Face Recognition and De-Identification in the Frequency Domain

by

Víctor E. Alonso Pérez

In partial fulfillment of the requirements for the degree of

DOCTOR IN SCIENCE WITH SPECIALITY IN ELECTRONICS

in the

National Institute of Astrophysics, Optics and Electronics

December 2016

Tonantzintla, Puebla

Advisor:

Dr. Rogerio Enríquez Caldera, INAOE

Co-Advisor:

Dr. L. Enrique Sucar Succar, INAOE

©INAOE 2016
The author gives INAOE the rights to reproduce and distribute partial or full copies of this Thesis
Abstract

When dealing with images obtained in surveillance context, video-based automatic identity recognition is considerably more difficult than still images-based. The difficulties increase due to several simultaneous and uncontrolled factors, such as severe variations of illumination, expression, pose, occlusion and motion. Furthermore, an automatic face detection process does not provide an accurate registration neither guarantees a full face localization, then as a result there are misalignment errors and/or wrong scales that affect the recognition performance. Moreover, not only recognizing face images is highly important to combat crime, but nowadays, it is also quite important to protect the privacy of subjects visible in some surveillance scenarios. Thus, to obtain the necessary rightful information and at the same time to preserve the rights for people’s privacy is a current challenge. In this dissertation, a general framework for such a challenge is developed.

Firstly, one goal is face recognition considering illumination and expression variations, relative poses, partial occlusions and spatial shifts. To this end, we propose a correlation filter capable of dealing with all those difficulties and achieving at the same time a higher margin of separability between genuine and impostor classes. Because the method uses both, modeling a face subspace in the frequency domain and the original image matrix, we propose a new $(2D)^2(PCA)$ based phase-only method that yields face images with higher quality while preserving edge information allowing to represent and extract more efficiently facial features. Experimental results, using the AR and YALE-B face databases, show that the proposed method achieves higher recognition accuracies than other methods in the frequency domain and in the space domain.

Secondly, another goal is to protect the individual’s identity by de-identifying face
images in such a way that human and machines cannot reliably identify subjects, but simultaneously preserve facial details to perform future classification tasks. Thus, a new real-time foveation-based technique in the discrete cosine transform (DCT) domain, inspired by the neuronal anatomy of the human visual system, is proposed as an alternative method for image obfuscation for face de-identification while simultaneously being aware of basic gender and face expression. The awareness-privacy trade-off at different obfuscation levels is quantified by using a support vector machine (SVM) and a Principal Component Analysis (PCA) method. More specifically, PCA is used to evaluate the de-identification process for privacy protection while SVM is used to quantify the de-identified images intelligibility. Therefore, the trade-off is accomplished by maximizing the classification accuracy for either gender or expression and minimizing the recognition rate. Comparative results using the facial subset of the FERET database show that the new technique achieves a good trade-off between privacy and awareness; with 30% of recognition rate and a classification accuracy as high as 88% (near the 90% baseline) obtained from the common figures of merit using the privacy-awareness map.
Resumen

Cuanto tratamos con imágenes obtenidas en un contexto de video vigilancia, el reconocimiento automático de la identidad de alguna persona en video es considerablemente más dificultoso y desafiante que en imágenes fijas. El incremento de estas dificultades se debe a varios factores simultáneos y no controlados, tal como variaciones severas de iluminación, pose, occlusión and movimiento. Asimismo, un proceso de detección automática de rostros no proporciona exactitud de registro ni garantiza la localización perfecta del rostro, y como resultado, hay errores de desalineación y/o de escalas incorrectas, que afectan el desempeño de reconocimiento facial. Además, no sólo es muy importante reconocer imágenes de rostros en video vigilancia para combatir el crimen o para salvaguardar la seguridad de los habitantes de sospechosos y/o terroristas, sino también es muy importante proteger la privacidad de los sujetos visibles en algunos escenarios de video vigilancia. De esta manera, hay un desafío en obtener información legítima necesaria y al mismo tiempo preservar los derechos de privacidad de las personas en el campo de visión. En ésta tesis, se desarrolla marco general de trabajo para estos desafíos.

En primer lugar, una meta es que el reconocimiento facial considere variaciones de iluminación, expresiones faciales, poses relativos, occlusiones parciales y desplazamientos espaciales. Para éste fin, proponemos un filtro de correlación capaz de tratar con todas esas dificultades y logar al mismo tiempo un mayor margen de separabilidad entre clases genuinas e impostoras. Puesto que éste método modela un subespacio en el dominio de la frecuencia y usa la matriz original de la imagen, proponemos un nuevo método de los espectros de fase basado en \((2D)^2(PCA)\) que produce imágenes de rostro con mayor calidad, preservando la información de borde permitiendo representar y extraer más eficientemente las características faciales. Resultados comparativos, usando las bases de datos AR y YALE-B, muestran que el método propuesto logra mayores tasas de reconocimiento facial que otros métodos en el dominio de la
frecuencia y en el dominio espacial.

En segundo lugar, la otra meta es proteger la identidad de los individuos mediante la de-identificación de rostros de tal forma que los humanos y computadoras junto con software de reconocimiento facial no puedan identificar a los sujetos, pero al mismo tiempo preservar muchos detalles faciales para realizar tareas adicionales de clasificación, que permitan coadyuvar las labores de los operadores de cámara. Así, una nueva técnica en tiempo real basada en foveado en el dominio de la Transformada Discreta Coseno (DCT), inspirada en la anatomía neuronal del sistema visual humano, se propone como un método alternativo para ofuscación de imágenes para la de-identificación de rostros mientras que al mismo tiempo se preservan detalles faciales para clasificación de género y expresión. El equilibrio entre conocimiento-privacidad en diferentes niveles de ofuscación se cuantifica por medio de un Máquina de Soporte Vectorial (SVM) y un método de Análisis de Componentes Principales (PCA). Resultados comparativos usando un subconjunto de rostros de la base de datos FERET muestran la nueva técnica logra un buen balance entre privacidad y conocimiento; con una tasa de reconocimiento del 30 % y una exactitud de clasificación del 88 % (respecto al 90 % de exactitud de clasificación) obtenido del mapa de privacidad-conocimiento.
# Contents

Abstract  
Resumen  

1. Introduction  
   1.1. Major Challenges in Face Recognition  
   1.2. Importance of Privacy  
      1.2.1. Keeping some Face Characteristics in the Proposed Face Method  
   1.3. Thesis Objectives  
      1.3.1. General Objective  
      1.3.2. Specific Objectives  
   1.4. Main Contributions  
   1.5. Thesis Outline  

2. Face De-Identification  
   2.1. Overview in Privacy Protection Methods  
   2.2. Ah-hoc Face De-identification Methods  
      2.2.1. Pixelation  
      2.2.2. Gaussian White Noise  
      2.2.3. Randomizing  
      2.2.4. Thresholding  
      2.2.5. Gaussian Blur  
      2.2.6. Eigenfeatures for Face De-Identification  
   2.3. Image Foveation Filtering  
      2.3.1. The Human Visual System
2.3.2. Photoreceptor Sampling ........................................... 25
2.3.3. Eccentricity ......................................................... 25
2.3.4. Spatial Frequency ..................................................... 26
2.3.5. The Contrast Sensitivity Function ............................... 27
2.4. Evaluation for Face De-Identification ............................. 31
  2.4.1. Preservation Information and Privacy ........................... 31

