network can perform classification on its own, the output layer of the network is not used in our experiments and 4096-dimensional descriptors are instead extracted from the FC-1 layer shown in **Table 1**. To extract features from an image, the image is preprocessed and fed to the CNN as a multidimensional array of pixel intensities. For RGB images, the input is a $224 \times 224 \times 3$ array, while for grayscale images the input is a $224 \times 224 \times 1$ array. Each convolutional layer performs a filtering operation on the preceding layer resulting in an activation volume which in turn becomes the input of the following layer. Pooling is used throughout the network to reduce the number of nodes by down sampling the activation maps using the max operation. The fully-connected layers of the network are used for learning the classification function. The feature descriptors are extracted from the output of the first fully-connected layers (FC-1), then L_2-normalized by dividing each component by the L_2-norm of the feature vector. The normalized features are then used for training and testing.

4.3. Classification Methods

The performance and interoperability of our authentication ("identification") method is evaluated across a suite of classifiers, which consists of nearest neighbor (NN) [16], linear discriminant analysis [17], and ensemble of subspace discriminant [3] classifiers. The nearest neighbor classifier uses the 1-nearest neighbor rule and Euclidean distances. To improve NN performance, we apply principal component analysis (PCA) (see Section 4.3.1) to the extracted features. Linear discriminant analysis (LDA) assumes that images for each subject are drawn from Gaussian distributions and the training data is used to learn linear boundaries that discriminate between human subjects. We again apply PCA to the extracted features prior to LDA to improve performance. The subspace discriminant method employs an ensemble of 200 weak learners based on decision trees.

Principal Component Analysis (PCA)
PCA is used to project the CNN features to a lower dimensional subspace [18] in order to improve the performance of the nearest neighbor and linear discriminant analysis classifiers. Given n CNN feature vectors X_1,

$X_2, \ldots, X_n$, one constructs the matrix $X = [X_1, X_2, \ldots, X_n]$ where each X_i is a 4096-dimension FC-1 feature vector. One computes then the mean vector M and subtracts it from the vectors X_i such that $X_i \rightarrow X_i$ - M. One then computes the covariance matrix $C = X^T X$ where X^T is the transpose of X. Finally, one computes the eigenvectors of C and selects only the top eigenvectors corresponding to the largest eigenvalues while preserving 95% of the data variability. The top eigenvectors are used to project the CNN features to the low-dimensional PCA subspace.

5. Experimental Results

We evaluate the performance of the proposed approach in the identification of subjects where there is a significant time lapse between the enrolled and probe images. For each subject, a subset of the corresponding images is used for enrollment while the remaining subset contains the probe images used for testing. In the case of linear and subspace discriminant models, the identity of each subject in the probe dataset is determined directly by the classification model. In the nearest neighbor classifier, probe images are assigned the identity of their respective nearest neighbors based on Euclidean distances. This section details the biometric datasets used, the performance of CNN descriptors and the classifier methods described earlier (see Sect. 4.3) for face recognition across time lapse, and as a unique characteristic, the ability for interoperability on the mix of FG-NET and MORPH.

5.1. Time Lapse Biometric Datasets

We use the FG-NET [4] (see **Figure 3**) and MORPH [5] (see **Figure 4**) datasets, which are publicly available. These datasets contain multiple images per subject reflecting variability in age, in addition to other variability such as pose, illumination and expression. Image normalization and preprocessing includes in-plane rotation and scale. FG-NET contains 1002 images of 82 subjects where subjects' ages vary between 0 and 69. MORPH contains 55,134 images of 13,000 subjects collected over four years. CNN descriptors were extracted from the datasets and used for both training and testing.

