26

Frequency Domain Face Recognition

Marios Savvides, Ramamurthy Bhagavatula, Yung-hui Li
and Ramzi Abiantun
Department of Electrical and Computer Engineering, Carnegie Mellon University
United States of America

1. Introduction

In the always expanding field of biometrics the choice of which biometric modality or modalities to use, is a difficult one. While a particular biometric modality might offer superior discriminative properties (or be more stable over a longer period of time) when compared to another modality, the ease of its acquisition might be quite difficult in comparison. As such, the use of the human face as a biometric modality presents the attractive qualities of significant discrimination with the least amount of intrusiveness. In this sense, the majority of biometric systems whose primary modality is the face, emphasize analysis of the spatial representation of the face i.e., the intensity image of the face. While there has been varying and significant levels of performance achieved through the use of spatial 2-D data, there is significant theoretical work and empirical results that support the use of a frequency domain representation, to achieve greater face recognition performance. The use of the Fourier transform allows us to quickly and easily obtain raw frequency data which is significantly more discriminative (after appropriate data manipulation) than the raw spatial data from which it was derived. We can further increase discrimination through additional signal transforms and specific feature extraction algorithms intended for use in the frequency domain, so we can achieve significant improved performance and distortion tolerance compared to that of their spatial domain counterparts.

In this chapter we will review, outline, and present theory and results that elaborate on frequency domain processing and representations for enhanced face recognition. The second section is a brief literature review of various face recognition algorithms. The third section will focus on two points: a review of the commonly used algorithms such as *Principal Component Analysis* (PCA) (Turk and Pentland, 1991) and *Fisher Linear Discriminant Analysis* (FLDA) (Belhumeur et al., 1997) and their novel use in conjunction with frequency domain processed data for enhancing face recognition ability of these algorithms. A comparison of performance with respect to the use of spatial versus processed and un-processed frequency domain data will be presented. The fourth section will be a thorough analysis and derivation of a family of advanced frequency domain matching algorithms collectively known as *Advanced Correlation Filters* (ACFs). It is in this section that the most significant discussion will occur as ACFs represent the latest advances in frequency domain facial recognition algorithms with specifically built-in distortion tolerance. In the fifth section we present results of more recent research done involving ACFs and face recognition. The final

Source: Face Recognition, Book edited by: Kresimir Delac and Mislav Grgic, ISBN 978-3-902613-03-5, pp.558, I-Tech, Vienna, Austria, June 2007

section will be detail conclusions about the current state of face recognition including further future work to pursue for solving the remaining challenges that currently exist.

2. Face Recognition

The use of facial images as a biometric stems naturally from human perception where everyday interaction is often initiated by the visual recognition of a familiar face. The innate ability of humans to discriminate between faces to an amazing degree causes researchers to strive towards building computer automated facial recognition systems that hope to one day autonomously achieve equal recognition performance. The interest and innovation in this area of pattern recognition continues to yield much innovation and garner significant publicity. As a result, face recognition (Chellappa et al., 1995; Zhao et al., 2003) has become one of the most widely researched biometric applications for which numerous algorithms and research work exists to bring the work to a stage where it can be deployed.

Much initial and current research in this field focuses on maximizing separability of facial data through dimensionality reduction. One of the most widely known of such algorithms is that of PCA also commonly referred to as *Eigenfaces* (Turk and Pentland, 1991). The basic algorithm was modified in numerous ways (Grudin, 2000; Chen et al., 2002, Savvides et al., 2004a, 2004b; Bhagavatula & Savvides, 2005b) to further develop the field of face recognition using PCA variants for enhanced dimensionality reduction with greater discrimination. PCA serves as one of the universal benchmark baseline algorithms for face recognition. Another family of dimensionality reduction algorithms is based on LDA (Fisher, 1936). When applied to face recognition, due to the high-dimensionality nature of face data, this approach is often referred to as *Fisherfaces* (Belhumeur et al., 1997). In contrast to *Eigenfaces*, *Fisherfaces* seek to maximize the relative between-class scatter of data samples from different classes while minimizing within-class scatter of data samples from the same class. Numerous reports have exploited this optimization to advance the field of face recognition using LDA (Swets, D.L. & Weng, J., 1996; Etemad & Chellappa, 1996; Zhao et al. 1998, 1999).

Another actively researched approach to face recognition is that of ACFs. Initially applied in the general field of *Automatic Target Recognition* (ATR), ACFs have also been effectively applied and modified for face recognition applications. Despite their capabilities, ACFs are still less well known than the above mentioned algorithms in the field of biometrics. Due to this fact most significant work concerning ACFs and face recognition comes from the contributions of a few groups. Nonetheless, these contributions are numerous and varied ranging from general face recognition (Savvides et al., 2003c, 2004d; Vijaya Kumar et al., 2006) large scale face recognition (Heo et al., 2006; Savvides et al., 2006a, 2006b), illumination tolerant face recognition (Savvides et al., 2003a, 2003b, 2004a, 2004e, 2004f), multi-modal face recognition (Heo et al., 2005), to PDA/cell-phone based face recognition (Ng et al., 2005).

However, regardless of the algorithm, face recognition is often undermined by the caveat of limited scope with regards to recognition accuracy. Although performance may be reported over what is considered a challenging set of data, it does not necessarily imply its applicability to real world situations. The aspect of real world situations that is most often singled out is that of scale and scope. To this end, large scale evaluations of face recognition algorithms are becoming more common as large scale databases are being created to fill this need. One of the first and most prominent of such evaluations is the *Face Recognition Technology* (FERET) database (Phillips et al., 2000) which ran from 1993 to 1997 in an effort to develop face recognition algorithms for use in security, intelligence, and law enforcement.

Following FERET, the *Face Recognition Vendor Test* (FRVT) (Phillips et al., 2003) was created to evaluate commercially available face recognition systems. Since its conception in 2000, FRVT has been repeated and expanded to include academic groups in 2002 and 2006 to continue evaluation of modern face recognition systems. Perhaps the most widely known and largest evaluation as of yet is the *Face Recognition Grand Challenge* (FRGC) (Phillips et al., 2005) in which participants from both industry and academia were asked to develop face recognition algorithms to be evaluated against the largest publicly available database. Such evaluations have served to better simulate the practical real-world operational scenarios of face recognition.

