Face Recognition through Mismatch Driven Representations of the Face

Simon Lucey Tsuhan Chen

Advanced Multimedia Processing Laboratory, Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh PA 15213, USA
slucey@ieee.org, tsuhan@cmu.edu

Abstract

Performance of face verification systems can be adversely affected by a number of different mismatches (e.g. illumination, expression, alignment, etc.) between gallery and probe images. In this paper, we demonstrate that representations of the face used during the verification process should be driven by their sensitivity to these mismatches. Two representation categories of the face are proposed, parts and reflectance, each motivated by their own properties of invariance and sensitivity to different types of mismatches (i.e. spatial and spectral). We additionally demonstrate that the employment of the sum rule gives approximately equivalent performance to more exotic combination strategies based on support vector machine (SVM) classifiers, without the need for training on a tuning set. Improved performance is demonstrated, with a reduction in false reject rate of over 30% *when compared to the single representation algorithm. Experiments were conducted on a subset of the challenging Face Recognition Grand Challenge (FRGC) v1.0 dataset.*

1. Introduction

The task of verifying the identity of a subject from a single gallery image is an extremely difficult task, considering the number of permutations probe images of the same subject can undergo (e.g. illumination, expression, alignment, etc.). As one can sometimes only have access to a single gallery image (e.g. in a surveillance scenario), one must rely on prior subject independent knowledge of between-class and within-class variation in order to obtain invariance. The form this prior should take is an open question in face recognition at the moment.

In this paper, we propose an approach to face recognition that embraces the use of two types of prior: namely *exemplar* and *representation*. Priors in the form of exemplar learning have found much success in face recognition, particularly approaches based on subspace methods [1, 2, 3, 4, 5]. These techniques approach the problem as trying to find the optimal subspace that maximizes the signal-to-noise (SNR) ratio, or related criteria, so that ratio of between- to within- subject variation is maximized.

In theory, if one had an infinite amount of development[1] examples there would be no need to rely on a representation prior, as the dependencies in the data should provide all necessary discriminative information. However, in practice there is only ever a finite number of development examples making the representation prior in the training of a face recognition system extremely important. In essence, the representation prior acts as a regularization function (although no regularization function in practice is explicitly used) to ensure generalization between the development and evaluation set.

Representation priors are often based on heuristics or the known physics of the problem domain in which the pattern recognition task (i.e. face recognition) is being performed. Some common representation priors/assumptions employed in face recognition are:

Alignment prior: Facial images contain inherent alignment uncertainty, such that smoothing/blurring operations or the synthesis of multiple alignments encourage better matches [6, 7].

Reflectance prior: Some local spatial frequency bands of face images are less prone to mismatch than others (e.g. low-frequency bands contain more illumination variation than higher-frequency bands) [8, 9].

Parts prior: Some local areas/patches of face images are less prone to mismatch than others (e.g. eyes and nose are sometimes less susceptible to mismatch than other areas of the face such as the mouth) [2, 10].

In most circumstances however, it is very difficult to ascertain which (if any) of these assumptions hold for an individual gallery and probe face image comparison. In these

[1] In this paper, we refer to the development set as the set of face images that are used to train the classifier with subject-independent data. The evaluation set is the set of gallery and probe images used during the verification process.

circumstances it is often wise to employ multiple representations of the face to try and "hedge your bets" over which representation will perform best.

The employment of multiple representations has already found its way into face recognition literature of late. Shan et al. [7] demonstrated improved face recognition performance through the concatenation of multiple Gabor filter responses of a single image into a vector. The authors then applied discriminant analysis to a development set of multi-representation vectors, where improved face recognition performance was cited in testing. Similarly, Del la Torre et al. [3] demonstrated improved face recognition in mismatched conditions using up to 150 different representations of the face (e.g. Gabor filter responses, gradient maps, Laplacian maps, etc.). Hitherto, the selection of these representations has been in most cases very adhoc, based only on simple heuristics.

In this paper, we propose that multiple representations of the face should be based on the mismatch one anticipates between the gallery and probe sets. Specifically, we propose two categories of representation namely: *parts* and *reflectance* corresponding to two of our three representational priors/assumptions stated earlier. These two categories of representation can naturally deal with heterogeneous[2] mismatches encountered in the face image both spatially and spectrally. We ignore the representation prior associated with alignment error in this paper as the dataset we conducted our experiments on seemed to suffer minimally from this type of mismatch.