3. Foveation for Face De-Identification ................................. 33
  3.1. Introduction ........................................................... 33
  3.2. Foveation for Face De-Identification ............................. 33
    3.2.1. Foveation Model ................................................. 34
    3.2.2. DCT Domain Foveation ......................................... 34
  3.3. Experimental Results ............................................... 36
    3.3.1. Test Data .......................................................... 37
    3.3.2. Experiment 1: Performance Analysis of Privacy Protection . 38
    3.3.3. Experiment 2: Performance Analysis of Information Preservation 41
    3.3.4. Experiment 3: Privacy and Awareness Map .................... 42
    3.3.5. Discussion ......................................................... 44
    3.3.6. Experiment 4: Additional Experiments using (2D)²PCA .... 45
      Performance Analysis of Privacy Protection ....................... 45
      Performance Analysis of Awareness ................................ 46
      Privacy and Awareness Map ........................................... 47
      Discussion ............................................................. 48

4. Face Recognition .......................................................... 51
  4.1. Overview of Face Recognition Methods ............................ 51
  4.2. Face Recognition in Subspaces .................................... 57
    4.2.1. Principal Component Analysis ................................ 58
    4.2.2. Eigenfaces and Related Techniques ............................ 60
    4.2.3. Linear Discriminant: Fisherfaces ............................. 61
      Independent Component Analysis .................................. 63
    4.2.4. Two-Dimensional Two-Directional PCA ....................... 63
      Two-Dimensional Principal Component 2DPCA ..................... 65
      Alternative 2DPCA ..................................................... 67
    4.2.5. Kernel-PCA Method .............................................. 68
## 4.3. Correlation Filters

- 4.3.1. Correlation pattern recognition for biometrics

## 5. Hybrid Two-directional Two-dimensional-PCA Correlation Filter for Face Recognition

- 5.1. Introduction
- 5.2. CSS-Based Hybrid (2D)$^2$PCA-Correlation Filter Bank
- 5.3. Phase-Only Fourier Synthesis
- 5.4. Using (2D)$^2$ PCA in the Frequency Domain
- 5.5. Hybrid (2D)$^2$PCA-Correlation Filter
- 5.6. Experimental Results
  - 5.6.1. Experiments on the Yale-B database
  - 5.6.2. Experiments on the AR database
  - 5.6.3. Discussion

## 6. Conclusions and Future Work

- 6.1. Face Recognition
- 6.2. Face De-Identification

## Bibliography

## Appendix

- A. MATLAB Code
  - A.1. Hybrid (2D)$^2$ PCA Correlation Filter
    - A.1.1. YALE-B face database

- B. Published Works
  - B.1. Conference Proceedings
  - B.2. International Journals
Chapter 1

Introduction

Face recognition is a fundamental task that human beings perform effortlessly in their daily lives. However, this is not the case for computer-based systems because building such systems with comparable capabilities has proven to be difficult. Although evidence from psychological and neurobiological experiments, provides insights into how we might code and recognize faces (either in a holistically way or in local features analysis), there is still a great deal of work to achieve an automatic comparable computational reality. A main challenge is to identify fast and accurately a person by taking its image and matching its face against the known face images in a database.

Factors such as lighting conditions, body motions, viewpoints, poses (changes in the egocentric rotation angles yaw, pitch and roll), low resolution and facial expressions contribute to make difficult the before mentioned problem resolved in everyday life. Nevertheless, face recognition computational systems are in great demand, partly because of perceived benefits in areas such as crime-fighting and public-safety, and also partly because of potential applications to ensure higher security such as user authentication for access control to physical and virtual spaces.

A general statement of the problem can be formulated as follows: Given a still image or sequence of video images of a scene, it should be feasible to identify or recognize one or more persons from the scene using a face database. Moreover, the search can be reduced by using available information such as gender, race and age. In order to accomplish this goal, several steps must be performed:

- i) Face localization, which segments the face area from the background. Face detection provides a coarse estimate of the location and scale of the face. This module can be helped by a face landmarking module or face alignment module
in order to locate fine facial marks such eyes, nose mouth, and facial outline.

- **ii) Face normalization.** Because state-of-the-art recognition methods are expected to recognize face images with varying pose and illumination, it normalizes the face geometrically and photometrically. The first process has the purpose to provide invariant features under geometric transformations such as face scaling, rotation, shifting and cropping so the interocular distance is kept constant and irrelevant information such as hair and background is removed. The latter process has the purpose to reduce the illumination variations while still keeping distinguishing features.

- **iii) Face feature extraction** extracts salient information robust with respect to geometrically and photometric variations that is useful for distinguishing faces of different persons.

- **iv) Face matching.** The features extracted from the input image are matched against one or many of the enrolled faces in the face databases under study [1]. When a face recognition system operates in verification mode (authentication), it involves a one-to-one match that compares a query face image against an enrollment face image whose identity is being claimed. In other words, the individual in the unknown image claims an identity. The matcher output; *yes* or *not*. When the system operates in face identification mode, it involves a one-to-many function that compares a query face against multiple faces in the enrollment database in such a way that the query image is associated to one of those in the database. It is assumed that the individual in the unknown image is in the gallery. In this case, when the top match is found, the matcher output is the identity of the input image and when the match score is below a threshold, identity is unknown.

The accuracy of the face recognition system depend on both, the features that are extracted to represent the face, and on the correct face localization and geometrical normalization process.
1.1. Major Challenges in Face Recognition

Face recognition has been considered as one of the major biometrics technologies, due to a number of significant advantages over other identification methods such as fingerprint and iris, where the most important is that the faces can be captured at distance by simply glancing at a camera and even, in a covert manner stressing the advantages of being natural and nonintrusive. An example would be the CCTV cameras placed to monitor specific areas looking for known criminals or tracking suspected people. Thus, research into robust face recognition system has received significant attention, especially during the past several years because they can be used for homeland security, law enforcement, human-computer applications, and for specific applications related to identification or verification - such as access control, surveillance, ID licensing.

Despite that current recognition systems have reached a certain maturity, their success is limited because uncontrolled conditions imposed by many real applications. In the constrained conditions, for example where lighting, frontal pose, stand-off, facial wear and facial expression can be controlled, automatic face recognition can surpass human recognition performance, especially when the database contains a large number of faces. Under those conditions, the best commercial and academic systems achieve verification accuracies comparable to those of fingerprint recognizers. Nevertheless, face recognition under unconstrained conditions is still considerably more challenging because the performance of current algorithms degrades significantly when tested across illumination, pose, expression, motion, spatial shifts, partial occlusion and resolution. In addition, an automatic face detector, in turn, may not guarantee that the location of the face is perfect where faces can be at a wrong scale or misaligned [2]. This low performance on face recognition prevents systems from being widely deployed in real applications and limiting them to be used only in particular scenarios [3].