3. Subspace Modelling Methods

Image data, and particularly facial image data is typically represented in a very high dimensional space, thus a significant amount of data needs to be processed requiring significant computation and memory. In this case, we try to reduce the overall dimensionality of the data by projecting it onto a lower dimensional space that still captures most of the variability and discrimination. Several techniques have been proposed for the latter option such PCA, and *Fisher Discriminant Analysis* (FLDA) (Belhumeur et al., 1997).

3.1 Principal Component Analysis

PCA is among the most widely used dimensionality reduction technique. It enables us to extract a lower dimensional subspace that represents the principal directions of variations of the data with controlled loss of information. Also known as the *Karhunen Loeve Transform* (KLT) or *Hotelling Transform*, its application in face recognition is most commonly known as *Eigenfaces*.

The aim of PCA is to find the principal directions of variation within a given set of data. Let $\mathbf{X}$ denote a $d \times N$ matrix containing N data samples of dimension d vectorized along each column. PCA looks for $k < d$ principal components projections such that the projected data $\left\{y_i = \boldsymbol{\omega}_i^{\mathrm{T}}\mathbf{X}\right\} \in \mathbb{R}^D$ has maximum variance. In other words, we look for the d unit norm direction vectors $\boldsymbol{\omega}_i \in \mathbb{R}^D$ that maximize the variance of the projected data or equivalently best describe the data. These projection vectors form an orthogonal basis that best represent the data in a least-squared error sense. The variance is defined as

$$\begin{aligned}\mathrm{Var}(\mathbf{y}) &= \mathrm{Var}\left(\boldsymbol{\omega}^{\mathrm{T}}\mathbf{x}\right) \\ &= \mathbf{E}\left[\left(\boldsymbol{\omega}^{\mathrm{T}}\mathbf{x} - \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\mu}\right)^2\right] \\ &= \boldsymbol{\omega}^{\mathrm{T}}\mathbf{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\right]\boldsymbol{\omega} \\ &= \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\omega}\end{aligned} \tag{1}$$

such that $\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} = 1$, and $\boldsymbol{\Sigma}$ is defined as

$$\boldsymbol{\Sigma} = E\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\right] \tag{2}$$

where $\boldsymbol{\mu} = E[\mathbf{x}]$. We can estimate the covariance matrix $\hat{\boldsymbol{\Sigma}}$ and the mean $\hat{\boldsymbol{\mu}}$ from the N available data samples as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^{\mathrm{T}} \\ &= \frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{T}}\end{aligned} \tag{3}$$

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{4}$$

where $\mathbf{X}$ now denotes the zero-mean data matrix. To maximize this objective function under the constraint $\|\boldsymbol{\omega}\| = 1$, we utilize the following Lagrangian optimization:

$$L(\boldsymbol{\omega}, \boldsymbol{\lambda}) = \boldsymbol{\omega}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}\boldsymbol{\omega} - \boldsymbol{\lambda}\left(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} - 1\right) \tag{5}$$

To find the extrema we take the derivative with respect to $\boldsymbol{\omega}$ and set the result to zero. Doing so we find that:

$$\hat{\boldsymbol{\Sigma}}\boldsymbol{\omega}_i = \boldsymbol{\lambda}_i\boldsymbol{\omega}_i \tag{6}$$

Premultiplying Eq. (6) by $\boldsymbol{\omega}_i^{\mathrm{T}}$ we get more insight

$$\boldsymbol{\omega}_i^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}\boldsymbol{\omega}_i = \boldsymbol{\lambda}_i\boldsymbol{\omega}_i^{\mathrm{T}}\boldsymbol{\omega}_i \longrightarrow \boldsymbol{\omega}_i^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}\boldsymbol{\omega}_i = Var\{\mathbf{y}_i\} = \boldsymbol{\lambda}_i \tag{7}$$

This corresponds to a standard eigenvalue-eigenvector problem, hence the name *Eigenfaces*.

The directions of variation we are looking for are given by the eigenvectors $\boldsymbol{\omega}_i$ of $\hat{\boldsymbol{\Sigma}}$, and the variances along each direction are given by the corresponding eigenvalues $\boldsymbol{\lambda}_i$ as shown from the above equation. Thus we first choose the eigenvectors (or *Eigenfaces*) with the largest eigenvalues. Moreover, because the covariance matrix is symmetric and positive semi-definite, the eigenvectors produced from Eq. (6) will yield an orthogonal basis. In other words, PCA is essentially a transformation from one coordinate system to a new orthogonal coordinate system which allows us to perform dimensionality reduction and represent the data in the least squared error sense. We apply PCA to face images taken from the Carnegie Mellon University Pose-Illumination-Expression (CMU PIE) No-Light database (Sims et al., 2003) to visualize the resulting *Eigenfaces*. Figure 1 shows the mean image followed by the first 6 dominant *Eigenfaces* computed from this dataset.

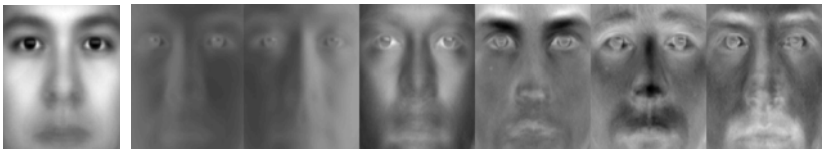


Figure 1. From left to right: PIE No-Light database mean face image followed by the first 6 *Eigenfaces*

3.2 Fisher Linear Discriminant Analysis

Despite its apparent power, PCA has several shortcomings with regards to discriminating between different classes primarily because PCA is optimal for finding projections that are optimal for representation but not necessarily for discrimination.

