

Face Recognition Across Time Lapse: On Learning Feature Subspaces

Brendan Klare and Anil K. Jain*

Dept. of Computer Science and Engineering
Michigan State University
East Lansing, MI, U.S.A.

{klarebre, jain}@cse.msu.edu

Abstract

There is a growing interest in understanding the impact of aging on face recognition performance, as well as designing recognition algorithms that are mostly invariant to temporal changes. While some success has been made on this front, a fundamental questions has yet to be answered: do face recognition systems that compensate for the effects of aging compromise recognition performance for faces that have not undergone any aging? The studies in this paper help confirm that age invariant systems do seem to decrease performance in non-aging scenarios. This is demonstrated by performing training experiments on the largest face aging dataset studied in the literature to date (over 200,000 images from roughly 64,000 subjects). Further experiments conducted in this research help demonstrate the impact of aging on two leading commercial face recognition systems. We also determine the regions of the face that remain the most stable over time.

1. Introduction

Face recognition technology is being used in a number of applications, including de-duplication of identification cards, and verification of prisoner identities [6]. A major factor for the success of face recognition in these scenarios is the ability to mostly constrain factors known to be detrimental to face recognition performance, such as facial pose and expression, illumination, and background. However, another variate that is known to greatly impact face recognition performance is the alteration in facial structure and appearance that occurs through the aging process [20].

Unlike the pose, expression, and illumination, aging factors cannot be constrained in order to improve face recognition performance. For example, many years may pass before a released prisoner is recidivate, resulting in a large

time lapse between the mug shot image in the gallery and the current booking image (probe). Similarly, a U.S. passport is valid for ten years, and most state driver's licenses only need to be renewed every five to ten years. Thus, in many critical applications the success of face recognition technology may be impacted by the large time lapse between a probe image and its true mate in the gallery.

Over the past five years, there has been a growing interest in understanding the impact of aging on face recognition performance and proposing solutions to mitigate any negative impact from aging. A major contributor to these advances has been the availability of the MORPH database by Ricanek et al. [21]. The MORPH database consists of two albums, which, in total, contains roughly 100,000 images of about 25,000 subjects. The MORPH dataset has facilitated studies on synthetic aging [16, 22], age invariant face recognition [11, 12, 16], age estimation [5], and aging analysis [17]. A broader examination of facial aging methods in the literature can be found in the summary provided by Ramanathan et al. [19].

Various approaches for improving face recognition performance in the presence of aging can be dichotomized into two groups. The first contains generative synthesis methods which seek to learn an aging model that can estimate the appearance of an aged face from an input image. While these methods have shown some success in mimicking the aging process [16, 22], generative methods are challenging due to the large number of parameters that must be estimated. Synthesis methods also rely on the appearance of the face in order to simulate the aging process, and the results of these methods can suffer from the minor pose and illumination variations that are encountered in large datasets. Further, synthesis methods do not handle the problem of face recognition and need a separate face engine to perform matching. Of course, this also speaks to one of the advantages of synthetic aging methods: they can be easily integrated with existing face recognition engines.

An alternative solution to improving face recognition performance across time lapse is through discriminative

*A.K. Jain is also with the Dept. of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea.