5.2. Time Lapse Biometric Performance on FG-NET

After image preprocessing as described in section 4.1, we extract 4096-dimensional feature vectors from each of the 1002 images in FG-NET, then train the proposed suite of classifiers. We use 5-fold cross-validation such that for each fold, 80% of the images of each subject are used for training and 20% are used for testing. The performance reported is the average performance found for each fold. **Figure 5** shows the recognition performance ("rank-1 identification") of each of the classifiers described in Section 4.3. The best performance is 80.6% and was achieved using an ensemble of subspace discriminant learners (ESD) [3]. No PCA dimensionality reduction was applied to the ESD learner in order to achieve maximum performance. One of our experiments showed that PCA has a negative effect on the ESD learner performance due to the loss of information. The performance of the linear classifier was slightly lower than ESD, however, it has the additional benefit of requiring significantly less time to train.

5.3. Time Lapse Biometric Performance on MORPH

We repeat the same experimental setting as in the previous subsection (for FG-NET) using a random subset of 1002 images from MORPH corresponding to 267 subjects. The results are shown in **Figure 6**. The best performance is 92.2% and was achieved again using the ensemble of subspace discriminant classifiers. The performance on MORPH was better than FG-NET due to lower intra-class variability in the dataset.

5.4. Time Lapse Interoperability on Mixed (FG-NET and MORPH) Dataset

Using the same experimental setup as for FG-NET and MORPH, we construct a dataset composed of a 50/50 mix of images from the 2 datasets. Such biometric functionality and corresponding experiments is novel in nature and is thus more important for interoperability purposes. The results obtained show good cross-modal generalization and support a high-level of interoperability. **Figure 7** shows recognition performance for each classifier. The best performance is 86.9% and was achieved again using an ensemble of subspace discriminant learners [3].

Figure 3. FG-NET images [4].

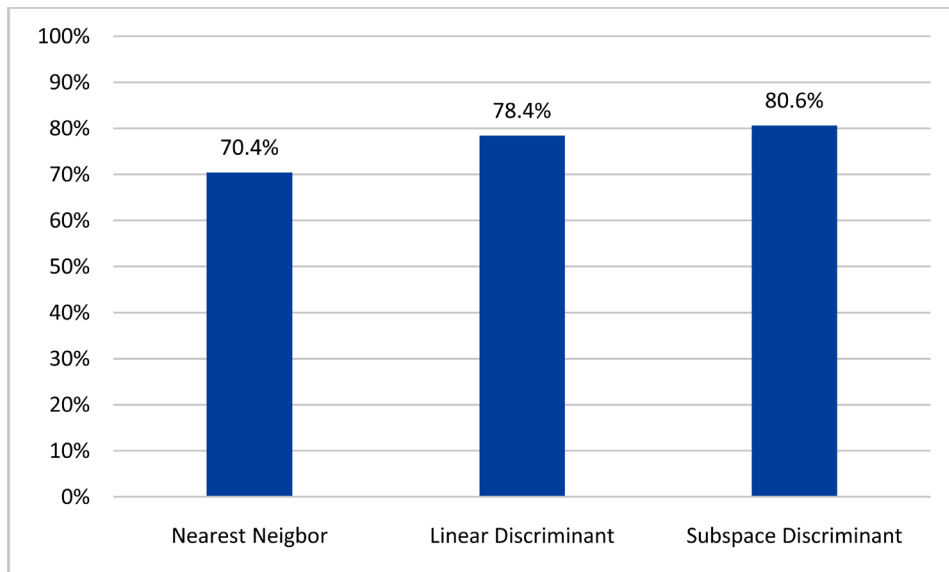**Figure 4.** MORPH images [5].