We also propose a strategy in this paper for combining these representations effectively based on previous work in classifier combination [11]. In this work, we demonstrate that a combination strategy based on the *sum* rule performs similarly to more exotic combination functions based on a support vector machine (SVM) classifier, with the added advantage that the sum rule does not require a separate tuning set to find the combination function.

2. Discriminant Analysis

In this paper, we will be constructing subspaces that attempt to maximize,

$$\frac{|\mathbf{V}^T\mathbf{S}_b\mathbf{V}|}{|\mathbf{V}^T\mathbf{S}_w\mathbf{V}|} \tag{1}$$

where $\mathbf{V}$ is the concatenation of column vectors that span the discriminative subspace. Provided one can obtain good estimates of the between $\mathbf{S}_b$ and within $\mathbf{S}_w$ scatter matrices, the solution to $\mathbf{V}$ can be found efficiently through eigen-analysis and simultaneous diagonalization which we shall refer to as linear discriminant analysis (LDA). However, in practice the scatter matrices $\mathbf{S}_b$ and $\mathbf{S}_w$ are usually under ranked due to the "small sample size" (SSS) problem, as the number of training samples is far less than the dimensionality of those samples.

A number of techniques can be employed to circumvent the SSS problem [4], with the two most common techniques being:

- The application of a principal component analysis (PCA) step to ensure that $\mathbf{S}_b$ and $\mathbf{S}_w$ are fully ranked in the reduced space where canonical LDA is then applied; a process we refer to as F-LDA in our results section or *Fisherfaces* in literature [1].
- Through eigen-analysis the principal subspaces of $\mathbf{S}_b$ and $\mathbf{S}_w$ are found directly, where the null space of $\mathbf{S}_w$ is subtracted from the principal space of $\mathbf{S}_b$ before finding its principal subspace. This process, outlined in [12], satisfies the maximization of Equation 1, but has the added advantage of preserving additional discriminatory information that may be discarded in the blind application of PCA in the Fisherface approach. We shall refer to this approach as null-space LDA (N-LDA).

For a baseline we will also be presenting results for the whitened *Eigenfaces* [13] approach, which can be interpreted as the maximization of Equation 1 where one assumes that $\mathbf{S}_b$ is an identity matrix and $\mathbf{S}_w$ is the total scatter matrix of all variation in the training samples. This differs to the canonical Eigenface approach, where the values of $\mathbf{S}_b$ and $\mathbf{S}_w$ are effectively the reverse, but the whitened approach has demonstrated superior performance in the task face recognition [14, 15]. We shall refer to this technique as whitened PCA (W-PCA).

After we map our gallery and probe images into the desired subspace, we are still left with the problem of which gallery and probe images are most similar. We obtain an estimate of the *a posteriori* probability as,

$$P(c|\mathbf{x}_r) = \text{logsig}\, d_{\cos}(\mathbf{x}_r^{\{c\}}, \mathbf{x}_r) \tag{2}$$

and,

$$d_{\cos}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^T\mathbf{b}}{\sqrt{\mathbf{a}^T\mathbf{a}} \cdot \sqrt{\mathbf{b}^T\mathbf{b}}} \tag{3}$$

where $\mathbf{x}_r$ is the subspace representation r of the probe image and $\mathbf{x}_r^{\{c\}}$ is the subspace representation r of the gallery image for client c. The multiple representations, from which each subspace is generated from, is described in the next section. The cosine distance was employed due

[2] We introduce the term heterogeneous mismatch in this paper to describe a mismatch where only a portion of the signal is corrupted. Heterogeneous mismatch can be dealt with by ignoring the portion of the signal where the corruption has occurred (i.e. maximizing the SNR). Homogeneous mismatch refers to the situation where the entire signal is corrupted, leaving the only avenue for successful pattern recognition with some type of transform to lessen/remove the corruption.

to its good performance cited [15, 16] in previous work in comparison to other distances like the Euclidean or Mahalanobis. The logsig $d = 1/(1 + \exp(-d))$ function is used in Equation 2 to ensure that all resultant match-scores are scaled between zero and one.