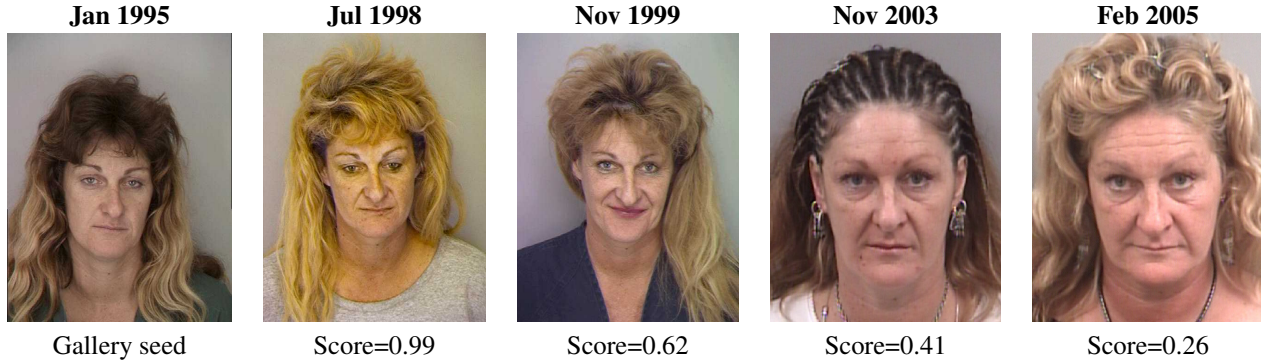


Figure 1. Multiple images of the same subject are shown, along with the match score (obtained by a leading face recognition system) between the initial gallery seed and the image acquired after a time lapse. As the time lapse increases, the recognition score decreases. This phenomenon is a common problem in face recognition systems. The work presented in this paper (i) demonstrates this phenomenon on the largest aging dataset to date, and (ii) demonstrates that solutions to improve face recognition performance across large time lapse impact face recognition performance in scenarios without time lapse.

learning methods [11, 18, 14, 12]. Such methods seek to find the weighted combination of features that are most stable across a particular time lapse. Discriminative approaches are able to leverage both the wide range of facial feature representations [10], as well as the family of learning methods in face recognition. Beginning with Belhumeur et al.’s FisherFaces method [3], discriminative learning approaches have been critical to the advancement of face recognition over the past two decades.

Li et al. used a discriminative random subspace method that outperformed a leading commercial face recognition engine on the MOPRH dataset [11]. This work helped demonstrate that a face recognition system could be trained to improve performance in the presence of aging. While these contributions helped advance the state of the art in face recognition in the presence of time lapse, they also raise a new question regarding the design of face recognition systems: *does the learned subspace for face recognition across time lapse impact the face recognition performance in a non-aging scenario?* In other words, while we can improve face recognition performance in the presence of a large time lapse between the probe and gallery images, we do not want to decrease the performance on two images with minimal time lapse.

The contributions of the research presented in this paper are motivated by the need to answer the question posed above. This question is answered by providing the largest study to date on the impact of aging on face recognition performance. Leveraging a dataset of mug shot images from 200,000 subjects, we demonstrate (i) a degradation in face recognition performance from two leading commercial-of-the-shelf (COTS) face recognition systems (FRS) on match sets partitioned by the amount of time lapse occurring between the probe and gallery images, and (ii) training to improve performance on a particular time lapse range impacts

performance on other time lapse ranges. These findings indicate that face recognition systems should update face templates after a certain time interval has passed from the original acquisition date in order to maximize the benefit of age invariant subspaces without impacting face recognition in non-aging scenarios.

The remainder of the paper is outlined as follows. In Section 2, we discuss the face dataset used in this research. In Section 3, we revisit the random subspace framework and discuss how it was adopted for this work. Section 4 presents experiments on the impact of training on age invariant face recognition, as well as the computational demands in undertaking such a large scale study.

2. Dataset

This study leverages a set of 200,000 mug shot images from roughly 64,000 subjects collected in the state of Florida, U.S.A. Each image contains a subject id and an image acquisition date, which enables the time lapse between any two images to be determined. The 200,000 images are a subset of a larger 1.5 million image dataset available to us, and the 200,000 images were selected so that different time lapse ranges were equally represented in this study.

The time lapse ranges analyzed in this study were (i) 0 to 1 year time difference between a probe and gallery image, (ii) 1 to 5 year time difference, (iii) 5 to 10 year difference, and (iv) more than 10 years difference. Training sets for each of the time lapse ranges listed above are generated so that each range has 8,000 subjects. The only exception is the 10+ year time lapse range, where only around 2,000 subjects were available for training. Similarly, test sets were generated to represent each of the time lapse ranges above. For each time lapse range, 12,000 subjects were used for testing. However, the 10+ year time lapse test set contained

only 2,000 subjects. For each subject in the study, their oldest face image was used as the gallery seed image. Multiple probe images that fell within the time lapse range for a subject were often available as well. For example, the 1 to 5 year test set contained 12,000 gallery images and 33,443 probe images, where each probe image was taken between one to five years after the corresponding gallery image.