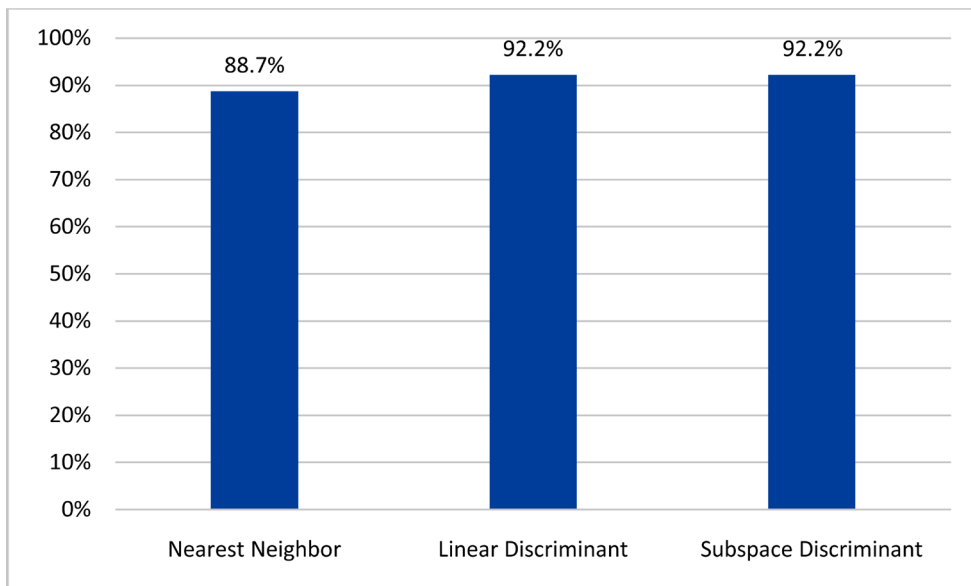**Figure 5.** Time lapse performance on FG-NET using various classification methods.

Figure 6. Time lapse performance on MORPH using various classification methods.
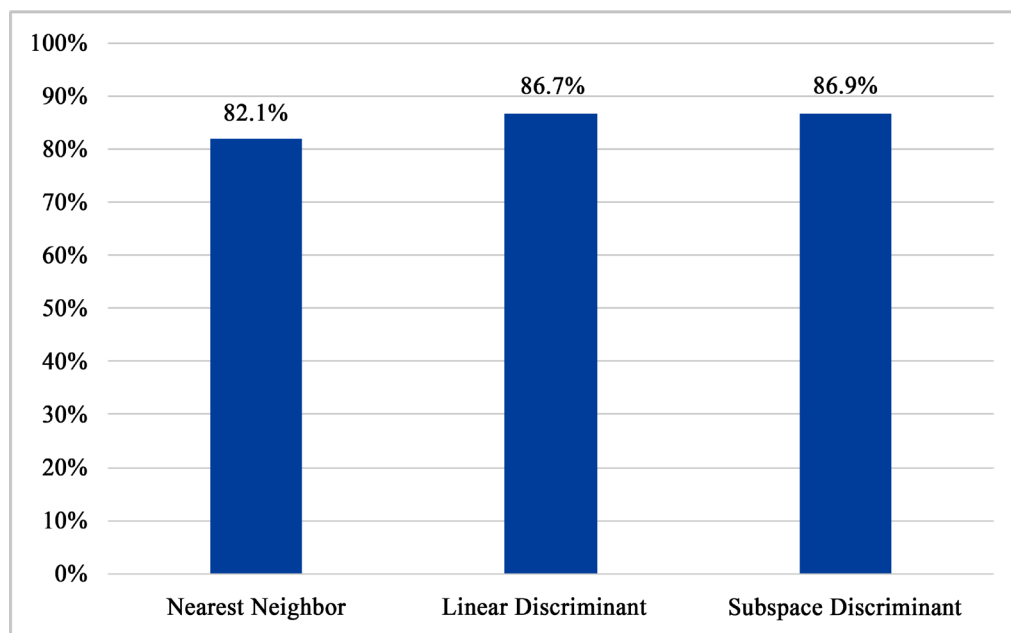
Figure 7. Time lapse performance on mixed (FG-NET and MORPH) dataset using various classification methods.

6. Comparative Time Lapse Performance Evaluation

Table 2 summarizes the performance results of the 3 classification methods on FG-NET, MORPH, and the mixed (FG-NET and MORPH) dataset. The ensemble of subspace discriminants provides the most robust classification method across all datasets. Despite their simplicity, the nearest neighbor and linear discriminant classifiers achieve reasonably good performance in the presence of variability due to aging and other factors such as pose, illumination and expression.