First developed for taxonomic classifications, LDA (Fisher, 1936) tries to find the optimal set of projection vectors $\boldsymbol{\omega}_i$ that maximize the projected between-class scatter while simultaneously minimizing the projected within-class scatter. This is achieved by maximizing the criterion function equal to the ratio of the determinant of the projected scatter matrices as defined below:

$$J_{\mathrm{FLDA}}(\mathbf{W}) = \frac{\left|\mathbf{W}^{\mathrm{T}}\mathbf{S}_{\mathbf{B}}\mathbf{W}\right|}{\left|\mathbf{W}^{\mathrm{T}}\mathbf{S}_{\mathbf{W}}\mathbf{W}\right|} \tag{8}$$

Where $\mathbf{S}_{\mathbf{B}}$ and $\mathbf{S}_{\mathbf{W}}$ are defined as

$$\mathbf{S}_{\mathbf{B}} = \sum_{i=1}^{c}(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^{\mathrm{T}} \tag{9}$$

$$\mathbf{S}_{\mathbf{W}} = \sum_{i=1}^{c}\sum_{j=1}^{N_i}\left(\mathbf{x}_j^i - \boldsymbol{\mu}_i\right)\left(\mathbf{x}_j^i - \boldsymbol{\mu}_i\right)^{\mathrm{T}} \tag{10}$$

where N_i, $\boldsymbol{\mu}_i$, and $\boldsymbol{\mu}$ are the number of training images for i^{th} class, the mean of the i^{th} class, and the global mean of all classes respectively. To maximize the Fisher criterion we follow a similar derivation to that of Eq. (5) yielding the following generalized eigenvalue-eigenvector problem:

$$\mathbf{S}_{\mathbf{B}}\boldsymbol{\omega}_i = \boldsymbol{\lambda}\mathbf{S}_{\mathbf{W}}\boldsymbol{\omega}_i \tag{11}$$

whose standard eigenvalue-eigenvector problem equivalent is

$$\mathbf{S}_{\mathbf{W}}^{-1}\mathbf{S}_{\mathbf{B}}\boldsymbol{\omega}_i = \boldsymbol{\lambda}_i\boldsymbol{\omega}_i \tag{12}$$

When applying FLDA to face recognition, the data dimensionality d is typically greater than the total number of data samples N. This situation creates rank deficiency problems in $\mathbf{S}_{\mathbf{W}}$. More specifically, note that $\mathbf{S}_{\mathbf{B}}$, being the sum of c outer product matrices has at most rank $c-1$. Similarly, $\mathbf{S}_{\mathbf{W}}$ is not full rank but of rank $N-c$ at most (when $N << d$). To avoid this singularity condition, one can perform PCA on the data to reduce its dimensionality to $N-c$ and then perform FLDA as shown in Eq. (13). The final resulting basis is called *Fisherfaces* (Belhumeur et al., 1997) as given by Eq. (14).

$$\mathbf{W}_{\mathrm{FLDA}} = \arg\max_{\mathbf{W}} \frac{\left|\mathbf{W}^{\mathrm{T}}\mathbf{W}_{\mathrm{PCA}}^{\mathrm{T}}\mathbf{S}_{\mathbf{B}}\mathbf{W}_{\mathrm{PCA}}\mathbf{W}\right|}{\left|\mathbf{W}^{\mathrm{T}}\mathbf{W}_{\mathrm{PCA}}^{\mathrm{T}}\mathbf{S}_{\mathbf{W}}\mathbf{W}_{\mathrm{PCA}}\mathbf{W}\right|} \tag{13}$$

$$\mathbf{W}_{\mathrm{Fisherface}}^{\mathrm{T}} = \mathbf{W}_{\mathrm{FLDA}}^{\mathrm{T}}\mathbf{W}_{\mathrm{PCA}}^{\mathrm{T}} \tag{14}$$

3.3 Frequency Domain Extensions

It has been shown (Oppenheim et al., 1980) that phase information of an image holds the most salient information. In (Hayes et al., 1981), it is shown that one can reconstruct the original signal up to a scale factor given only phase information of the signal. This concept was exploited in face recognition to improve performance over standard algorithms (Savvides et al., 2004b). Figure 2 shows images of two different subjects; each image is split in Fourier domain between magnitude and phase. Figure 2 shows that when the first subject's Fourier magnitude spectrum is coupled with the second subject's Fourier phase spectrum, the resulting image in spatial domain shows significantly more similarity to the second subject compared to the first subject.

(a) (b) (c) (d)

Figure 2. (a) Original image of first subject (b) Original image of second subject (c) Spatial domain image synthesized from combination of Fourier magnitude spectrum of first subject with Fourier phase spectrum of second subject (d) Spatial domain image synthesized from combination of Fourier magnitude spectrum of second subject with Fourier phase spectrum of first subject

However, performing PCA in the frequency domain alone does not constitute any breakthrough, this is because the eigenvectors obtained in the frequency domain are merely the Fourier transform of their spatial domain counterparts. We begin this derivation by defining the standard 2-D *Discrete Fourier Transform* (DFT) pair which is fundamental to the rest of our discussion. Given an 2-D discrete input signal $x[m,n]$ of size $M \times N$ we denote its Fourier transform as $X[k,l]$ whose Fourier transform pair is defined as follows:

$$\begin{aligned} & x[m,n] \underset{F^{-1}}{\overset{F}{\rightleftarrows}} X[k,l] \\ & X[k,l] = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} x[m,n] e^{\frac{-i2\pi km}{M}} e^{\frac{-i2\pi ln}{N}} \\ & x[m,n] = \frac{1}{MN}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1} X[k,l] e^{\frac{i2\pi km}{M}} e^{\frac{i2\pi ln}{N}} \end{aligned} \tag{15}$$

where $i = \sqrt{-1}$, operator F is defined as the forward DFT, and the operator F^{-1} is the inverse DFT.

The estimated covariance matrix of the data in Fourier domain $\hat{\mathbf{\Sigma}}_f$ is given by Eq. (16) where $\mathbf{F}$ is the $d \times d$ Fourier transform matrix containing the DFT basis vectors. The estimated covariance matrix of the data in Fourier domain is given as

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_f &= \frac{1}{N}\sum_{i=1}^{N}\left\{\mathbf{F}\left(\mathbf{x}_i-\hat{\boldsymbol{\mu}}\right)\right\}\left\{\mathbf{F}\left(\mathbf{x}_i-\hat{\boldsymbol{\mu}}\right)\right\}^{+} \\ &= \mathbf{F}\hat{\boldsymbol{\Sigma}}_s\mathbf{F}^{-1}\end{aligned} \tag{16}$$

As was with standard PCA, the eigenvectors $\boldsymbol{\omega}_f$ of $\hat{\boldsymbol{\Sigma}}_f$ are given by

$$\mathbf{F}\hat{\boldsymbol{\Sigma}}_s\mathbf{F}^{-1}\boldsymbol{\omega}_f = \boldsymbol{\lambda}\boldsymbol{\omega}_f \tag{17}$$

Premultiplying each side by $\mathbf{F}^{-1}$ we get

$$\hat{\boldsymbol{\Sigma}}_s\mathbf{F}^{-1}\boldsymbol{\omega}_f = \boldsymbol{\lambda}\mathbf{F}^{-1}\boldsymbol{\omega}_f \tag{18}$$

Comparing Eq. (18) to Eq. (6) we conclude that $\boldsymbol{\omega}_s = \mathbf{F}^{-1}\boldsymbol{\omega}_f$ where $\boldsymbol{\omega}_s$ is an *Eigenface* in spatial domain. We have thus proved that modeling data in the frequency domain does not bring any advantages so far. This fact brings to doubt the usefulness of such a transform with respect to PCA and FLDA without any further processing. However, the ability to distinguish using the magnitude and phase spectrums is the key advantage of the Fourier domain. By modelling the subspace of the phase and magnitude spectrums separately, we can gain further insight and properties of the data otherwise unattainable in the space domain.