All parameter validation in this work was performed using the training set. This was done by using the first half of the training set to train on different parameter values and the second half of the training set to determine the optimal parameter combination (with respect to face recognition performance). Thus, the second half of the training set also served as a validation set.

The analysis performed on the dataset is the largest such study to date. Further, because the images are pulled from a larger pool of images, the dataset is also unique in that it controls the time lapse variate so that the same number of subjects are available to analyze 5 to 10 years aging as 0 to 1 year aging (for example). As such, measuring the performance of COTS FRS on this dataset will provide an unprecedented demonstration of how commercially available face recognition technology performs in the presence of aging. Because both Li et al. [11] and Ling et al. [12] have been able to surpass the performance of COTS FRS by performing discriminative learning on face images with time lapse, it is generally accepted that face recognition performance suffers greatly as the time between image acquisition increases.

We analyzed the performance of two commercial-of-the-shelf face recognition systems: Cognitec’s FaceVACS SDK [1], and PittPatt’s Face Recognition SDK [1]. Both matchers were competitive participants in the latest NIST sponsored Multi-Biometrics Evaluation (MBE) [4]. Results in this work list the matchers as “COTS 1” and “COTS 2” in order to make anonymous each matcher’s performance relative to the other.

Figure 2 shows the matching accuracies of the two COTS matchers as a function of the time lapse between the probe and gallery image on the test sets mentioned above. The decrease in performance as the time lapse increases clearly shows the difficulty face recognition systems have with age variation.

3. Random Subspace Face Recognition

In this work, we adopt a random subspace linear discriminant analysis (RS-LDA) algorithm, based on Wang and Tang’s original face subspace method [23]. More recently, Li et al. [11] have used a variant of this approach to improve face recognition in the presence in aging. Klare and Jain have also demonstrated the benefit of RS-LDA on a heterogeneous face recognition scenario [8].

RS-LDA is based on the FisherFace linear discriminant

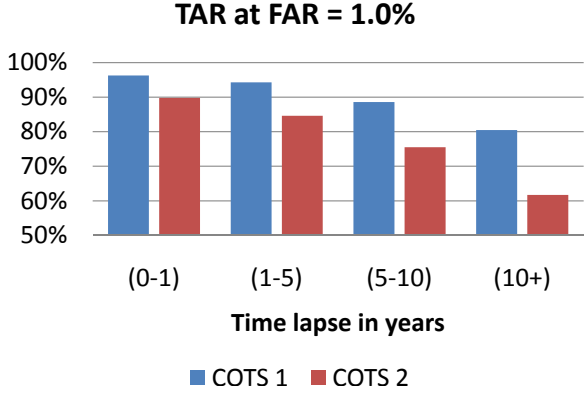


Figure 2. The performance of two commercial face recognition systems as a function of time lapse between probe and gallery images.

analysis algorithm [3], where a linear subspace Ψ is learned from the original feature space by solving the generalized eigenvalue problem $S_b \cdot \Psi = \Lambda \cdot S_w \cdot \Psi$ with the between-class and within-class matrices S_b and S_w built from a set of training images. In RS-LDA, multiple subspaces Ψ_b , $b = 1 \dots B$, are learned using both randomly sampled subsets of the original feature space as well as randomly sampled subjects from the set of training instances. The motivation for using RS-LDA over standard LDA is due to degenerative properties that often manifest in S_w (which must be full rank to solve $S_w^{-1} \cdot S_b$). While Level 2 facial feature representations [10] (such as the local binary patterns [15] used in this work) offer improved recognition accuracies, they also increase the dimensionality of the facial feature vectors. This in turn increases the likelihood that S_w is degenerate, and further necessitates the need for a method such as RS-LDA. Other LDA variants offer solutions to this small sample size problem [13, 7]; however, RS-LDA is preferred due to the ease of implementation and wider range of successful applications in face recognition [23, 11, 8, 9].

The approach used in this work is mostly based on the method by Li et al. [11]; however, we had to modify their method in order to reduce the computational requirements because the number of images handled in this experiment is an order of magnitude larger than their work. Again, the intent of this work is not to provide a method that can improve on commercially available face recognition technology (this capability has already been demonstrated [11, 12]). Instead, we wish to understand how training a face recognition system to improve recognition accuracies on a particular time lapse scenario performs on scenarios with a larger or smaller amount of time lapse than training time lapse.