3. Representations

In this section, we discuss the multiple representations used in our face verification system, and the motivation for their employment due to their insensitivity to some forms (i.e. spatial and spectral) of mismatch.

3.1. Parts Representations

In cognitive science [17], theories abound over whether humans recognize faces based on component parts or holistic representations. We use the term "monolithic" in this paper to describe the holistic vectorized representation of the face based purely on pixel values within an image array, which can be associated with the holistic mechanism used in a human face recognition system. Similarly, we use the term "parts" to denote a representation of the face that can be considered as an ensemble of image patches of the image array.

A monolithic representation offers the advantage that dependencies across the entire face can be learned, but as a consequence the representation is prone to spatial heterogeneous mismatch (e.g. eye closed, mouth open, beard, etc.). Parts representations have an advantage over monolithic representations in that they are able to assume varying dependencies between other patches within an image. In face recognition literature, there has been a plethora of work conducted on methods that attempt to recognize faces based on component parts of the face rather the monolith. Notable work by Brunelli et al. [10] and Moghaddam et al. [2] cited good recognition performance by trying to align the face to a set of salient anchor points/regions (eg. eyes, nose, mouth). Images in the gallery set were used to create modular templates for comparison with salient regions of the probe images. Both Brunelli and Moghaddam noted superior performance by trying to align the image in a modular manner, rather than holistically, as long as the salient regions had been localized to a satisfactory accuracy. As an in-depth review of parts vs. monolithic face recognition literature is beyond the scope of this paper, the reader is directed to the following article [5].

In our work, we have proposed four different parts representations of the face for verification. Specifically, *left-side*, *right-side*, *eyes* and *nose*. A graphical depiction of the four parts representations can be seen in Figure 1. These four representations were chosen due to their natural invariance to different types of spatial heterogeneous mismatch. An advantage of the parts representations, used in this paper, are that they are extracted solely with reference to the eye positions and the between eye distance in the image. Other parts approaches [2, 10] require strict alignment with other reference points on the face (e.g. mouth and nose), which can sometimes see a degradation in performance if not located properly. For all parts representations in this paper we perform histogram equalization to account for some lighting variability, and ensure the resultant vector of the area is statistically normalized to have zero mean and unit norm before performing discriminant analysis.

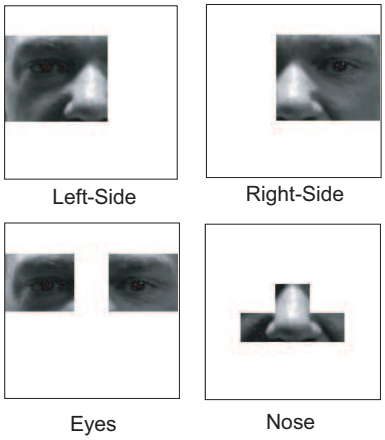

Figure 1: Representations used in verification process.

3.2. Reflectance Representations

An image acquired by a camera can be thought of as the product of two components, reflectance and illumination [8].

$$I(x, y) = L(x, y)R(x, y) \tag{4}$$

Illumination is the amount of light falling on the object due to the light source. Reflectance is the amount of light reflected from the surface of the object. A number of techniques exist for dealing with illumination variation [8, 9][3] especially given multiple images of the subject under different illuminations (e.g. shape from shading). However, in most circumstances one only has a single image of the subject for a single type of illumination. In these circumstances solving Equation 4 for reflectance is an ill-posed problem.

In this paper we will employ two primary estimates of reflectance:

Retinex: Land [18] proposed that an image can be decomposed into reflectance and illumination by estimating the illumination as a low pass version of the original

[3] It is outside the scope of this paper to discuss all possible illumination normalization approaches applied to the task of face recognition. Readers are encouraged to look at Short et al. [8] for a review on this topic.