Figure 8 shows the performance of competing methods on the FG-NET dataset. A description of each method is provided in **Table 3**. It can be observed that our approach consistently outperforms state-of-art methods including commercial face recognition engines such as FaceVACS commercial face recognition engine. This demonstrates the effectiveness and interoperability of automatic CNN-based feature extraction across datasets and classification methods.

7. Motivation and Impact and Directions for Future Research

The use of robust age-invariant recognition is important for biometrics-based security applications. Our method uses deep learning for the automatic extraction of CNN features and shows significant improvement in face recognition across time lapse. Age-invariant methods can further reduce the overall operational cost of biometrics security systems by minimizing the need for reenrollments due to time lapse.

Our methodology is holistic where the face image is considered as a single component rather than a collection of semantically separate parts, e.g., eyes, nose, and mouth. Recent research in parts-based recognition has shown that partitioning the face into distinct components can lead to performance improvement. Future work combining CNN features and recognition-by-parts can significantly enhance performance. Towards that end, one should perform visualization driven ablation and sensitivity studies [26] to assess the relative importance of different parts and intermediate CNN layers.

Traditional classification methods assume that the distribution of training and test data remains unchanged. However, in real-world face recognition subject to time lapse, the distributions of samples used for training and testing may be quite different. This challenge is known as covariate shift and is inherent in face recognition systems operating in uncontrolled settings in the wild. An interesting direction for future research is to design CNN driven robust classification methods that can cope with covariate shift and minimize the generalization error. Again visualization driven ablation and sensitivity studies [26] can overcome covariance shift using principled importance weighting.

Table 2. Time lapse performance on FG-NET and MORPH using various classification methods.

	FG-NET	MORPH	FG-NET and MORPH
Nearest Neighbor	70.4%	88.7%	82.1%
Linear Discriminant	78.4%	92.2%	86.7%
Subspace Discriminant	80.6%	92.2%	86.9%

Table 3. Overview of competing time lapse methods on FG-NET.

Publication	Feature Extraction Method	Classification Method	Rank-1
Mahalingam *et al.* (2010) [19]	Feature points extracted using Local Feature Analysis (LFA)	Graph matching	50%
	LBP		42%
Park *et al.* (2010) [20]	Face VACS Commercial FR Engine		38%
Li *et al.* (2011) [21]	SIFT, MLBP	Multi-feature Discriminant Analysis	47.5%
	Face VACS Commercial FR Engine		26%
Mahalingam *et al.* (2011) [22]	Gabor phase and magnitude encoded through Local Gabor Binary Pattern (LGBP) histograms	Histogram intersection	34%
Yadav *et al.* (2013) [23]	LBP using oldest images	Bacteria foraging fusion algorithm	64.5%
	LBP using youngest images		31.2%
Wang *et al.* (2013) [24]	C1-S texture and shape features based on HMAX model	Nearest neighbor with Mahalanobis distance	33.9%
Yang *et al.* (2014) [25]	Local Gabor Binary Pattern Histogram Sequence (LGBPHS)	Factorized Graph Matching	64.47%
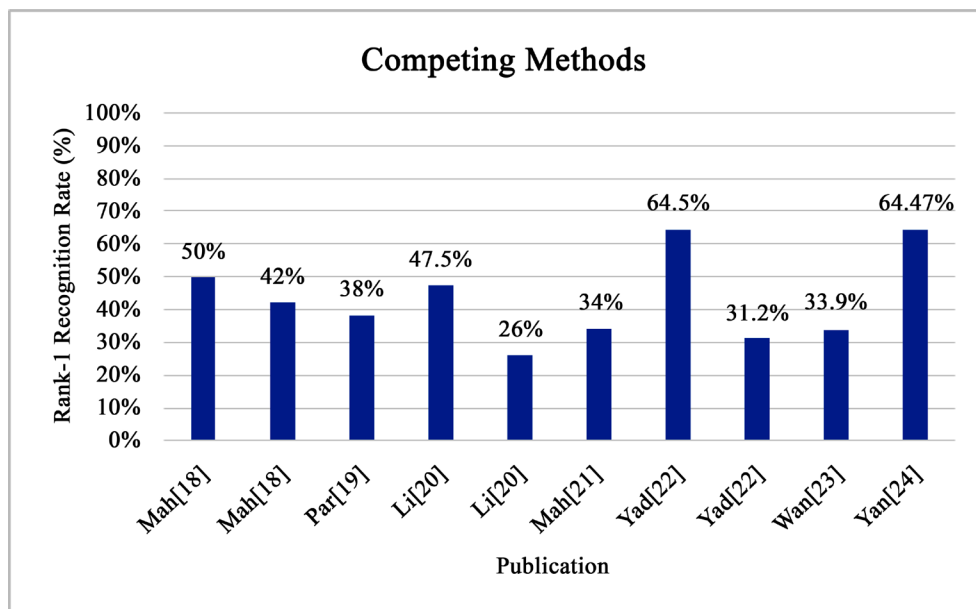
Figure 8. Comparative time lapse performance of competing methods on FG-NET.