3.1. Face Representation

We represent face images in this experiment with multi-scale local binary patterns (MLBP), which is the concatenation of local binary patterns [15] of radii 1, 3, 5, and 7. Ahonen et al. first demonstrated the effectiveness of representing face images with LBP descriptors [2].

In order to represent a face with MLBP feature descriptors, the face is first geometrically normalized using the eye locations to (i) perform planar rotation so the angle between the eyes is 0 degrees, (ii) scale the face so the inter-pupillary distance between eyes is 75 pixels, and (iii) crop the face to 250x200 pixels. Once geometrically normalized, MLBP feature descriptors are densely sampled from patches of size 24x24 across the face, with an overlap of 12 pixels. In total, this yields 285 MLBP descriptors representing the face. The size of the patch (24x24) was selected by using the training set to perform parameter validation.

To reduce the total feature vector size, principal component analysis (PCA) was performed on one half of the training set to learn a subspace for each of the 285 MLBP feature sampling locations. The second half of the training set was used to determine the minimum energy variation that needed to be retained without impacting face recognition performance. It was determined that 98% of the variance could be retained without impacting the recognition performance. The original MLBP descriptor is 236 dimensional ($4 \cdot 59$). After PCA dimensionality reduction, the descriptor size, on average, was reduced to 99 dimensions at each of the 285 sampling locations. After the dimensionality of the MLBP descriptor for each face patch was reduced, all descriptors are concatenated together, resulting in a feature vector of dimensionality $d = 28,187$. Without this PCA step, the feature dimensionality would have been 67,260.

3.2. Random Subspaces

B random LDA subspaces Ψ_b are learned from B random samples of the d -dimensional feature space. The eigenvalues corresponding to each feature dimension from the previous PCA step were used to weight the random sampling so that features with higher variation energy will have a higher likelihood of being selected. The benefit of this approach was confirmed by evaluation on the validation set. The number of features sampled with the weighted random sampling was controlled by the parameter ρ ($0 < \rho < 1$) in order to select $d' = \rho \cdot d$ features at each stage $b = 1 \dots B$. Additionally, from the N training subjects available, $N' < N$ were randomly sampled to build the between-class scatter matrix $S_{\text{Btwn}}^b \in \mathbb{R}^{d', d'}$ and the within-class scatter matrix $S_{\text{Wthn}}^b \in \mathbb{R}^{d', d'}$ at each stage b . Finally, we learn the subspace Ψ_b as

$$\Psi_b = \underset{\Psi'}{\operatorname{argmax}} \frac{\|\Psi'^T \cdot S_{\text{Btwn}}^b \cdot \Psi'\|}{\|\Psi'^T \cdot S_{\text{Wthn}}^b \cdot \Psi'\|} \quad (1)$$

After learning the set of B subspaces Ψ_b , $b = 1 \dots B$, a new face image is represented as the concatenation of the each of the B subspace projections. The dissimilarity between two faces is then measured by the L-2 norm distance.

Despite reducing the feature dimensionality and only using a ρ percent of the ($d' = \rho \cdot d$) features, reasonable values of ρ yield a feature vector that is too large to accurately solve Eq. 1. To resolve this, a second PCA step was applied at each stage b to perform feature reduction on the d' dimensional feature vector. This second PCA step was performed by retaining $0 < p < 1$ percent of the variance in the training instances at stage b .

The parameters in the RS-LDA framework are the number of training subjects at each stage (N'), the percentage of features to sample at each stage (ρ), the number of stages (B), and the percentage of variance in the PCA step for each stage (p). Using the training set for validation to find the highest recognition accuracies, the following parameters values were selected: $N' = 300$, $\rho = 0.45$, $B = 20$, and $p = 0.95$.

4. Experiments

Figure 2 shows the negative correlation between face recognition accuracy and the amount of time lapse between probe and gallery image capture. A strong case has been made to handle this problem by training discriminative face recognition systems [11, 12]. Here we will use the random subspace framework developed in Section 3 to understand if training a face recognition system to improve performance on aging impacts the standard face recognition scenarios.