3.3.1 Phase Spectrum

It has been shown (Savvides et al., 2004b) that by performing PCA on the phase spectrum alone and disregarding the magnitude spectrum the resulting subspace is more robust with respect to illumination variation. The resulting principal components derived from this new subspace are termed *Eigenphases* in analogy to *Eigenfaces*. It was shown that *Eigenphases* outperform *Eigenfaces* and *Fisherfaces* when trying to recognize not only full faces but also partial or occluded faces as depicted in Figure 4.

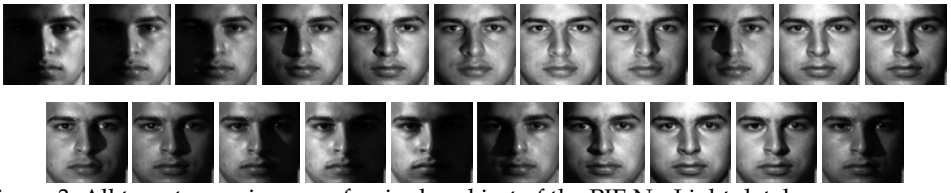
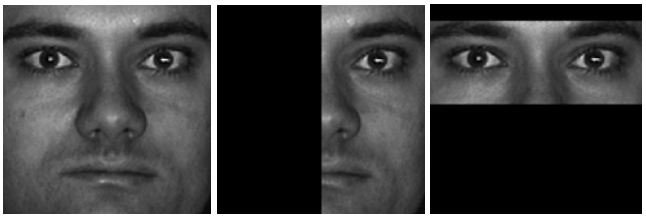
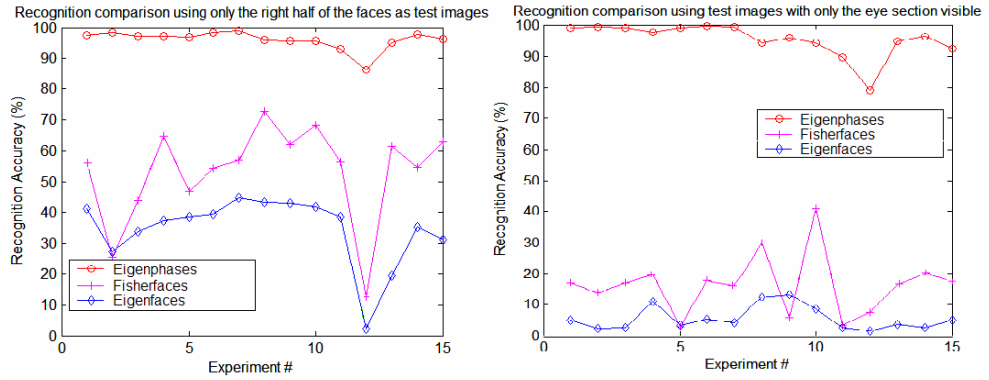


Figure 3. All twenty-one images of a single subject of the PIE No-Light database



(a) (b) (c)

Figure 4. Various occlusions on an example PIE No-Light subject (a) full face (b) right half-face (c) eye section



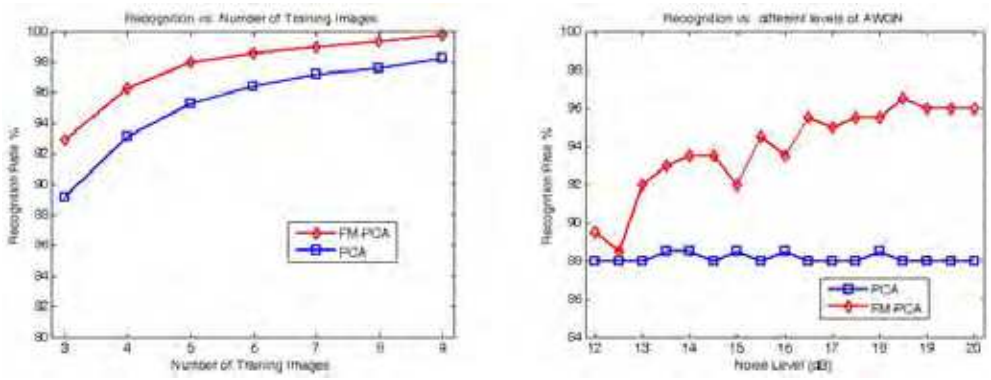
(a) (b)

Figure 5. Rank-1 identification rates obtained by *Eigenphases*, *Eigenfaces*, and *Fisherfaces* for two different experiments each using different types of partial faces. (a) right half face (b) eye-section face

In this work, comparisons between Rank-1 identification rates obtained from *Eigenphases*, *Eigenfaces*, and *Fisherfaces* are made when using whole and partial faces. Training is done on multiple subsets of the PIE database while testing is performed over the whole database. Fifteen different training subsets each representing different types of illumination with the first seven having the most or harshest illumination variation with the remaining eight containing near frontal lighting which are considered the most neutral lighting conditions. Figure 5 depicts the recognition rates obtained with the three different methods using half-faces and eye-sections. These results show that not only do *Eigenphases* outperform *Eigenfaces* and *Fisherfaces* for all experiments by a wide margin, but they also demonstrate minimal performance degradation for half-faces and eye-section faces. This added occlusion robustness is a very attractive property in real-world applications where missing data and poor data quality are common problems.