Face recognition has been analysed from the point of view of face subspace or manifolds. For instance, the eigenface or Principal Component Analysis (PCA) method [4] derives a small number (typically 40 or lower) of principal components or eigenfaces from a set of training faces images. Given the eigenfaces as basis for the face subspace, a face image is compactly represented by a low dimensional feature vector and can be reconstructed as linear combinations of eigenfaces. Subspace modeling techniques have significantly advanced the face recognition technology.
In addition to the large appearance variability in face images due to lighting, pose, and other factors, other three major problems are present in automatic face recognition [1]:

- **Large variability in facial appearance.** Facial appearance is also subject to several factors, including facial pose (or camera viewpoint), facial expression, aging, and illumination. In addition to these, various imaging parameters, such as aperture, exposure time, lens aberrations, and sensor spectral response also increase intra-subject variations. All of these factors are confounded in the image data, so the intra-subject (images belonging to the same subject) variations caused by these factors are almost always larger than inter-subjects variations due to change in face identity [5]. This variability makes it difficult to extract the intrinsic information about the face identity from a facial image and has encouraged many researchers to work in different fields to face this difficult endeavor.

- **High dimensionality and small sample size.** Another challenge in facial recognition is the ability to be generalized. For instance, a face image of size 64 $\times$ 64 pixels resides in a 4096-dimensional feature space. The number of example face images per person (typically fewer than 10 and sometimes just one) available for learning the manifold is usually much smaller than the dimensionality of the image space, and as a result a system trained on a small number of examples may not generalize well for unseen instances of a given face.

- **Complex nonlinear manifolds.** By modeling variations in facial appearance in a PCA subspace by the first principal components, the manifold of all the faces is highly nonlinear and nonconvex, as shown in Figure 1.1. Thus, more complex (nonlinear and nonconvex) trajectories are expected in the original image space. Linear methods such as PCA [4], independent component analysis (ICA) [6], and linear discriminant analysis (LDA) [7] project the data linearly from a high-dimensional space (the image space) to a low-dimensional subspace. As such, they are unable to preserve the nonconvex variations of the face manifold necessary for distinguishing between two different persons. In a linear subspace, Euclidean distance and, more generally, the Mahalanobis distance do
1.1 Major Challenges in Face Recognition

Figure 1.1: Nonlinearity and nonconvexity of face manifolds in a PCA subspace spanned by the first three principal components under a) Translation, b) Rotation, c) Scaling, and d) Gamma transformations. Each plot depicts the manifolds of three individuals (in three colors). A curve represents a sequence of 11 transformed face images in the PCA space, and thus there are 64 curves for each individual [1].
not perform well for discriminating between face and nonface manifolds and between manifolds of different individuals.

- **Low resolution (LR).** It considers to recognize faces images from small size or poor quality with varying pose, illumination, expression and other factors. Face recognition in low-resolution deals with several issues, such as super-resolution (SR) for face recognition, resolution-robust features, unified features spaces, and face detection at a distance. From classification perspective, LR in face images will cause the dimensional mismatch problem due to different resolutions between gallery images and probe ones.

In order to deal with the above mentioned challenges, two main strategies have been explored:

- **i)** Extracting invariant and discriminative face features in which a good set of features is deemed if the face manifold is simpler (e.g., more linear and convex). Also, two processing stages are considered:

  1. Normalizing the face images geometrically and photometrically (illumination correction).
  2. Extracting features in the normalized images, looking for stable processes with respect to geometric and photometric variations (such as Gabor filters or local binary pattern (LBP)).

- **ii)** Constructing a robust face classifier that is able to deal with non-linearities, and also to be able to generalize properly in the constructed feature space. The two stages of processing may be designed jointly in order to obtain high performance in face recognition.

### 1.2. Importance of Privacy

As it was mentioned before, recent technological developments in surveillance video cameras along with facial recognition software permit to track and identify people while in the field of view. Such systems not only can help to prevent crime and law enforcement, but they also represent a potential threat to the general public’s
privacy because the storage and future analysis of the information associated with the captured images. Thus, public concerns due to CCTV based systems are on the rise due to possible misuses by human monitor operators - personal abuse, voyeurism and even criminal purposes - of the captured data [8].

1.2.1. Keeping some Face Characteristics in the Proposed Face Method

Despite of how fair may appear the use of CCTV smart systems, the subject’s identity is being treated by identifying people’s face through face recognition software. Thus, there is a challenge to save the necessary rightful information for performing the surveillance tasks and, at the same time, to preserve the right of privacy for people. Then, several efforts have been performed to integrate into camera systems two complementary actions: to obfuscate the face images for privacy protection and to simultaneously save the de-identified face intelligibility for surveillance tasks.

Therefore, additionally to the previous work of identification in this research, and because the face is already in the frequency domain, an alternative method of de-identifying faces, inspired by human eye’s fovea, is proposed.

It is well-known that the human visual system has higher spatial resolution at the fovea that anywhere else on the retina and that such resolution, smoothly degrades in all directions from the fovea according to the eye eccentricity. On one hand, our alternative method for face de-identification uses the geometry of the eccentricity to remove the perceived intelligibility of a foveated image and, on the other hand, we pursue to quantify the remained intelligibility (e.g., facial expression and gender) in the obfuscated face. Finally, a privacy-awareness map is yielded by evaluating the trade-off between privacy and the preserved information of de-identified faces.

1.3. Thesis Objectives

1.3.1. General Objective

- Develop a face recognition method in the frequency domain which allows to obtain an improvement face representation, while simultaneously dealing with several uncontrolled environment factors such as severe variations of illumina-
tion, expression, pose, occlusion and spatial shifts.

- Introduce an approach to protect the privacy rights of individuals through obfuscating their faces while simultaneously preserving some facial details without revealing the subject’s identity, which guarantees that face recognition software cannot reliably recognize de-identified faces.

1.3.2. Specific Objectives

- Propose and implement two-directional two-dimensional Principal Component Analysis $(2D)^2$PCA in the frequency domain on the phase-only spectrum of face images.

- Extend even further the hybrid-$(2D)^2$PCA subspace to develop, in an innovative manner, a shift-invariant hybrid $(2D)^2$PCA-correlation filter that performs phase matching through correlation.

- Evaluate the hybrid $(2D)^2$PCA-correlation filter performance for face recognition when it is applied on images from YALE-B and AR databases.

- Compare the recognition performance of several methods in the frequency domain ($(FM)^2$PCA, Hybrid PCA-correlation filter) and in the spatial domain (the PCA, IPCA and the $(2D)^2$PCA).

- Propose an innovative method of de-identifying images based on a Foveation technique in the Discrete Cosine Transform (DCT) domain.

- Evaluate the face de-identification ad-hoc techniques using the FERET face databases.

- Evaluate the privacy protection of the de-identified faces by using a Principal Component Analysis (PCA) recognizer and the two-directional two-dimensional Principal Component Analysis $(2D)^2$PCA.

- Measurement of data utility preservation such as gender and facial expression by using a Support Vector Machine (SVM) classifier.
1.4. Main Contributions

The main contributions of this work are:

- The development of a novel method for face recognition termed two-directional two-dimensional Principal Component Analysis (2D)$^2$PCA in the frequency domain on the phase-only spectrum of the face images. This new hybrid-(2D)$^2$PCA subspace not only increases the margin of separability between classes by representing more efficiently the face images of the phase-only synthesis in conditions of extreme illumination and partially occluded faces; but it is also computationally more efficient than the PCA alone and it increases the recognition rate.

Specifically, the proposed hybrid (2D)$^2$PCA-correlation filter has proven to be much more robust to illumination variations, partial occlusions and spatial shifts than the PCA only based correlation filter. Furthermore, when performing (2D)$^2$PCA to compute the eigenvectors of the phase-only spectra, the original matrix does not need to be transformed into a 1D long vector beforehand, as it is done in 1D-PCA, but instead two covariance matrices are directly constructed from rows and columns of the 2D image matrices. In fact, the (2D)$^2$PCA method is essentially the PCA method performed on the rows and columns of all images [9].