Using the training set splits discussed in Section 2, we trained five different RS-LDA systems using the algorithm presented in Section 3.

- The first RS-LDA system was trained on the 8,000 training subjects with 0 to 1 year time lapse between probe and gallery image.
- The second system was trained on 8,000 subjects with 1 to 5 year time lapse.
- The third system was trained on 8,000 subjects with 5 to 10 year time lapse.
- A fourth system was trained on 2,000 subjects with over 10 years time lapse (only 2,000 subjects were available with this time lapse).
- A final system was trained using 8,000 subjects with time lapse equally distributed amongst the four time lapse splits shown. Thus, this final system was trained on subjects with 0 year time lapse up to 17 years (17 years is the maximum time lapse in the 10+ set).

Test set: 0 to 1 year time lapse							
RS-LDA trained on (time lapse in years):					Baselines:		
(0-1)	(1-5)	(5-10)	(10+)	(All)	MLBP Only	COTS1	COTS2
94.5%	94.1%	93.1%	91.8%	94.1%	71.2%	96.3%	89.8%
# of Match Comparisons:					19,996		
# of Non-Match Comparisons:					239,572,034		
(a)							
Test set: 1 to 5 year time lapse							
RS-LDA trained on (time lapse in years):					Baselines:		
(0-1)	(1-5)	(5-10)	(10+)	(All)	MLBP Only	COTS1	COTS2
90.3%	90.5%	89.1%	87.7%	90.2%	62.9%	94.3%	84.6%
# of Match Comparisons:					33,443		
# of Non-Match Comparisons:					401,282,557		
(b)							
Test set: 5 to 10 year time lapse							
RS-LDA trained on (time lapse in years):					Baselines:		
(0-1)	(1-5)	(5-10)	(10+)	(All)	MLBP Only	COTS1	COTS2
75.2%	81.2%	82.0%	80.4%	81.3%	46.7%	88.6%	75.5%
# of Match Comparisons:					24,036		
# of Non-Match Comparisons:					215,795,208		
(c)							
Test set: 10+ year time lapse							
RS-LDA trained on (time lapse in years):					Baselines:		
(0-1)	(1-5)	(5-10)	(10+)	(All)	MLBP Only	COTS1	COTS2
65.6%	72.2%	72.4%	71.0%	71.2%	39.2%	80.5%	61.7%
# of Match Comparisons:					6,221		
# of Non-Match Comparisons:					12,995,669		
(d)							

Figure 3. The true accept rates at a fixed false accept rate of 1.0% across datasets with different amount of time lapse between the probe and gallery images. A set of RS-LDA subspaces were trained on a separate set of subjects with the different time lapse ranges tested above. In all cases except 10+ year time lapse, the RS-LDA subspaces performed the best when trained on the same amount of time lapse they were tested on. This examination suggests the need for multiple recognition subspaces depending on the time lapse.

Figure 3 shows the recognition accuracy on each of the four test sets using the five trained systems. The first test set (Fig. 3(a)) has 0 to 1 year time lapse between the probe and gallery images for 12,000 subjects. The results show that the best performance from the five trained systems is the system trained on 0 to 1 year time lapse. As the time lapse between the training set and the test set increases, the face recognition accuracy decreases. These results help provide the following answer to the question originally posed: *training a face recognition system to improve on face aging does seem to reduce its performance when facial aging has not occurred.*

The recognition performance on face images that have 1

to 5 years time lapse (Fig. 3(b)) show the best performance from the five RS-LDA systems is the system trained on 1 to 5 year lapse. However, the performance from 0 to 1 years time lapse training is not much lower. In fact, the difference between training and testing on 0 to 1 year and 1 to 5 year is rather minimal. This is likely due the fact that only minor aging changes have occurred in these time spans.

The recognition performance on face images with 5 to 10 years time lapse (Fig. 3(c)) shows how learning can help improve recognition accuracies in the presence of a large amount of aging. The true accept rate improves by nearly 7.0% when trained on the 5 to 10 year set than with the 0 to 1 year training set. Thus, the feature subspaces learned

on data with minimal aging did not generalize well to data with larger amounts of aging.