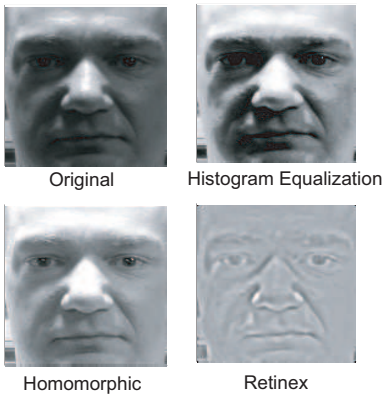

Figure 2: Representations used in verification process.

image. This is often referred to in literature [8, 9, 18] as the "retinex" theory. The reflectance could then be found by dividing the image by the illuminance function. A problem with this approach however is that "halo" effects are often visible at large discontinuities. To reduce halos we pass the reflectance estimate through a 2-D median filter. An example of a retinex face image can be seen in Figure 2.

Homomorphic: making similar assumptions to the retinex approach (i.e. L varies slowly while R can change abrubtly), a homomorphic filter is employed to high pass filter the logarithm of the image. The logarithm of the image is used under the assumption that this representation does a better job of mimicking the human visual system. An example of a homomorphic face image can be seen in Figure 2.

These two estimates of reflectance were chosen due to their ability to deal with differing types of spectral heterogeneous mismatch; note the difference in appearance between the homomorphic and retinex images in Figure 2 due to their differing spectral emphasis/demphasis. Other types of reflectance estimates were ignored (e.g. anisotropic techniques [9]) in this paper due to their computational complexity. In this paper, we will also use the image as an additional approximation of the reflectance after histogram equalization. An example of the original and histogram equalized images can be seen in Figure 2.

4. Combining representations

In this paper, we have chosen different representations of the face due to their natural ability to provide invariance to a number of different mismatches (i.e. for spatial mismatch, a parts representation offers some invariance; for spectral mismatch, a reflectance representation offers some invariance). The problem still remains on how to combine the output probabilities stemming from each representation. A number of classifier combination strategies exist [11, 19]; however many of them require additional training on a separate data-set to gain a synergetic combination function. Kittler et al. [11] demonstrated that the *sum* rule can obtain good performance in classifier combination, when the multiple classifiers are sufficiently diverse and produce match scores approximately representative of their true a posteriori probabilities. Another advantage of the *sum* rule is that it is completely data independent as it requires no tuning set to effectively fuse match scores. The application of the sum rule can be applied below,

$$\zeta(\boldsymbol{\ell}) = \sum_{r=1}^{R} P(c|\mathbf{x}_r) \tag{5}$$

where ζ is the final combined match-score and,

$$\boldsymbol{\ell} = [P(c|\mathbf{x}_1), \ldots, P(c|\mathbf{x}_R)] \tag{6}$$

is a vector of R probability estimates for each representation. In previous work [19] it has been demonstrated that the problem of classifier combination can be posed as a geometrical pattern recognition problem, where the match-scores from each classifier correspond to a separate dimension. Support vector machines (SVMs) have been demonstrated [19] to be extremely useful in this type of classifier combination strategy as they attempt to find the hyper-plane that maximizes the margin between positive (i.e. clients) and negative (i.e. imposters) score vectors (i.e. $\boldsymbol{\ell}$).

For a linear SVM the final match-score after combination is,

$$\zeta(\boldsymbol{\ell}) = \mathbf{w}^T \boldsymbol{\ell} + b \tag{7}$$

where $\mathbf{w}$ is the vector normal to the separating hyperplane and b is the bias. Both $\mathbf{w}$ and b were estimated so that they minimize the structural risk of a tuning set of match score vectors. Note that the SVM match-score formulation in Equation 7 is equivalent to the sum rule match score formulation in Equation 5 if one assumes $\mathbf{w} = \mathbf{1}$ and $b = 0$. Typically $\mathbf{w}$ is not defined explicitly, but through a linear sum of support vectors. As a result SVMs offer additional appeal as they allow for the employment of non-linear combination functions through the use of kernel functions such as the *radial basis function* (RBF), *polynomial*, *sigmoid* kernels. Please refer to [19] for additional information on SVM estimation and kernel selection. A major drawback of SVMs however, in comparison to the sum rule, is the need for a tuning set from which to calculate $\mathbf{w}$ and b.