Additionally, because the size of both image covariance matrices is equal to the width and height of the face image, (2D)$^2$PCA not only represents and extracts more efficiently the facial features but also can significantly improve the speed of image feature extraction. In contrast to the developed method, the recognition accuracy of all the methods mentioned above could not achieve good performance when the face images present large variations in lighting, occlusion, facial expression or spatial shifts.

- Extending further the hybrid-(2D)$^2$PCA subspace to achieve shift-invariance when the test input image is shifted, probably due to registration errors, then the correlation output is also shifted by the same amount of pixels. This function avoids changing the peak sharpness on the correlation plane, and therefore the classification decision is not affected by this shifting. Thus, we develop a hybrid (2D)$^2$PCA-correlation filter that performs phase matching with a built-in shift-invariance function.
An alternative method of de-identifying images based on DCT-domain foveation technique. This part of our work has two main advantages:

1) Presents a real-time foveation technique, inspired by the human visual system (HVS), proposed as an alternative method for image obfuscation, which not only protects the individual’s identity but also preserves information that can be used for subsequent tasks on surveillance cameras helping human operators. The main advantage over other ad-doc methods is that it not only keeps a balance between privacy and data preservation but also provides a computational speed-up. That is, the DCT-domain foveation for privacy protection can be incorporated into both DCT-based video codification and standard motion compensation techniques for low bit rate coding such as the H.263 or MPEG.4 video coding [10]. Specifically, in a macro-block based compression scheme, the disjoint foveation regions coincides with macro-blocks without requiring an extra processing at the coder. By applying DCT-domain based foveation is simplified both the cut-off frequency and contrast sensitivity computations because it is only computed for each macro-block according to the viewing distance.

2) Proposes a framework to simultaneously evaluate the privacy protection and information preservation of any given method without limiting or canceling its use, which allows to perform an analysis of the balance between privacy and awareness for risky situations.

1.5. Thesis Outline

The structure of this thesis is based on a general framework addressing on the one hand, appearance-based face recognition under uncontrolled factors using a face representation in the frequency domain and, on the other hand, face de-identification also represented in the frequency domain to preserve privacy of individuals.

Chapter 2 depicts an overview in privacy protection methods, focusing on ad-hoc face de-identification methods. We evaluated six different methods to fade out face images. Furthermore, we present the basic idea to foveate an image or filter it
in a similar way to foveation in the human visual system. The basic concepts, to understand both the privacy and awareness performances, are also presented.

Chapter 3 introduces the DCT-domain foveation method adopted for face de-identification. We perform a comparative analysis with the most commonly used distortion methods, and also present an analysis using the privacy-awareness map by using a face recognition software and a SVM classifier, respectively. We discuss our results and findings.

Chapter 4 presents an overview in face recognition methods as a specific research area, focusing on advance correlation filters (ACFs) for face recognition. We analyse some linear and nonlinear subspaces (PCA, KPCA, etc.) to represent face images, and perform face recognition. Furthermore, we also present an overview of correlation filters applied in the field of face recognition.

Chapter 5 depicts the proposed hybrid $(2D)\text{PCA-corr}$ filter method for face recognition. It explains how to yield a hybrid $(2D)\text{PCA-corr}$ filter bank based on class-specific subspaces built from the phase-only spectra of training images. Our experimental results, using public face databases, compare the proposed method along with other methods in the frequency and space domain.

Chapter 6 presents conclusions and open research work associated to both face de-identification and face recognition in the frequency domain methods.
Chapter 2

Face De-Identification

2.1. Overview in Privacy Protection Methods

Many different privacy protection techniques have been developed and applied to surveillance systems. In modern systems, and despite of undergoing processes to de-identify faces that were captured, it is a current challenge to keep a balance between privacy and awareness, this is, keeping the strictly necessary information without intruding in the actual identity of someone in the scene. It is desired that a face de-identification method can hold enough useful information for video surveillance tasks while concealing personal information. Thus, the de-identified face intelligibility can still serve to perform some classification enabling either a human or an automatic process to identify, for instance, the subject’s gender or facial expression.

Over the past few years, several techniques for face de-identification in image/video surveillance have been proposed independently of any compression scheme used, as shown in Figure 2.1. Among the most commonly used techniques for privacy protection are blurring and pixelation [11–16]. Both of these techniques, independently of the compression scheme, have been used to fade out either the face or body silhouette in an automated surveillance closed-circuit television (CCTV). For instance, Chimoni et al [13] proposed their PrivSurv, which is a video surveillance system that can adaptively protect privacy by determining the object’s privacy according to the closeness between the object and the viewer. Neustaedter et al [15] addressed the problem of how to develop and evaluate privacy-protecting strategies using techniques to protect privacy in a home media space under a user’s special interface design. In their work, they showed that video blurring was unable to balance privacy with awareness for risky situations.
Several of these techniques are evaluated in a subjective manner [14, 15, 17]. To the human eyes, a de-identified face image may look sufficiently obfuscated to protect privacy, but it could provide little protection if it is applied some face recognition software [18]. In other words, a subjective evaluation methodology to measure the privacy-awareness trade-off, should not to be applied when a face recognition software threatens privacy of people. In counterpart of the subjectively evaluated ad hoc methods, there exist other methods where authors evaluate whether or not their proposed methods defeat facial recognition systems [8, 16, 18–25]. This means that if ad-hoc techniques were not evaluated by a face recognition software, these would not be applied to perform a specific and previously required privacy protection task without first performing an analysis in balancing privacy protection and data utility preservation and thus, limiting or canceling out their use.

For example, Newton et al [18] proposed a privacy-enabling algorithm called \textit{k-same} which guarantees that face recognition software cannot reliably recognize de-identified faces, while preserving many facial details. Gross et al [22] introduced an extension to Newton’s algorithm, the so called \textit{k-same-select}, which integrates a data utility function to evaluate the trade-off between the privacy and the preserved information (for gender and expression) from the de-identified face. Meng et al developed an algorithm, \textit{k-same-furthest}, which maximizes the identity loss by de-identifying each original face with an aggregate face of a cluster that, identity-wise, is found...
furthest away from the original face. Thus, there have been efforts to find ways aiming to preserve some useful information when face images are obfuscated to conceal individual’s identity using naïve methods as well as privacy protection schemes based on k-anonymity [26].

Recently, advanced face de-identification techniques based on cryptography and coding have also been proposed, among them are found scrambling [8], cryptography [27,28], JPEG 2000 [29], watermarking [30–32], discrete wavelet transform [33], geometric transformation [19, 20] and face subspaces [21]. For instance, Phillips [21] presents an algorithm for privacy protection through the reduction of the number of eigenvectors used in reconstructing face images. The privacy operating characteristics (POC) is proposed to measure quantitatively the resulting trade-off between privacy and security. Boult [27] presented a cryptographically invertible obfuscation method for privacy protection called PICO. Although it allows to monitor actions of a person without the revealing the person’s identity, facial details such gender, race or facial expression are removed preventing or limiting its use for surveillance tasks.