The recognition results on aging over 10 years (Fig. 3(c)) is the only scenario in which the subspace trained on the same time lapse as tested on did not offer the highest results. However, the 10+ year subspace only had 2,000 subjects to train on while the other subspaces had 8,000 subjects. This could also be explained by the complex nature of face aging that manifests itself in different ways for different individuals, especially when the time lapse is large.

In each testing scenario, the subspace labeled (All) is the one trained on 8,000 subjects exhibiting all time lapse ranges studied here. While this subspace never had the top accuracy with respect to the other RS-LDA subspaces, it consistently performed well on all time lapses. This indicates that mixed training (with a wide set of age differences) is a viable solution when learning multiple subspace models is not reasonable (such as scenarios when image acquisition dates are unavailable).

The performance of COTS1 exceeded the RS-LDA system in each testing scenario. However, the RS-LDA system was purposely designed to be less complex to help facilitate the scope of this study. Including additional features such as the SIFT descriptors and multiple patch sizes that Li et al. used in their aging system [11] would have led to improved performance. Despite this, the validity of the training performed in RS-LDA is clearly established when examining the performance of the RS-LDA subspaces over the baseline MLBP only performance. MLBP only makes use of the initial MLBP feature representation to measure the (dis)similarity between the faces, but does not perform training. Through the use, of RS-LDA the recognition accuracy improves substantially.

The large time lapse dataset with a large number of subjects presented in this study also enabled us to examine which regions of the face remained the most persistent or retained the most discriminative power over time. To examine this stability, we measured the Fisher separability at each patch where the MLBP feature descriptors were computed. For a given face patch, we measured the Fisher separability as the ratio of the sum of eigenvalues from the between-class scatter to the sum of eigenvalues from the within-class scatter. This indicates the inherent separability provided by the Level 2 MLBP features at different regions of the face. These Fisher separability values at different time lapses are shown in Figure 4. The results show that while, as expected, the inherent separability decreases for each facial region as time increases, the mouth region has more discriminative information than the nose region, especially as time passes. This also confirms the discriminative information contained in the region of the face around the eyes. Such information could be useful in explicitly weighting different face regions.



Figure 5. Examples where training on different time lapses than tested on caused failures when matching the probe image (left) with the gallery image (right). (a) The subspace trained on 0 to 1 years time lapse yielded a Rank-1 match, while the subspace trained on 5 to 10 years time lapse yielded a Rank-13 match. (b) The subspace trained on 0 to 1 years time lapse yielded a Rank-881 match, while the subspace trained on 5 to 10 years time lapse yielded a Rank-1 match.

4.1. Computational Demands

Future work will attempt to leverage the additional face images contained in the 1.5 million mug shot image dataset available to us. However, one of the major difficulties we anticipate in this analysis is the computational demands for processing such a wide corpus of data. In this section, we briefly highlight some of the challenges of processing large scale data.

In this study, each of the roughly 120,000 test images used were enrolled in the Cognitec's and PittPatt's FRS. Once each face was enrolled, 869 million match comparisons were performed by each matcher to measure the performance on each time lapse data set.

The analysis of RS-LDA on MLBP feature representation required all 200,000 images in the study. This, in turn, required all images to be geometrically normalized using the eye locations automatically detected by the FaceVACS system. Once the images were aligned, the MLBP feature descriptors were extracted. With a 236-dimensional MLBP descriptor extracted at 285 patches across each face,