5. Database and Methodology

In our experiments, we chose to use a subset of the v1.0 FRGC [16] dataset. Specifically data from the challenging Experiment 4 [16] was employed, which had 152 subjects

Figure 3: Depiction of the appearance variation present in Experiment 4 of the FRGC (v1.0) dataset. One can see there is substantial mismatch in illumination, focus and expression between the gallery and probe images.

in the evaluation set (i.e. gallery and probe sets) and 182 subjects in the development set. For each subject in the evaluation set there were only 1 gallery image and 7 probe images. All images were geometrically normalized according to their eye-coordinates to give a cropped face image of 150×130 pixels. Examples of the gallery and probe images used in our experiments can be seen in Figure 3. As one can see in Figure 3, this is an especially challenging data-set as there is substantial mismatch between gallery and probe images in terms of illumination, focus and expression.

Verification is performed by accepting a claimant when his/her match-score is greater than or equal to Th and rejecting him/her when the match-score is less than Th, where Th is a given threshold. Verification performance is evaluated using two measures; being false rejection rate (FRR), where a true client is rejected against their own claim, and false acceptance rate (FAR), where an impostor is accepted as the falsely claimed client. The FAR and FRR measures increase or decrease in contrast to each other based on the threshold Th. The overall verification performance of a system is typically visualized in terms of a receiver operating characteristic (ROC) or detection error tradeoff (DET) curve. A simple measure for overall performance of a verification system is found by determining the equal error rate (EER) for the system, where FAR = FRR. In this paper, we also employ the FRR, given that FAR= 1%, as an additional criteria of performance; as this operating point is an additional good indicator of the overall performance of the system.

6. Results and Discussion

In this section, we shall first investigate which combination function should be used to fuse the match-scores from multiple representations. Second, we will demonstrate the overall improvement in verification performance compared

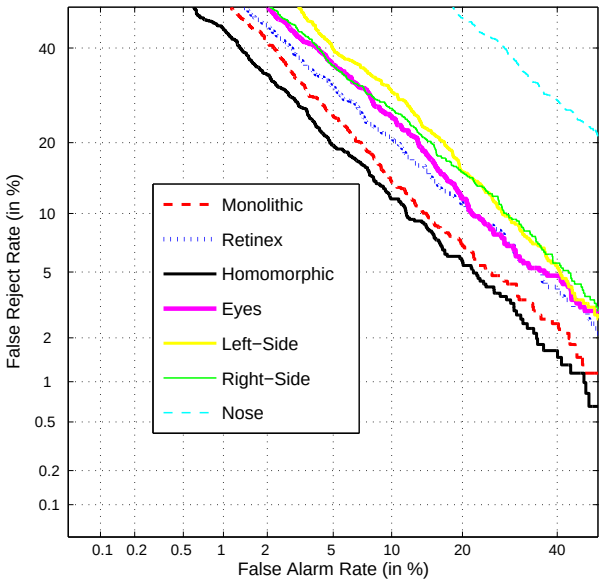

Figure 4: Verification performance of each different representation from both the parts and reflectance categories. Note: the monolithic and all parts representations have all been histogram equalized.

to other subspace algorithms employing a single representation.

6.1. Combination Results

Figure 4 shows the DET curves for each representation being explored in this paper. The homomorphic representation achieves the best performance overall, with the nose representation performing the worst overall. Representations from the reflectance category of representations (i.e. homomorphic, retinex, monolithic) seem to get the best performance in comparison to the parts category of representations (i.e. eyes, left-side, right-side, nose).

To explore the appropriate combination function for fusion with a SVM, we randomly split the FRGC evaluation set of 152 subjects into two equal sets $g1$ and $g2$. When the match-scores from $g1$ were used for evaluation, we trained the SVM with match-scores from set $g2$ and vice versa when the $g2$ set was used for evaluation. In this way we ensure the tuning set used to train the SVM is separate to the one used in evaluation. Results for the $g1$ and $g2$ sets are only reported in Table 1 with all other results in this paper referring to the complete FRGC v1.0 Experiment 4 dataset.