On the other hand, Dufaux et al [8] proposed a method for H.264/AVC which scrambles the quantized transform coefficients of each 4 × 4 blocks of a previously defined ROI by pseudo-randomly flipping their sign and applying a random permutation to rearrange the order of transform coefficients in those blocks. In order to distinguish between the scrambled ROI and the unscrambled background, two slice groups are defined using Flexible Macroblock Ordering. The main advantage of the ROI-based transform-domain method is that can be reversible through a secret encryption key. However, the scrambling method depends on the compression algorithm used, since it should be designed and implemented for every video encoder and decoder used, e.g., MPEG-4 [34] or H264/ACV [8]. Both encryption and scrambling methods fulfil in protecting privacy but require modification at the encoder.

Conversely, Korshunov et al [19, 20] introduced a distortion method based on a geometric transformation that is independent of a compression scheme. However, there is no evidence if the de-identified face preserves useful information (traits or behavioural) to be used for additional surveillance tasks.

Then, according to the literature review, several efforts in face de-identification, to protect privacy and preserve some useful information, have been done by implementing ad hoc methods as well as privacy protection schemes based on k-anonymity. It is also worth to point out the importance to evaluate either the level of obfuscation
or the level of preserved information and, correspondingly, to perform an analysis of the trade-off between privacy protection and data preservation.

In order to decide what type of method or filter performs better, an awareness-privacy map is yielded to compare how well a specific filter can protect the privacy for a given degree of classification accuracy. In this dissertation, awareness issues of the surveillance task and the effectiveness of privacy protection in de-identified face images are evaluated by measuring both face recognition and classification accuracies at the same time.
2.2. Ah-hoc Face De-identification Methods

This section is a necessary preface to the proposed alternative method for face de-identification to simultaneously deal with identification and privacy protection evaluation. In this part, several ad-hoc techniques commonly used for concealing the subject’s identity are described and, then we will be able to present the results whose analysis evaluates different face de-identification algorithms through face recognition software while simultaneously preserves sufficient awareness on both gender and facial expression. Thus, in following section we investigate six well-known methods of face de-identification, namely pixelation, blurring, thresholding, eigenfeatures and Gaussian white noise and randomizing noise [11–17,35].

Below we briefly review the commonly used ad-hoc methods for face de-identification, which are based on pixel operations. These distortion methods have been used to fade out the sensitive area of an individual, such as its face or body visible in the scene.

2.2.1. Pixelation

A face image, of total size of $N \times N$, is split into sub-blocks of size $p \times p$ and within each sub-block, the value of each pixel is replaced by the average pixel intensity value. In this technique, the larger the value of $p$, the greater is the loss of information in the face image. The parameter $p$ is varied to modify the sub-block size between 1 and 31. See Figure 2.2 for examples of face images resulting of applying pixelation with different values of $p$ to an original face image.

![Figure 2.2: De-identified images using the pixelation method at different levels.](image-url)
2.2.2. Gaussian White Noise

Adding noise to an pixel’s value $x$ is achieved by modeling the noise through a Gaussian distribution function $p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, and in some cases, generated by the Box-Muller Method [36]. The smaller the value of $\sigma$, the less information is removed. The parameter $\sigma$ is normalized in range from 0 to 1. Figure 2.3 shows the resulting example images of adding noise for various levels of $\sigma$.

![Figure 2.3: De-identified images using the Gaussian white noise method at different levels.](image)

2.2.3. Randomizing

A face image is transformed by randomly changing both, the pixel position and the pixel values depending upon the value of $r$, chosen in a range from 0 to a given maximum value for each pixel in the face image. Note that both position and pixel values are applied for the same set of images. Likewise as the Gaussian white noise method, while increasing of the value $r$, the greater is the loss of information. The maximum number of distorted pixels that can be altered in the face image are $m \times n$. See Figure 2.4 for examples of distorted face images.

2.2.4. Thresholding

A gray-scale face image is binarized by modifying the pixel values to either dark or bright, depending upon a threshold value $t$ between 0 and 255. Figure 2.5 shows example images of applying the thresholding approach with different values of the threshold $t$ to face images.
2.2 Ah-hoc Face De-identification Methods

![De-identified images using the randomizing method at different levels.](image1)

![De-identified images using the thresholding method at different levels.](image2)

2.2.5. Gaussian Blur

The blurred face image is computed as the convolution of the original face image with a symmetric Gaussian kernel: 
\[ G_\eta(x, y) = \frac{1}{2\pi\eta^2} \exp\left(-\frac{x^2+y^2}{2\eta^2}\right), \]
where \( x \) and \( y \) are the two-dimension Cartesian coordinates and \( \eta \) controls the width of the kernel. The parameter \( \eta \) is varied to control the neighborhood size between 1 and 31, respectively. Figure 2.6 shows typical blurred images at different values of distortion \( \eta \) applied to a face image.

2.2.6. Eigenfeatures for Face De-Identification

In order to attenuate the ability to determine person’s identity, the eigenfeatures or eigenfaces approach can be used to yield a degraded image through the reduction of the number of eigenvectors in the representation. This method requires to perform an unsupervised Principal Component Analysis (PCA), maximizing the variance of the
variability of each feature component, and discarding the least significant components resulting in a dimension reduction [37].

Given a training set of $N$ facial images $\{x_1, \ldots, x_n\}$, each input image is transformed into $n$-dimensional image space. Then, the total scatter matrix is defined as $C_s = \sum_{k=1}^{N}(x_k - \mu)(x_k - \mu)^T$, where $N$ is the number of sample images and $\mu \in \mathbb{R}^n$ is the mean image of all samples. In this method, the PCA linear transform maps the original $n$-dimensional image space into an $m$-dimensional feature space $y_k \in \mathbb{R}^m$, which is defined by the following equation [38]:

$$y_k = W^T x_k \quad k = 1, 2, \ldots, N$$

(2.2.1)

where $W^T \in \mathbb{R}^{n \times m}$ is a matrix with orthonormal columns and $m << n$.

Thus, the scatter of the transformed feature vectors $\{y_1, y_2, \ldots, y_N\}$ is $W^T C_s W$. The optimal projection $W_{opt}$ is chosen to maximize the determinant of the total scatter matrix of the projected samples $\{y_1, y_2, \ldots, y_N\}$:

$$W_{opt} = \{w_1, w_2, \ldots, w_m\} = \arg \max_w |W^T C_s W|$$

(2.2.2)

where $\{w_1, w_2, \ldots, w_m\}$ are eigenvectors of $C_s$ corresponding to the first $m$ largest eigenvalues.

The number of eigenvectors of $w$ can be obtained by setting a threshold:

$$\frac{\sum_{i=1}^{w} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq \theta.$$  

(2.2.3)

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are $n$ largest eigenvalues of $C_s$ and $\theta$ is a pre-set threshold value.
Thus, in the PCA representation, each image is represented by projecting it into the subspace according to the projection $W_{opt}$; when varying the eigenfeatures parameter $w$ which ranges from 1 to 200, different levels of privacy protection are provided. Sample images resulting of the reduction of the eigenvectors number $w$ are shown in Figure 2.7.

![Figure 2.7: De-identified images using the eigenfeatures approach at different levels.](image)
2.3. Image Foveation Filtering

The word fovea refers to a small pit in the retina of the eye, and it means ditch or pit. The fovea is located along the optical axis of the eye, which results in images at the center of gaze being focused directly onto the fovea. The human visual system has a higher resolution at the fovea - due to a higher density of photoreceptor (rods and cones) and ganglion cells - than anywhere else on the retina. This spatial resolution is smoothly degraded in all directions from the fovea, as shown in Figure 2.8. Also, the term foveation refers to this smoothly varying resolution.