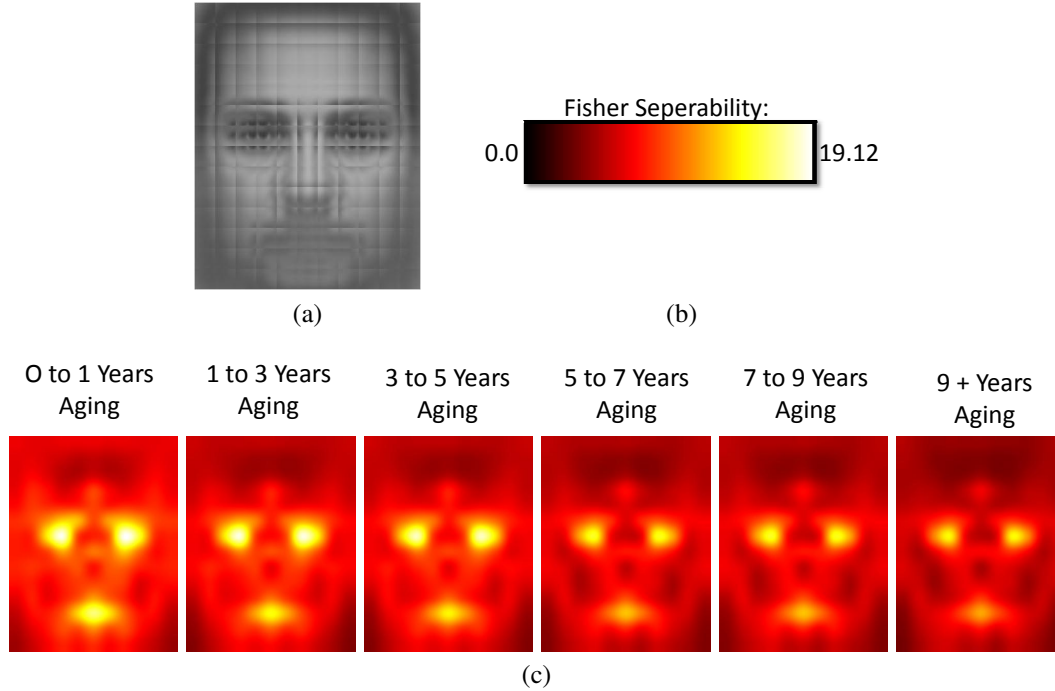


Figure 4. Discriminability of different face regions over time. (a) The mean pixel values at each patch where MLBP feature descriptors are computed. (b) The scale of the Fisher separability criterion. (c) The heat map showing Fisher separability values at each image patch across different time lapses. As time increases, the eyes and mouth regions seem to be the most stable sources of identifiable information.

roughly 48Gb of space was needed for storing these features.

For analyzing the RS-LDA performance on each of the five time lapse training sets, a total of 869 million test set comparisons needed to be performed five times, resulting in a total of 4.34 billion comparisons. Other computational demands arose from the training of the RS-LDA subspaces on sets of 8,000 subjects, performing parameter validation on combinations of the four different parameters¹ in the RS-LDA framework, and generating the ROC curves for each score matrix.

Machines with large amounts of RAM were also required to efficiently process the data. For example, the covariate analysis necessary for RS-LDA needed the MLBP features from all 8,000 subjects to be loaded into main memory. For testing, a major bottleneck occurred when reading from disk the MLBP feature descriptors from each of the 12,000 subjects. This made it necessary to keep the MLBP features in memory while each of the 20 random subspaces were processed (as opposed to releasing the memory as each image was projected into one of the subspaces).

Efficient code design helped overcome some of these computational challenges. However, this study was pri-

¹Recognition accuracies from training on roughly 10,000 subjects and testing on 10,000 subjects was explored on over two hundred parameter combinations.

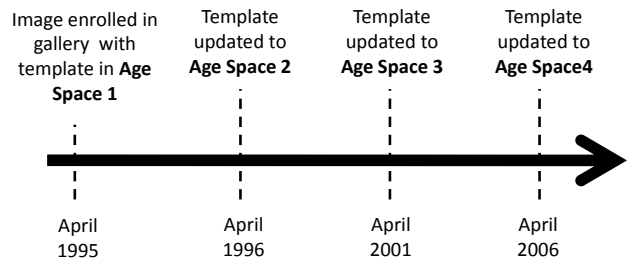


Figure 6. The ability to improve face recognition performance by training on the same time lapse being tested on suggests face recognition systems should update templates over time. For example, at fixed intervals from the original acquisition date the template is updated to reside in a subspaces trained for the time lapse that has occurred since acquisition. Probe images would be projected into each subspace and matched in the subspace corresponding to each gallery image.

marily made possible by Michigan State University's High Performance Computing Center (HPCC), which provides a cloud computing service. Up to 40 different compute nodes, each with over 10gb of RAM, were used at the same time to meet the computational demand of this experiment.