8. Conclusion

This paper advances a novel approach for age-invariant face recognition using automatic, highly discriminative and interoperable deep learning driven CNN descriptors across both single source and mixed (multiple source) biometric datasets. The paper illustrates the feasibility and utility of a pre-trained CNN for automatic (rather than hand-crafted) feature extraction yielding performance comparable to state-of-the-art methods on the challenging FG-NET and MORPH datasets. Additional merit of the deep learning/CNN methodology comes from

the interoperability aspect where performance holds steady around 80% - 90% across single source and mixed datasets. Future venues for R&D include the coupling of domain adaptation and transfer learning, on one side and recognition-by-parts rather than holistic biometric authentication, on the other side. Recognition-by-parts will serve as the counterpart to subspace discrimination, weak learners, and ensemble methods.

References

[1] El Khiyari, H., Abate, A.F., De Marsico, M. and Wechsler, H. (2012) Biometric Interoperability across Training, Enrollment, and Testing for Face Authentication. *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications* (*BIOMS*), Salerno, 14 September 2012, 1-8. http://dx.doi.org/10.1109/BIOMS.2012.6345777

[2] Parkhi, O.M., Vedaldi, A. and Zisserman, A. (2015) Deep Face Recognition. *Proceedings of the British Machine Vision Conference* (*BMVC*), Swansea, 7-10 September 2015.

[3] Ho, T.K. (1998) The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 832-844.

[4] Face and Gesture Recognition Working Group (2000) http://www-prima.inrialpes.fr/FGnet/

[5] Ricanek Jr., K. and Tesafaye, T. (2006) MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. *IEEE* 7*th International Conference on Automatic Face and Gesture Recognition*, Southampton, 2-6 April 2006, 341-345.

[6] Lanitis, A., Taylor, C.J. and Cootes, T.F. (1999) Modeling the Process of Ageing in Face Images. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, **1**, 31-136. http://dx.doi.org/10.1109/iccv.1999.791208

[7] Lanitis, A. and Taylor, C.J. (2000) Towards Automatic Face Identification Robust to Ageing Variation. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, 28-30 March 2000, 391-396.

[8] Lanitis, A., Taylor, C.J. and Cootes, T.F. (2002) Toward Automatic Simulation of Aging Effects on Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 442-455. http://dx.doi.org/10.1109/34.993553

[9] Wang, J., Shang, Y., Su, G. and Lin, X. (2006) Age Simulation for Face Recognition. 18*th Int. Conf. Pattern Recognition* (*ICPR*'06), **3**, 913-916. http://dx.doi.org/10.1109/ICPR.2006.230

[10] Biswas, S., Aggarwal, G., Ramanathan, N. and Chellappa, R. (2008) A Non-Generative Approach for Face Recognition across Aging. 2*nd IEEE International Conference on Biometrics*: *Theory*, *Applications and Systems*, Arlington, 29 September-1 October 2008, 1-6. http://dx.doi.org/10.1109/btas.2008.4699331