![Image showing the density of cone cells and ganglion cells with eccentricity from the fovea](image.png)

Figure 2.8: The density of cone cells and ganglion cells with eccentricity from the fovea (the point of maximum resolution). It is located at zero degrees eccentricity, and the eccentricity increases equally in all directions away from the fovea [41].

The basic idea to foveate an image is to sample or filter the image in a manner similar to the foveation in the human eye. Figure 2.9 shows an example of a foveated image at different viewing distances. The point of maximum resolution or the fixation point in a foveated image is commonly called fovea because this is the point in the scene that is being focused onto the fovea, whereas the periphery of an image is called non-foveal. In other words, the fixation point selects which regions will be represented in high resolution and which regions will be represented in low resolution. Although the human eye has a single anatomical fovea, multiple fixation points can be made to create resolution distributions and to better capture the content of a foveated image.
2.3 Image Foveation Filtering

Figure 2.9: An example of a foveated image at different viewing distances. The fixation point in the foveated image is marked with a white circle.
2.3.1. The Human Visual System

This section presents an overview of the anatomy of the human visual system (HVS) and studies the low level vision processing. In short, the eye is the primary imaging device in the HVS, as shown in Figure 2.10, and is on average, 24 mm long by 22 mm across.

![Figure 2.10: The basic anatomy of the human eye. The highest concentration of photoreceptors in the retina is found in the fovea, and it is aligned along the optical axis of the cornea and lens [41].](image)

The sclera or white part is the main body of the eye. The cornea is transparent and is located at the front of the eye; its function is to focus the light entering the eye. The pupil is a hole that allows light to enter the eye. And the iris adjusts the diameter of the pupil in order to control the amount of light entering the eye. In low light, the iris dilates or increases the diameter of the pupil, which allows to enter more light into the eye. In bright light, it reduces the diameter of the pupil allowing less light into the eye. Similar to the cornea, the lens focuses the light entering the eye, but the focusing power of the lens is adjustable via the zonular fibers.

Thus, photoreceptors in the retina are responsible for converting impinging light into electrical signals. Such electrical signals from the retina are pooled and transmitted, along the optic nerve, to the brain for processing.
2.3 Image Foveation Filtering

2.3.2. Photoreceptor Sampling

The Shannon-Nyquist Sampling Theorem specifies the necessary sampling rate $f_s$ to unambiguously reconstruct a continuous signal (with a maximum frequency $f_{\text{max}}$) as,

$$f_s > 2f_{\text{max}}$$

(2.3.1)

where the sampling frequency $f_s$ must be greater than twice the maximum frequency $f_{\text{max}}$.

Therefore, the human eye cannot resolve signals whose frequency content is greater than $f_{\text{cutoff}}$,

$$f_{\text{cutoff}} = \frac{1}{2C_s}$$

(2.3.2)

where $C_s$ is the spacing between cones. Thus, near the fovea, $C_s$ is small and the corresponding $f_{\text{cutoff}}$ is near or above the diffraction limit. However, in the periphery the cones are more sparsely spaced, $C_s$ is greater, and the sampling limit is lower than that imposed by diffraction limit. The diffraction limit imposed by the pupil restricts the resolvability of high frequency signals. The maximum resolvable frequency is given by the Rayleigh Criterion

$$\sin \theta_R = 1.22 \frac{\lambda}{d},$$

(2.3.3)

where $\theta_R$ is the maximum resolvable frequency, $\lambda$ is the wavelength of the light, and $d$ is the diameter of the pupil. Eq. 2.3.3 shows that the maximum resolvable frequency actually depends on the wavelength, or color, of the light.

Thus, the diffraction limit is a byproduct of the wave nature of light, where a wave passing through a small aperture causes constructive and destructive interference on the imaging plane. The result is that the image of a point of light produces a blur, which is known as the point spread function. This function is conceptually similar to an impulse response function, and characterizes the filtering effects of the optical system.

2.3.3. Eccentricity

Eccentricity is defined as the angular distance from the fovea, along the retina, expressed in degrees. The eccentricity of a point on the retina is the angle between the optical axis and the line that intersects the nodal point of the lens and the fovea.
and thus the fovea has an eccentricity of zero degrees, and this eccentricity increases equally in all directions away from fovea. It is also worth noting that eccentricity is circulant symmetric. That is, a point located five degrees toward the temple and five degrees toward the nose are indistinguishable in terms of eccentricity.

2.3.4. Spatial Frequency

Spatial frequency measurements related to the human eye are measured in cycles per degree because spatial frequencies of a signal projected onto the retina from a particular source are dependent on the distance between the observer and the source. Then it should be considered that if the observer moves away from a scene or object, its projected spatial frequency content expands towards the higher frequencies (because the projected image gets smaller and closer to the fovea), meanwhile if the observer moves toward the source, its frequency content moves to a lower spatial frequency range.

Although most anatomical and psychophysical frequency measurements are taken with respect to cycles per degree, digital image frequencies are expressed in terms of cycles per pixel, then it takes at least two pixels to represent a single cycle. Therefore, the maximum frequency on a computer monitor is 0.5 cycles per pixel. Frequencies above this rate cannot be represented without being aliased to lower frequencies. Therefore, whenever applying psychophysical measurements to algorithms that deal with digital images, it is necessary to translate cycles per degree to cycles per pixel.

This conversion depends on the size of the pixels as well as the observers distance from the screen or object. Thus, the size of the pixel in radians is given approximately by

\[ \tan \theta = \frac{w}{v} \]  \hspace{1cm} (2.3.4)

where \( w \) is the width of the pixel in some measurement unit, \( v \) is the distance of the viewer from the screen in the same units, and \( \theta \) is the angle subtended by the pixel in radians. For small values of \( \theta \), \( \tan \theta \approx \theta \). Ignoring gaps between pixels, the number of pixels per unit distance \( N \), is approximately equal to \( \frac{1}{w} \). It is generally easiest to measure the width of the monitor and divide the horizontal resolution by the width.
to find $N$. Therefore, the number of radians per pixel is approximately:

$$\theta = \frac{1}{N\nu}$$  \hspace{1cm} (2.3.5)

and the number of pixels per degree is then

$$\frac{\pi N\nu}{180}$$  \hspace{1cm} (2.3.6)

The equivalent cycles per degree can be calculated by multiplying the cycles per pixel by this value:

$$\frac{Cycles}{Degree} = \frac{Cycles}{Pixel} \times \frac{\pi N\nu}{180}$$  \hspace{1cm} (2.3.7)

This conversion is fairly standard throughout the literature, however it is worth noting that it is only an approximation. For observers placed far from the screen and viewing images located directly on the fovea, the errors incurred by this approximations are small. However, when the subject is placed very close to the screen or when viewing images in the periphery, it may be necessary to use more precise calculations. Handling frequencies in the periphery in a correct way is difficult because, unless the screen is curved, the distance from the eye to the screen increases with eccentricity.