3.3.2 Magnitude Spectrum

In contrast, if PCA is performed on the magnitude spectrum only, it has been shown (Bhagavatula & Savvides, 2005a) that the resulting subspace holds many advantages over spatial subspaces. Using the Olivetti Research Laboratory (ORL) database, which is noted for significant pose variation, it was shown that the *Fourier Magnitude Principal Component Analysis* (FM-PCA) subspace yielded higher recognition rates across a range of experiments. These experiments included varying the number of training images whose comparison to spatial domain PCA or *Eigenfaces* is illustrated in Figure 6 (a). It was also shown that FM-PCA is more robust to noise as demonstrated in Figure 6 (b). This was verified by corrupting the testing images with varying levels of *Additive White Gaussian Noise* (AWGN). In similar fashion, it was demonstrated that *Fourier Magnititude Fisher Linear Discriminant Analysis* (FM-FLDA) clusters data better than traditional *Fisherfaces* with decreased within-class scatter and increased between-class scatter. FM-FLDA yields higher recognition rates for varying image sizes and resolutions in comparison to spatial FLDA or *Fisherfaces* as tabulated in Table 1.



(a) (b)

Figure 6. Comparisons of identification rates of spatial domain PCA and FM-PCA under varying conditions (a) varying number of training images (b) varying degrees of AWGN noise corrupting the testing images

In addition to increased performance, Fourier Magnitude feature subspaces hold another key advantage. They are shift invariant, as a direct result of the properties of Fourier transform. If the image is shifted in the spatial domain, that shift will translate into a linear-phase change in frequency domain and not in its magnitude. This makes Fourier Magnitude subspaces robust to errors in registration, where the input images are not correctly centred which could cause significant recognition errors. To demonstrate this property, face recognition experiments have been done (Bhagavatula & Savvides, 2005a) by shifting images in both horizontal and vertical directions up to ±5 pixels. These results verify that FM-FLDA and FM-PCA recognition accuracies are not affected, while their spatial domain counterparts are severely affected.

Image Size	32 × 32	64 × 64	112 × 92	128 × 128
FM-Fisher	80.8%	83.2%	84.6%	84.4%
Traditional Fisher	77.7%	78.5%	77.3%	74.0%

Table 1. Recognition accuracies with different image resolutions

4. Advanced Correlation Filters (ACFs)

4.1 Advanced Correlation Filter Basics

The previous sections of this chapter have shown the power of frequency domain representations of data when used in conjunction with techniques and algorithms usually applied to spatial domain representations. However, none of the preceding concepts have been derived from a purely frequency domain approach. By developing algorithms whose focus is on the frequency domain representation of information we can achieve significant gains in performance. One such family of algorithms that have and are still being developed is that of *Correlation Filters* (CFs). CFs have a long and rich history in optics, automatic target recognition, and pattern recognition in general. More recently a new family of CF's termed ACFs (Vijaya Kumar, 1992) have evolved to become the cutting edge of this general family of algorithms. The numerous and varied types of ACFs offer many attractive qualities such as

shift invariance, normalized outputs, and noise tolerance. Their derivations require some knowledge in such fields as linear algebra, signal processing, and detection and estimation theory. We will assume that readers will have sufficient background in these fields and only elucidate on background information when is necessary. We will also now limit our discussion to two-dimensional applications which include facial recognition using grayscale imagery.

To begin the discussion we define a few fundamental terms and conventions that will be used repeatedly for the span of this section. The application of a CF or ACF to an input image will yield a correlation plane. The centre or origin of correlation plane will be considered to be the spatial position (0, 0). Analysis of the correlation plane to some metric of performance or confidence will usually involve calculation and identification of the largest value or peak in the correlation plane.. The simplest CF is the *Matched Filter* (MF), commonly used in applications such as communication channels and radar receivers where the goal is detecting a known signal in additive noise. The concept of noise is a very important aspect of pattern recognition problems. To characterize noise we define the quantitative measure of *Power Spectral Density* (PSD). Using this characterization of noise the MF is developed with the goal of maximizing the *Signal-to-Noise-Ratio* (SNR). Fundamentally this is equivalent to describing a filter whose application to an input signal will minimize the effect of specific type of noise while maximizing the output value when presented with the desired input signal. We will not develop the MF, however multiple other sources provide detailed derivations for varying applications and should be consulted for more information. We will use this fundamental concept of maximizing the response of the desired signal or pattern and minimizing the effects of noise as a guideline in our derivation of ACFs.

One of the fundamental differences between typical CFs and ACFs is the ability to synthesize ACFs from multiple instances of training data or in the case of face recognition, multiple facial images and by doing so, to be able to recognize all instances which are present in the training data. The desire or hope here is that the training data sufficiently represents or captures the potential distortion or variation that might be presented to the recognition system. With respect to face recognition systems this is an extremely desirable quality because the human face is subject to numerous variations both intrinsic and extrinsic. By allowing such variations to be at least partially represented through the use of representative training data we can increase both performance and robustness of face recognition systems.

4.2 Correlation Basics

Before we can derive any ACF we must first lay the framework of correlation with respect to 2D imagery. The standard definition of discrete 2-D correlation between an input 2-D signal $x(m,n)$ and a 2-D filter $h(m,n)$ resulting in 2D correlation output plane $y(m,n)$ is as follows:

$$\begin{aligned} y(m,n) &= x(m,n) \otimes h(m,n) \\ &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x(m+k, n+l) h(k,l) \\ &= \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} x(k,l) h(k-m, l-n) \end{aligned} \tag{19}$$

We will only consider the case of discrete correlation as this is the case of interest in face recognition systems although the analog domain provides some desirable qualities and generalizations. However, for our purposes the desired properties of both correlation and the Fourier transform are present in the discrete domain. Using the Fourier transform and its properties as discussed previously we can express Eq. (19) in the frequency domain as

$$\begin{aligned} y(m,n) &= x(m,n) \otimes h(m,n) \\ &= F^{-1}\left\{ X(k,l) \cdot H^{*}(k,l) \right\} \end{aligned} \tag{20}$$

where $X(k,l)$ and $H(k,l)$ are the 2-D Fourier transforms of $x(m,n)$ and $h(m,n)$ respectively. The symbols F^{-1}, $\cdot$, and * represent the inverse Fourier transform, the element by element (point to point) multiplication of the two 2-D signals, and the element by element conjugation respectively. Correlation in the frequency domain is vastly preferred to correlation in the spatial domain with regards to the number computational floating point operations required.