In Table 1, one can see that similar performance was seen in terms of EER and FRR irrespective of the kernel used to train the SVM. This result demonstrates that our assumption of using a linear combination function, as per the sum rule,

Set	g1		g2	
	EER %	FRR %	EER %	FRR %
SVM (Linear)	*7.93*	*25.38*	*7.57*	*17.90*
SVM (Polynomial)	*7.92*	*24.86*	*7.77*	*17.53*
SVM (RBF)	*7.92*	*25.59*	*7.77*	*17.87*
SVM (Sigmoid)	*7.93*	*25.08*	*7.77*	*18.09*
Sum	*8.23*	*27.30*	*7.44*	*18.40*

Table 1: Verification results for different combination strategies. Strategies employ the alternate score set (i.e. $g1$ or $g2$) in training to the set used in evaluation (note: the sum rule required no training). Verification results reported for EER % (FAR = FRR) and FRR % (FAR = 1%). Results demonstrate that the sum rule gives similar performance to more exotic strategies without the need for additional training.

is validated as there is no apparent benefit in using more exotic combination functions. Further, when compared directly to the sum rule there is minimal improvement in performance in terms of EER and FRR.

In Figure 5 one can see a break down of verification performance for each representation category (i.e. combined parts and reflectance representations) as well as the overall combined performance. The sum rule was used in all cases as the combination function. Interestingly in Figure 5, one can see that the parts representations, when combined, gives better performance than the combined reflectance representations. This is a reverse of the verification performance trend seen in Figure 4 where the individual verification results for the reflectance representations were superior to those seen for the parts representations. This result is important as it demonstrates that the individual performance of multiple weak classifiers/verifiers often has no bearing on final fusion performance, provided that those weak verifiers are sufficiently diverse and the correct combination function is chosen. Combining both categories of representations, provides superior performance to either representation category individually.

6.2. Overall Results

In Table 2, we can see a break down of performance for different discriminant analysis subspace algorithms. The single representation algorithms, denoted by the absence of a (MR), used a monolithic representation of the face. In agreement with previous literature [1, 15], the F-LDA and N-LDA algorithms outperform the baseline W-PCA algorithm by a substantial amount. The F-LDA and N-LDA algorithms obtain almost identical performance in terms of EER and FRR, with the N-LDA getting slightly superior performance in terms of FRR. One can see for all single representation algorithms the best FRR performance is still over 50%. When multiple representations are employed using the N-LDA algorithm and sum rule combination function, there is a substantial decrease in both EER (-3.75%) and FRR (-32.77%). This result goes some way to demonstrating the importance of representation for generalization in the task of face verification.

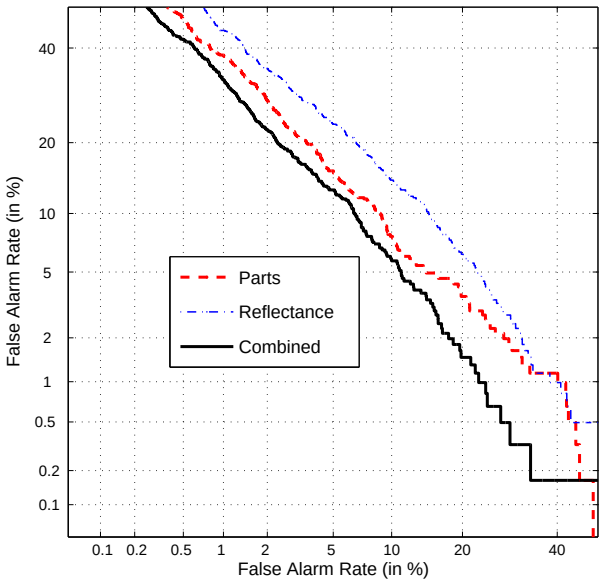

Figure 5: Verification performance of each category of representation (i.e. combined parts and reflectance representations). Note: the combination of all representations (using the sum rule) obtains superior performance to all other representations individually.

7. Conclusion

In this paper, we were able to demonstrate that the employment of multiple forms of subject independent prior knowledge, exemplar and representation, is superior to relying on a single form of prior (i.e. exemplar). In our work, the representation prior took the form of an algorithm that employed multiple representations of the face, based on their insensitivity to anticipated mismatches that may be encoun-

Algorithm	EER %	FRR %
W-PCA	*18.09*	*86.77*
F-LDA	*11.88*	*55.70*
N-LDA	*11.48*	*51.68*
N-LDA (MR)	*7.73*	*22.93*

Table 2: Verification results for different subspace algorithms. Note: the employment of multiple representations (denoted by MR) with the best performing subspace technique (N-LDA) improves FRR performance by an additional 32.77%. Verification results reported for EER % (FAR = FRR) and FRR % (FAR = 1%). Representations from both the parts and reflectance categories were combined for the final N-LDA (MR) algorithm.

tered. The word “anticipated” however, is a term that can often draw criticism, as it is subjective and is open to interpretation.