### 2.3.5. The Contrast Sensitivity Function

The contrast sensitivity function measures the ability of an observer to properly discern visual stimuli variations in both contrast and frequency. The contrast of an image patch can be computed by using the root mean square (RMS) of the intensities $I_{ij}$ corresponding to each pixel $(i, j)$

$$\tilde{I} = \frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} I_{ij},$$  \hspace{1cm} (2.3.8)

and the standard deviation

$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ij} - \tilde{I})^2}$$  \hspace{1cm} (2.3.9)
where $M$ and $N$ are the width and height of the image, respectively. Figure 2.11 illustrates how sensitivity varies with contrast and frequency. Contrast varies along the $Y$-axis while frequency varies along the $X$-axis. As it can be seen from the figure, a sinusoid signal has a higher visibility at higher levels of contrast (which are in the middle frequency range) that when the sinusoid is at the lower or higher ends of the frequency spectrum.

For a given frequency, the contrast at which the sinusoid is just barely visible is called the contrast threshold. Psychovisual experiments found that this contrast threshold varied not only with frequency but also with eccentricity. A model that provided a good fit to the available psychophysical data was proposed by Geisler et al. [39]:

$$CT(f, e) = CT_0 \exp\left(\frac{\alpha f (e_2 + e)}{e_2}\right),$$  \hspace{1cm} (2.3.10)

where $f$ is spatial frequency in cycles per degree, $e$ is retinal eccentricity in degrees, $CT_0$ is the minimal contrast threshold, $\alpha$ is a spatial frequency decay constant, and $e_2$ is a half-resolution eccentricity constant.

The contrast sensitivity function $CS(f, e)$, which attempts to quantify how noticeable the frequency components of a given contrast are to human viewer, is defined as the contrast threshold reciprocal $CS(f, e) = 1/CT(f, e)$.

The contrast and the contrast sensitivity function parameters also provided a good fit for the data from similar psychovisual responses (experiments were carried out by different research groups). However, this model is not perfectly accurate for everyone due to variations between individuals. For this model, the best fitting parameters values are $\alpha = 0.106$, $e_2 = 2.3^\circ$, and $CT_0 = \frac{1}{64}$.

Then, for a given eccentricity (considering a normalized contrast with a maximum value set to 1) we can calculate a cut-off frequency, by setting $CT = 1$ and solving for $f$

$$f_c(e) = \frac{e_2 \ln\left(\frac{1}{CT_0}\right)}{\alpha (e + e_2)},$$  \hspace{1cm} (2.3.11)

where $f_c$ is the cut-off frequency in cycles per degree, and is in fact, the spatial frequency that a viewer will be unable to perceive regardless of contrast.

For eccentricities close to zero, the cut-off frequency can exceed the highest frequency that can be represented on the screen without aliasing. Such a frequency (the maximum frequency on a computer monitor) is 0.5 cycles per pixel.
2.3 Image Foveation Filtering

That way, we can now compute the corresponding pixels per degree by using Eq. (2.3.6) and thus the maximum representable frequency in cycles per degree is:

\[
\frac{1}{2} \times \text{pixels per degree} = \frac{\pi N v}{360}
\]  

(2.3.12)

Thus, the cut-off frequency for a particular pixel \( x \) at a viewing distance \( v \) is calculated by:

\[
f_c(e) = \min \left( \frac{e_2 \ln \left( \frac{1}{CT_0} \right) \pi N v}{\alpha (e + e_2)}, \frac{\pi N v}{360} \right).
\]

(2.3.13)

Figure 2.12 illustrates how the eccentricity depends on the contrast threshold and contrast sensitivity function. As it can be seen from figure, given a contrast threshold, the contrast sensitivity quantifies how noticeable the frequency components are to a human observer. Therefore, the frequency response of the human visual system is modeled by both the cut-off frequency and contrast sensitivity functions, and therefore they are used in foveated techniques as models of the frequency response of the human visual system.
Figure 2.12: The contrast sensitivity (brightness indicates the strength of contrast sensitivity) with respect to foveal distance in pixels. The solid white line corresponds to the contrast threshold (the cut-off frequency). The a), b), c) and d) figures are for $N = 512$ and viewing distance $v = 1, 3, 6,$ and $16$ of the image width, respectively.
2.4. Evaluation for Face De-Identification

This section presents the basic concepts used to evaluate both the face de-identification and the information preservation processes. From this concepts, we will obtain a privacy-awareness map, which represents the trade-off between privacy protection and data preservation.

2.4.1. Preservation Information and Privacy

In order to understand how the comparative analysis and performance evaluation is done, it is essential to consider beforehand the following definitions [22]:

**Image Sets.** A gallery \( G = \{G_1, G_2, \ldots, G_n\} \) is a set of \( N \) face images of known individuals related to a face recognition software, and a probe \( P = \{P_1, P_2, \ldots, P_L\} \) is a set containing \( L \) face images corresponding to unknown subjects. All images consist of subjects with one single image per person.

**Face Recognition.** Given the gallery \( G \) and probe \( P \) image sets, we define a facial recognition function as \( f_{FR} : P \rightarrow G \), which assigns a test image \( p \in P \) with exactly one image \( g \in G \). A face recognition rate \( R_{f_{FR}} : P \times G \rightarrow [0, \ldots, 1] \) is a function that computes the percent of correctly matched face images using a face recognition function \( f_{FR} \).

**Face De-identification.** We define a de-identification function as \( f_d : \Psi_o \rightarrow \Psi_d \) such that \( f_d(\Gamma) = \Gamma^d \) with \( \Gamma \neq \Gamma^d \), where \( \Psi_o = \{\psi_1 \ldots \psi_i\} \) and \( \Psi_d = \{\psi^d_1 \ldots \psi^d_i\} \) are the original face images set and the de-identified face images set, respectively.

Finally, because an essential part of this research is to establish a quantitative comparison of the simultaneous evaluation of performance on de-identification and preservation of information on face images we define the following measurements.

**Face De-identification Performance.** Given a face de-identification method, \( f_d \), a gallery set, \( G \), a probe set, \( P \), and a face recognition function, \( f_{FR} \), we assess the performance of the de-identification function, \( f_d \), afforded by privacy protection models by measuring face recognition software performance, in the rank-1 recognition rates. So that, \( R_{f_{FR}}(f_{FR}, f_d(P), G) < R_{f_{FR}}(f_{FR}, P, G) \) for different face recognition functions, \( f_{FR} \), and face de-identification functions, \( f_d \).

**Preserved Information Performance.** Let \( \Psi_{o,d} = \{\psi_1 \ldots \psi_i\} \) be a face image set, and let \( \Phi : \Psi_{o,d} \rightarrow \{c_1, c_2, \ldots c_k\} \) be a data preserving utility function, which
associates each face image in $\Psi_{o,d}$ with exactly one class $c_i, i = 1, ..., k$, assuming the correct class $c_i$ for each face image is known. Now, let $f_{eval} : \Phi \rightarrow \{0,1\}$ be an evaluation function which assigns a score of correct class association for a data preserving utility function $\Phi$ and a face image set $\Psi_{o,d}$.