4.3 Synthetic Discriminant Functions

One of the first ACFs to incorporate such a composite design is the Synthetic Discriminant Function filter (Hester & Casasent, 1980). The design of the *Synthetic Discriminant Function* (SDF) filter is that the filter is created such that it yields a correlation plane whose output at the origin yields a pre-specified value. By introducing such a constraint on the output we not only allow for normalized comparisons but also a degree of discrimination into our filters. This framework refers to the ability to use a single filter to recognize different patterns or classes with sufficient discrimination as opposed to using a single filter for each class or image sample (as with the case of MFs). For example, in a two class problem we would like to design a filter yields an output value of 1 for class 1 while yielding an output value of 0 for class 2. We can achieve this by constraining the correlation plane outputs (at the origin) to be 1 for all training data from class 1 and 0 for all training data from class 2.

Our derivation of the SDF filter begins with an outline of the basic variables and problem definition. Let us assume that we have N facial training images $x_i(m,n)$ of size $d_1 \times d_2$. Define u_i to be the output value of the correlation plane $y_i(m,n)$; that is the result of applying the filter $h(m,n)$ to the training image $x_i(m,n)$. Please note that the output value of the correlation plane is considered to be the value of the correlation plane at the origin or equivalently $y_i(0,0)$. Thus we can define the following equation,

$$u_i = y_i(0,0) = \sum_{m=1}^{d_1} \sum_{n=1}^{d_2} x_i(m,n) h(m,n), \quad 1 \le i \le N \tag{21}$$

The above equation explicitly demonstrates the correlation operation and the constraint on the correlation plane output value at the origin. However, for convenience we can rewrite the above equation into a more compact vector format. Suppose we take a training image $x_i(m,n)$ (of dimensions d_1 x d_2) and place its entries (vectorize) from left to right and top to bottom into a column vector $\mathbf{x}_i$ of length $d = d_1$ x d_2 and similarly for $h(m,n)$ into column vector $\mathbf{h}$ whose length is also d. We can now express Eq. (21) in the following form,

$$u_i = \mathbf{x}_i^{\mathrm{T}}\mathbf{h}, \quad 1 \le i \le N \tag{22}$$

where $^{\mathrm{T}}$ is the transpose operation. We now have a system of N linear equations which encourages us to express them as the product of a matrix and a vector in order to take advantage of matrix algebra. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]$ be matrix of size $d \times N$ whose columns are the training image vectors. Likewise, let $\mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]^{\mathrm{T}}$ be a column vector of length N whose entries are the desired output values. Now we can express this system of linear equations as the following matrix vector product:

$$\mathbf{u} = \mathbf{X}^{\mathbf{T}}\mathbf{h} \tag{23}$$

A unique solution for **h** can be found by assuming that **h** is a linear combination of the training images, i.e. the columns of **X**. In matrix vector form this can be represented as

$$\mathbf{h} = \mathbf{X}\mathbf{a} \tag{24}$$

where **a** is a column vector of length N whose entries are weightings for the linear combination of the columns of **X**. Substituting Eq. (24) into Eq. (23) we form the following equation:

$$\mathbf{u} = \mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{a} \tag{25}$$

From the above equation we can uniquely determine **a** to equal

$$\mathbf{a} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{u} \tag{26}$$

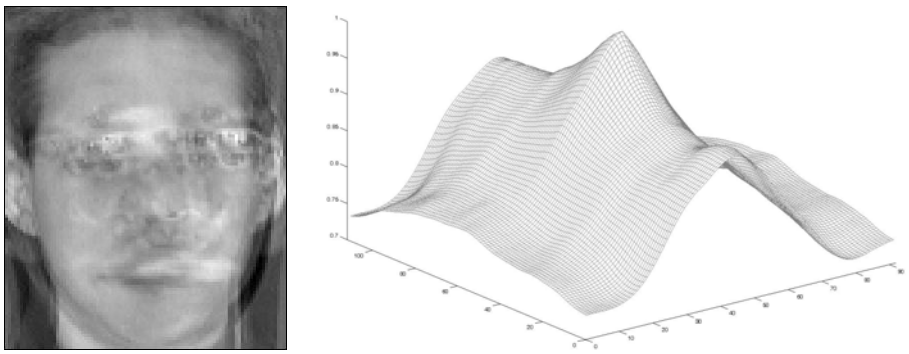
where $^{-1}$ is the standard matrix inverse. Subsequent substitution of the above equation into Eq. (24) yields a solution for the SDF filter **h** which is as follows:

$$\mathbf{h} = \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{u} \tag{27}$$

Eq. (27) expresses the SDF filter **h** as a column vector of length d in the space domain as opposed to the frequency domain.

We use the SDF filter to demonstrate some key characteristics of correlation in general and also some specific qualities of composite correlation. The images shown in Figure 7 are those of a set of training images taken from the ORL face database. We have used these training images to design an SDF filter whose correlation with any of the training images will yield a correlation plane whose output value, i.e. peak will equal 1. Figure 8 (a) shows the resulting SDF filter point spread function (2D-impulse response) , while Figure 8 (b) demonstrates the result of correlating the filter to one of the training images.

Figure 7. Facial training images taken from single subject in the ORL database



(a) (b)

Figure 8. (a) SDF filter derived from training images in Figure 7 (b) Mesh plot of correlation plane produced from application of SDF filter to one of the training images

As can be seen in these figures, the design of the filter guarantees a correlation plane whose peak equals 1 when applied to one of the training images. We make special note of the fact that we no longer specify the value of 1 to be at the origin but merely be the value of the peak (maximum value in the correlation plane) which corresponds to the location of the detected pattern. This consideration reflects the fact that correlation is a shift-invariant operation assuming the pattern of interest is still completely contained within the input image.