In our work, we endeavored to employ representations of the face based on an understanding of the physics of the domain. Specifically, we wanted to employ representations that had differing insensitivities to heterogeneous mismatches (i.e. mismatches that did not corrupt the entire signal) both spatially and spectrally in the face image. As a result we proposed two categories of face representation: namely *parts* and *reflectance*. Using these two categories of representation we were able to demonstrate substantial performance improvement on the challenging FRGC v1.0 Experiment 4 face dataset, through the employment of discriminant analysis and the sum rule. The sum rule proved to be approximately equivalent to other more exotic nonlinear combination functions (i.e. SVM combination function), but did not require tuning data for training.

In future work, we would like to improve performance further through the use of more exotic forms of discriminant analysis (e.g. kernel discriminant analysis). We would also like to construct a more formalized framework for obtaining multiple representations, that is not as reliant on heuristics and subjectivity.

References

[1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 711–720, 1997.

[2] B. Moghaddam and A. Pentland, “Probabilistic visual learning for object recognition,” *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 696–710, 1997.

[3] F. De la Torre, R. Gross, S. Baker, and B. V. K. Vijaya Kumar, “Representional Oriented Component Analysis (ROCA) for Face Recognition with One Sample Image per Training Class,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[4] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, “Face recognition using LDA-based algorithms,” *IEEE Trans. Neural Networks*, vol. 14, pp. 195–200, January 2003.

[5] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Computing Surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[6] A. M. Martínez, “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class,” *IEEE Trans. PAMI*, vol. 24, no. 6, pp. 748–763, 2002.

[7] S. Shan, W. Gao, Y. Chang, B. Cao, and P. Yang, “Review the strength of gabor features for face recognition from the angle of its robustness to mis-alignment,” in *International Conference on Pattern Recognition (ICPR)*, 2004.

[8] J. Short, J. Kittler, and K. Messer, “A comparison of photometric normalisation algorithms for face verification,” in *International Conference on Automatic Face and Gesture Recognition (FGR)*, pp. 254–259, 2004.

[9] R. Gross and V. Brajovic, “An image preprocessing algorithm for illumination invariant face recognition,” in *Audio- and Video-based Biometric Person Authentication (AVBPA)*, pp. 10–18, 2003.

[10] R. Brunelli and T. Poggio, “Face recognition: Features versus templates,” *IEEE Trans. PAMI*, vol. 15, no. 10, pp. 1042–1052, 1993.

[11] J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining classifiers,” *IEEE Trans. PAMI*, vol. 20, pp. 226–239, March 1998.

[12] L. F. Chen, H. Y. M. Liao, J. C. Lin, M. T. Ko, and G. J. Yu, “A new LDA-based face recognition system which can solve the small sample size problem,” *Pattern Recognition*, vol. 33, no. 10, pp. 1713–1726, 2000.

[13] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[14] X. Wang and X. Tang, “A unified framework for subspace face recognition,” *IEEE Trans. PAMI*, vol. 26, pp. 1222–1228, September 2004.

[15] J. Ruiz-del-Solar and P. Navarrete, “Towards a generalized eigenspace-based face recognition framework,” in *4th Int. Workshop on Statistical Techniques in Pattern Recognition*, (Windsor, Canada), August 2002.

[16] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the face recognition grand challenge,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.

[17] J. W. Tanaka and M. J. Farah, “The holistic representation of faces,” in *Perception of Faces, Objects, and Scenes* (M. A. Peterson and G. Rhodes, eds.), ch. 2, pp. 53–74, Oxford University Press, Inc., 2003.

[18] E. Land and J. McCann, “Lightness and retinex theory,” *Journal of the Optical Society of America*, vol. 61, pp. 1–11, 1971.

[19] J. Fierrez-Aguiliar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, “Fusion strategies in multimodal biometric verification,” in *International Conference of Multimedia and Expo (ICME)*, vol. 3, pp. 5–8, 2003.