[11] Ling, H., Soatto, S., Ramanathan, N. and Jacobs, D.W. (2010) Face Verification across Age Progression Using Discriminative Methods. *IEEE Transactions on Information Forensics and Security*, **5**, 82-91. http://dx.doi.org/10.1109/TIFS.2009.2038751

[12] Klare, B. and Jain, A.K. (2011) Face Recognition across Time Lapse: On Learning Feature Subspaces. 2011 *International Joint Conference on Biometrics* (*IJCB*), Washington DC, 11-13 October 2011, 1-8. http://dx.doi.org/10.1109/ijcb.2011.6117547

[13] Wang, X. and Tang, X. (2006) Random Sampling for Subspace Face Recognition. *International Journal of Computer Vision*, **70**, 91-104. http://dx.doi.org/10.1007/s11263-006-8098-z

[14] Vedaldi, A. and Lenc, K. (2015) MatConvNet: Convolutional Neural Networks for MATLAB. *Proceedings of the* 23*rd ACM International Conference on Multimedia*, Brisbane, 26-30 October 2015, 689-692. http://dx.doi.org/10.1145/2733373.2807412

[15] Fasel, I., Fortenberry, B. and Movellan, J.R. (2005) A Generative Framework for Real-Time Object Detection and Classification. *Journal of Computer Vision and Image Understanding*, *Special Issue on Eye Detection and Tracking*, **98**, 182-210. http://dx.doi.org/10.1016/j.cviu.2004.07.014

[16] Cover, T. and Hart, P. (1967) Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, **13**, 21-27. http://dx.doi.org/10.1109/TIT.1967.1053964

[17] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D. (1997) Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, 711-720. http://dx.doi.org/10.1109/34.598228

[18] Jolliffe, I.T. (2002) Principal Component Analysis. 2nd Edition, Springer, New York.

[19] Mahalingam, G. and Kambhamettu, C. (2010) Age Invariant Face Recognition Using Graph Matching. 4*th IEEE International Conference on Biometrics: Theory Applications and Systems* (*BTAS*), Washington DC, 27-29 September 2010, 1-7. http://dx.doi.org/10.1109/btas.2010.5634496

[20] Park, U., Tong, Y. and Jain, A.K. (2010) Age-Invariant Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32**, 947-954. http://dx.doi.org/10.1109/btas.2010.5634496

[21] Li, Z., Park, U. and Jain, A.K. (2011) A Discriminative Model for Age Invariant Face Recognition. *IEEE Transactions on Information Forensics and Security*, **6**, 1028-1037. http://dx.doi.org/10.1109/TIFS.2011.2156787

[22] Mahalingam, G. and Kambhamettu, C. (2011) Can Discriminative Cues Aid Face Recognition across Age? *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, Santa Barbara, 21-25 March 2011, 206-212.

[23] Yadav, D., Vatsa, M., Singh, R. and Tistarelli, M. (2013) Bacteria Foraging Fusion for Face Recognition across Age Progression. *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (CVPRW), Portland, 23-28 June 2013, 173-179. http://dx.doi.org/10.1109/cvprw.2013.33

[24] Wang, S., Xia, X., Huang, Y. and Le, J. (2013) Biologically-Inspired Aging Face Recognition Using C1 and Shape Features. 5*th International Conference on Intelligent Human-Machine Systems and Cybernetics* (*IHMSC*), **2**, 574-577. http://dx.doi.org/10.1109/ihmsc.2013.285

[25] Yang, H., Huang, D. and Wang, Y. (2014) Age Invariant Face Recognition Based on Texture Embedded Discriminative Graph Model. *IEEE International Joint Conference on Biometrics* (*IJCB*), Clearwater, 29 September-2 October 2014, 1-8.

[26] Zeiler, M.D. and Fergus, R. (2014) Visualizing and Understanding Convolutional Networks. In: Fleet, D., Pajdla, T., Schiele, B. and Tuytelaars, T., Eds., *Computer Vision—ECCV* 2014, Springer, Zurich, 818-833. http://dx.doi.org/10.1007/978-3-319-10590-1_53