5. Conclusions

This paper presents studies on the largest face aging dataset to date. These results demonstrate that (i) face recognition systems degrade as the time lapse between

probe and gallery faces images, (ii) training to improve face recognition performance in the presence of aging can lower performance in non-aging scenarios, and (iii) the best performance on a particular amount of time lapse is achieved by training a system on that particular time lapse.

The findings presented suggest a periodic update of face templates (see Figure 6). With significant time lapse, updating the face template to reside in a subspace designed to capture the most discriminative features will help improve the recognition performance in the presence of aging without compromising performance in cases where only a minimal amount of aging has occurred.

Our future work will involve expanding the study to incorporate additional subjects and face images. Further, the impact of training subspaces not only on the time lapse, but the subject's age as well (i.e. train separate subspaces for aging in young people and aging in old people) will be investigated.

Acknowledgement

The authors would like to express their gratitude for the strong support they received from Scott McCallum and the Pinellas County Sheriff's Office. The authors would also like to thank Unsang Park and Karl Ricanek for their valuable feedback. This research was partially supported by the U.S. Federal Bureau of Investigation's Biometric Center of Excellence. Anil Jain's research was also partially supported by the World Class University program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008).

References

- [1] FaceVACS Software Developer Kit, Cognitec Systems GmbH, <http://www.cognitec-systems.de>.
- [2] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 28(12):2037–2041, Dec. 2006.
- [3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 19(7):711–720, 1997.
- [4] P. J. Grother, G. W. Quinn, , and P. J. Phillips. Mbe 2010: Report on the evaluation of 2d still-image face recognition algorithms. *National Institute of Standards and Technology, NISTIR, 7709*, 2010.
- [5] G. Guo and G. Mu. Human age estimation: What is the influence across race and gender? In *Proc. of IEEE Conference on Computer Vision & Pattern Recognition*, 2010.
- [6] A. K. Jain, B. Klare, and U. Park. Face recognition: Some challenges in forensics. In *Proc. of Automatic Face and Gesture Recognition*, 2011.
- [7] H. Y. Jie, H. Yu, and J. Yang. A direct LDA algorithm for high-dimensional data – with application to face recognition. *Pattern Recognition*, 34:2067–2070, 2001.
- [8] B. Klare and A. Jain. Heterogeneous face recognition: Matching NIR to visible light images. In *Proc. International Conference on Pattern Recognition*, 2010.
- [9] B. Klare and A. Jain. Heterogeneous face recognition using kernel prototype similarities. *IEEE Trans. Pattern Analysis & Machine Intelligence (under review)*, 2011.
- [10] B. Klare and A. K. Jain. On a taxonomy of facial features. In *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2010.
- [11] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *IEEE Trans. Information Forensics and Security (to appear)*, 2011.
- [12] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs. Face verification across age progression using discriminative methods. *IEEE Trans. Information Forensics and Security*, 5(1):82–91, 2010.
- [13] J. Lu, K. Plataniotis, and A. Venetsanopoulos. Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. *Pattern Recognition Letters*, 26(2):181–191, 2005.
- [14] G. Mahalingam and C. Kambhamettu. Age invariant face recognition using graph matching. In *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2010.
- [15] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 24(7):971–987, 2002.
- [16] U. Park, Y. Tong, and A. Jain. Age-invariant face recognition. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 32(5):947–954, May 2010.
- [17] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King. Aspects of age variation in facial morphology affecting biometrics. In *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems*, 2007.
- [18] N. Ramanathan and R. Chellappa. Face verification across age progression. *IEEE Trans. Image Processing*, 15(11):3349–3361, 2006.
- [19] N. Ramanathan, R. Chellappa, and S. Biswas. Computational methods for modeling facial aging: A survey. *Journal of Visual Languages & Computing*, 20(3):131–144, 2009.
- [20] K. Ricanek, A. Sethuram, E. K. Patterson, A. M. Albert, and E. J. Boone. *Craniofacial Aging*. Wiley, 2008.
- [21] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Proc. of Automatic Face and Gesture Recognition*, 2006.
- [22] J. Suo, S.-C. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Trans. Pattern Analysis & Machine Intelligence*, 32(3):385–401, 2010.
- [23] X. Wang and X. Tang. Random sampling for subspace face recognition. *Int. Journal Computer Vision*, 70(1):91–104, 2006.