4.4 Minimum Average Correlation Energy Filter

Our discussion and development of the SDF filter has motivated us to address the issue of sidelobes whose presence is significant detriment to performance of any ACF. As such we will now derive the *Minimum Average Correlation Energy* (MACE) filter (Mahalanobis et al., 1987) whose design will not only allow us to achieve constrained peaks as in the SDF filter but also suppress sidelobes in order to yield sharp distinct peaks. This is fundamentally a minimization of the sidelobe heights. One approach is to minimize the correlation plane energy which will subsequently suppress sidelobes. We define the term *Average Correlation Energy* (ACE) for the same N training images in the previous section as

$$\text{ACE} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{d_1}\sum_{n=1}^{d_2}\left|y_i(m,n)\right|^2 \tag{28}$$

where the variables d_1, d_2, and $y_i(m,n)$ retain their definitions from our development of the SDF filter. Eq. (28) can be represented in the frequency domain by applying Parseval's Theorem. Letting $Y_i(k,l)$ be the 2-D Fourier transform of $y_i(m,n)$ we express Eq. (28) as

$$\text{ACE} = \frac{1}{N \cdot d}\sum_{i=1}^{N}\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\left|Y_i(k,l)\right|^2 \tag{29}$$

where d again is the total dimensionality of a training image. Since $y_i(m,n)$ is the result of the correlation between an input image $x_i(m,n)$ and our MACE filter $h(m,n)$ we can use Eq. (20) to rewrite the above equation into the following form:

$$\text{ACE} = \frac{1}{N \cdot d} \sum_{i=1}^{N} \sum_{k=1}^{d_1} \sum_{l=1}^{d_2} \left| X_i(k,l) \right|^2 \left| H(k,l) \right|^2 \tag{30}$$

It should be noted that it is at this point in the derivation where the role of the frequency domain representations of both the data and the filter are fundamental to the filter design. Later ACF designs will also utilize the quantitative measure of ACE along with other such measures. For now let us to proceed to again represent Eq. (30) in matrix vector form. Let **h** be a column vector of length d whose elements are taken from $H(k,l)$ and $\mathbf{X}_i$ be a diagonal matrix of size $d \times d$ whose non-zero elements are taken from $X_i(k,l)$. Using these frequency domain terms we can express Eq. (30) as

$$\text{ACE} = \frac{1}{N \cdot d} \sum_{i=1}^{N} \left(\mathbf{h}^{+} \mathbf{X}_i \right) \left(\mathbf{X}_i^{*} \mathbf{h} \right) \tag{31}$$

where the symbol $^{+}$ indicates the conjugate transpose. We can compress this expression further by defining a new diagonal matrix **D** of size $d \times d$ as follows:

$$\mathbf{D} = \frac{1}{N \cdot d} \sum_{i=1}^{N} \mathbf{X}_i \mathbf{X}_i^{*} \tag{32}$$

This allows us to express the quantity of ACE in very concise manner as

$$\text{ACE} = \mathbf{h}^{+} \mathbf{D} \mathbf{h} \tag{33}$$

Our goal in the design of the MACE filter is the minimization of the ACE of the training images while still satisfying the peak constraints we have specified. To accomplish this we must express these constraints in the frequency domain as well. Due to the fact that inner products in the frequency domain (at the origin only) are equivalent to inner products in the spatial domain, we can rewrite the peak constraints expressed in Eq. (23) as

$$\mathbf{X}^{+} \mathbf{h} = d \cdot \mathbf{u} \tag{34}$$

where **X** is a matrix of size $d \times N$ whose columns are the vector representations of the FTs of the training images. Thus, the filter **h** which minimizes Eq. (33) while satisfying the constraints expressed in Eq. (34) is our MACE filter. This constrained optimization can be solved using Lagrange multipliers, which can be found in the original paper (Mahalanobis et al., 1987), which yield the final solution to the frequency domain filter **h**:

$$\mathbf{h} = \mathbf{D}^{-1} \mathbf{X} \left(\mathbf{X}^{+} \mathbf{D}^{-1} \mathbf{X} \right)^{-1} \mathbf{u} \tag{35}$$

The notation and form of the solution allows for simple and efficient calculation of the filter in column vector form from which a simple reshaping operation can be done to recover the 2-D frequency domain filter of size $d_1 \times d_2$. Correlation of the filter with an input image now requires one less Fourier transform as the filter is already represented and stored in the

frequency domain. Using the same training images from our derivation of the SDF filter we can create a MACE filter whose output correlation planes will not contain the problematic sidelobes.

Visualizing the point spread function of the MACE filter itself does not reveal much insight without more significant analysis, but the goals of ACE minimization and constrained peaks are achieved as shown in Figure 9. Not only is the peak equal to 1 as specified, but the sidelobes are drastically suppressed when compared to those in the SDF filter's correlation plane in Figure 8 (b). Noise tolerance can be built in as discussed in the next section.

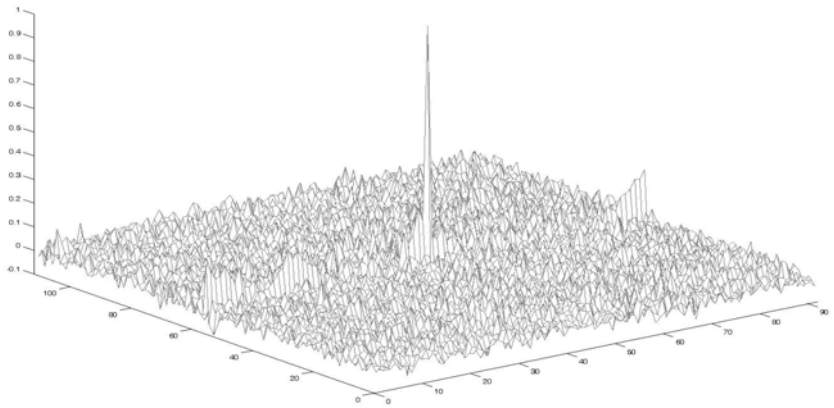


Figure 9. Mesh plot of correlation plane produced from application of MACE filter to one of the training images

4.5 Minimum Variance Synthetic Discriminant Function

Through our derivations of the SDF and MACE filters we have shown that in order to achieve high discriminative ability in our filters we must be able to control the correlation plane through constraints and sidelobe energy minimizations. However, in any practical application we must always take into consideration the factor of noise introduced from varying sources. Whether it is sensor noise or noise caused by background clutter, the presence of noise can have significant impact on any face recognition system. As such we would like to introduce into our ACF designs some degree of noise tolerance. Let us formalize the problem with the following equation:

$$\begin{aligned}(\mathbf{x}+\mathbf{v})^{\mathrm{T}}\mathbf{h} &= \mathbf{x}^{\mathrm{T}}\mathbf{h}+\mathbf{v}^{\mathrm{T}}\mathbf{h}\\ &= u+\delta\end{aligned} \tag{36}$$

where $\mathbf{x}$ is an image vector and $\mathbf{v}$ is the additive noise vector whose responses to the filter vector $\mathbf{h}$ are u and δ respectively. The variations in the outputs of our filter are due to δ and therefore δ is the quantity we wish to suppress. For the rest of the derivation we will assume that our noise processes are stationary. We will also assume that our noise is zero mean without any loss of generality. To suppress the effect of variation in our filter outputs due to noise we aim to minimize the variance of the output noise term δ. Denote this variance as the *Output Noise Variance* (ONV) whose definition is

$$\begin{aligned} \text{ONV} &= E\left\{\delta^2\right\} \\ &= E\left\{\left(\mathbf{v}^{\mathrm{T}}\mathbf{h}\right)^2\right\} \\ &= E\left\{\mathbf{h}^{\mathrm{T}}\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{h}\right\} \\ &= \mathbf{h}^{\mathrm{T}}E\left\{\mathbf{v}\mathbf{v}^{\mathrm{T}}\right\}\mathbf{h} \\ &= \mathbf{h}^{\mathrm{T}}\mathbf{C}\mathbf{h} \end{aligned} \tag{37}$$

where **C** is the covariance matrix of the input noise. We take note of the independence of ONV from the image vector **x** which implies that its definition is identical for all images of interest.

Let us now consider the training images we used in developing the SDF filter whose derivation focused on achieving certain constraints placed on output peak values. We would now like to not only achieve those same constraints expressed in Eq. (23) but also minimize the ONV amongst our training images. This formulation lends itself to the use of Lagrange minimization almost identical to that used in the formulation of the MACE filter to yield the following filter solution:

$$\mathbf{h} = \mathbf{C}^{-1}\mathbf{X}\left(\mathbf{X}^{+}\mathbf{C}^{-1}\mathbf{X}\right)^{-1}\mathbf{u} \tag{38}$$

The above filter is referred to as the *Minimum Variance Synthetic Discriminant Function* (MVSDF) filter (Vijaya Kumar, 1986). While the MVSDF filter does achieve minimum ONV amongst its training images, it does not suppress ACE and as such suffers from unsuppressed sidelobes. In later ACF designs we will show how to achieve an optimal tradeoff between ONV and ACE minimization in order to provide varying degrees of simultaneous noise tolerance and sidelobe suppression.

4.6 Maximum Average Correlation Height Filter

All of the ACFs we have described to this point have been designed with some constraint or optimization in mind that is meant to introduce distortion tolerance into our filters. However, this is but one way and perhaps not the best way to create distortion tolerance. There is no formalized relationship between the constraints we have described so far and the degree of distortion tolerance incorporated into the filter. A more intuitive approach is to remove these constraints to allow for more solutions. In essence this is akin to generalizing the solution space which will hopefully contain solutions to non-training images. This would result in a greater degree of distortion tolerance when compared to ACFs derived using hard constraints.

To address the issue of distortion tolerance it is necessary to first quantize the amount of distortion present in a set of filtered images. To this end we define the *Average Similarity Measure* (ASM) over a set of N filtered images $y_i(m,n)$ as

$$\text{ASM} = \frac{1}{N}\sum_{i=1}^{\mathrm{N}}\sum_{m}\sum_{n}\left(y_i(m,n) - \overline{y}(m,n)\right)^2 \tag{39}$$

where we define $\bar{y}(m,n)$ as the average image whose exact definition is

$$\bar{y}(m,n) = \frac{1}{N}\sum_{j=1}^{N} y_j(m,n) \tag{40}$$

ASM is a measure of the average variation amongst a set of correlation surfaces. As was with previous ACFs we recognize the fact that the above spatial domain equation is equivalently expressed in the frequency domain by applying Parseval's theorem. Let $Y_i(k,l)$ be the 2D-Fourier transform of $y_i(m,n)$ and $\bar{Y}(k,l)$ be the 2D-Fourier transform of $\bar{y}(m,n)$. Also, because we are primarily concerned with the frequency domain let us express $Y_i(k,l)$ and $\bar{Y}(k,l)$ as the column vectors $\mathbf{y}_i$ and $\bar{\mathbf{y}}$ respectively. Eq. (39) is equivalently represented in the frequency domain as

$$\begin{aligned} \text{ASM} &= \frac{1}{N\cdot d}\sum_{i=1}^{\text{N}}\sum_{k=1}^{d_1}\sum_{l=1}^{d_2}\left|Y_i(k,l)-\bar{Y}(k,l)\right|^2 \\ &= \frac{1}{N\cdot d}\sum_{i=1}^{\text{N}}\left|\mathbf{y}_i-\bar{\mathbf{y}}\right|^2 \end{aligned} \tag{41}$$

We must now introduce the filter itself into this metric to allow for optimization with respect to the filter coefficients. Let us consider the ASM over a set of correlation surfaces which are the result of filtering a set N training images $x_i(m,n)$ with the filter $h(m,n)$. As such let us express the Fourier transforms of the i^{th} training image and the filter as $X_i(k,l)$ and $H(k,l)$ respectively. Also, define $\bar{X}(k,l)$, the average Fourier transform of the N training images, as

$$\bar{X}(k,l) = \sum_{i=1}^{N} X_i(k,l) \tag{42}$$

We proceed by representing $X_i(k,l)$, $\bar{X}(k,l)$, $H(k,l)$ as column vectors $\mathbf{x}_i$, $\bar{\mathbf{x}}$, and $\mathbf{h}$ respectively. Let us now define the diagonal matrices $\mathbf{X}_i$ and $\bar{\mathbf{X}}$ whose non-zero elements are taken respectively from $\mathbf{x}_i$ and $\bar{\mathbf{x}}$. Using these matrices we can express $\mathbf{y}_i$ and $\bar{\mathbf{y}}$ as

$$\mathbf{y}_i = \mathbf{X}_i^*\mathbf{h} \tag{43}$$

$$\bar{\mathbf{y}} = \bar{\mathbf{X}}^*\mathbf{h} \tag{44}$$

Substituting the above equations in to Eq. (41) we have the following equivalent expression:

$$\begin{aligned} \text{ASM} &= \frac{1}{N\cdot d}\sum_{i=1}^{\text{N}}\left|\mathbf{X}_i^*\mathbf{h}-\bar{\mathbf{X}}^*\mathbf{h}\right|^2 \\ &= \frac{1}{N\cdot d}\sum_{i=1}^{\text{N}}\mathbf{h}^+\left(\mathbf{X}_i-\bar{\mathbf{X}}\right)\left(\mathbf{X}_i-\bar{\mathbf{X}}\right)^*\mathbf{h} \\ &= \mathbf{h}^+\mathbf{S}\mathbf{h} \end{aligned} \tag{45}$$