For example, utility functions for both expression and gender classifications are defined. Given a facial expression classifier $f_{EX} : \Phi \rightarrow \{l_1, l_2, \ldots l_i\}$ assigning facial expression labels (e.g., neutral and smile) to the test set, the expression utility function $f_{eval_{EX}}$ is defined as $R_{f_{EX}} = f_{eval_{EX}}$. Similarly, given a gender classifier $f_{GE} : \Phi \rightarrow \{l_{male}, l_{female}\}$, that assigns gender labels to test set, the gender utility function $f_{eval_{EX}}$ is defined as the classification accuracy rate $R_{f_{GE}} = f_{eval_{GE}}$. 
Chapter 3

Foveation for Face De-Identification

3.1. Introduction

This chapter presents the DCT-domain based foveation method proposed for face de-identification and compares six ad-hoc methods previously depicted in section 2.2. We evaluate the performance of both the privacy and the preserved information in faces de-identified. In section 3.2 we present the DCT-domain foveation method used to provide privacy. In section 3.3 we present the experimental results using two face recognition algorithms to measure the privacy protection: the PCA standard and the $(2D)^2$PCA. To quantify the preserved information (either gender or expression), we use a Support Vector Machine (SVM) classifier.

3.2. Foveation for Face De-Identification

Since the spatial resolution of the human eye varies with eccentricity from the fovea or point of fixation, one promising way to obfuscate a face in images and video is through a foveation technique by decreasing the spatial resolution at the fovea as a function of eccentricity according to the viewing distance. Such a process allows to remove the perceived image intelligibility in a foveated image in order to protect the privacy of the subjects.

In other words, our goal is to do exactly what the foveation does but using its function in reverse form, that is, to decrease the spatial resolution at the fovea according to the viewing distance. Therefore, in regions near the point fixation, the face
image will be degraded, and consequently in regions in the periphery, the face will be also degraded. Thus, more visual distortion may be achieved in order to form a de-identified image by varying the viewing distance values from the foveation point given by the viewer.

In the next section, it is explained the alternative method based on the DCT-domain image foveation technique in order to simultaneously preserve privacy and information in faces images.

### 3.2.1. Foveation Model

A model that fits the empirically contrast sensitivity of HVS as function of retinal eccentricity was proposed by Geisler et al [39]. Experimentally, they found out that the contrast threshold varied not only with frequency but also with eccentricity:

\[ CT(f, e) = CT_0 \exp \left( \frac{\alpha f (e_2 + e)}{e_2} \right) \]

(3.2.1)

where \( CT \) is the visible contrast threshold to perceive a sinusoid of spatial frequency \( f \) in cycles per degree at a retinal eccentricity \( e \) in degrees, and \( CT_0 \), \( \alpha \) and \( e_2 \) are the minimum contrast threshold, the spatial frequency decay constant, and the half-resolution eccentricity constant, respectively. The contrast sensitivity function \( CS(f, e) \) is defined as a contrast threshold reciprocal \( CS(f, e) = 1/CT(f, e) \).

The best fitting parameter values given in [39] are \( CT_0 = \frac{1}{64} \), \( \alpha = 0.106 \) and \( e_2 = 2.3^\circ \). This parameters in our framework were adopted because the contrast threshold in Eq. 3.2.1 is fitted to the human contrast sensitivity data measured as a function of spatial frequency and retinal eccentricity [40].

### 3.2.2. DCT Domain Foveation

Foveation filtering in the DCT domain is an image processing technique that produces images of spatially varying resolution or foveated images through spatially varying low-pass filters. This implementation employs the given cut-off frequency in Eq. 3.2.2.

An approximated foveation model, simplifying cut-off frequency and contrast sensitivity computations by introducing disjoint DCT-domain regions [10], was adopted to achieve our goal for face de-identification. In this model each macro-block of si-
ze $M \times M$ is assumed to have a constant cut-off frequency, in such a way that the proper cut-off frequency can be found by computing the square distance from the center of the region to the fixation point. For instance, by setting the smallest cut-off frequency yields a larger distance affecting the periphery of the the perceived image quality. Conversely, for regions near the point of maximum resolution a higher cut-off frequency is computed which corresponds to a higher quality image [41].

For a given eccentricity, $e$, from the foveation region point, one way of finding the local maximal perceptual $f_c$ is the use of Eq. 3.2.1. Thus, all spatial frequencies higher than $f_c$ will be invisible in the area beyond the given eccentricity, $e$, regardless of their contrast. By setting the left side of Eq. 3.2.1 to the maximum contrast $CT = 1.0$ and solving for $f$, the local cut-off frequency is obtained as [40,41]:

$$f_c(e) = \frac{e_2}{\alpha(e + e_2)} \ln \frac{1}{CT_0} \left(\frac{cycles}{degree}\right)$$

(3.2.2)

where $f_c$ is the cut-off frequency in cycles per degree.

![Figure 3.1: Viewing geometry: a human observer observing an image fixates at $(x_f, y_f)$. Eccentricity is the angle indicating where the image point $(x, y)$ falls on the retina [10].](image)

It is assumed that the line from the fovea to the point of fixation in the image is perpendicular to the image plane, and the position of foveation point and the viewing distance, $v$, from the eye to the image plane are known, as shown in Figure 3.1. More specifically, given a point $x = (x, y)^T$ and a foveation point $x_f = (x_f, y_f)^T$ (measured in pixels) in the image, the distance from $x$ to $x_f$ is given by:

$$d(x) = ||x - x_f||_2 = \sqrt{(x - x_f)^2 + (y - y_f)^2}$$

(3.2.3)

$$e(v, x) = \frac{180}{\pi} \tan^{-1} \left( \frac{d(x)}{Nv} \right)$$

(3.2.4)
where \( e(v, x) \) is the eccentricity for each macro-block, \( N \) and \( v \) are the image width and the viewing distance in pixels, respectively. Eq. 3.2.4 allows to remove undetectable high spatial frequencies by mapping the visual spatial frequency \( f_c \) (cycles/degree) to the digital frequency \( f_d \) with respect to the viewing distance. Now, the cut-off frequency model from Eq. 3.2.2 can be converted into the image pixel domain by multiplying the cycles per degree:

\[
\frac{\text{Cycles}}{\text{Pixel}} = \frac{\text{Cycles}}{\text{Degree}} \times \frac{180}{\pi N v} \tag{3.2.5}
\]

To perform the foveation filtering over an image, it is split into macro-blocks of size \( M \times M \) in the DCT domain and afterwards a weighting mask \( W(k_1, k_2) = w(k_1)w(k_2), k = 1, \ldots, M \) is applied to the DCT coefficients using triangular-transition weights without multipliers:

\[
w(k) = \begin{cases} 
1, & 0 \leq k \leq k_c \\
0.5, & k = k_c + 1 \\
0, & \text{otherwise}
\end{cases} \tag{3.2.6}
\]

where \( k_c \) denotes the index of the highest subband that is not suppressed [10]. More specifically, the designed weights of Eq. 3.2.6 approximate a low pass filter and reduce complexity by eliminating the need to perform multiplications, and they are computed according to the eccentricity of each macro-block with respect to the fixation point.

A foveated image of fixed size 64 \( \times \) 64 for face de-identification is obtained by applying the DCT-domain based foveation method, as shown in Figure 3.2, for different viewing distances \( v = 3, 4, 6, 8 \) and 15 times of the image width, respectively. Note that the higher value of \( v \), the greater is the loss of information over the face image. Here, the foveation point \( \vec{x}_f = [32, 32] \) is at the center of the face image, the macro-block size is \( b = 16 \times 16 \) pixels, and the observed image width is \( N = 64 \) pixels.

### 3.3. Experimental Results

In this section, the proposed methodology to achieve a trade-off between privacy protection and information preservation is analyzed through a comparative study with the most common ad-hoc methods. Each face de-identification method is tested with
3.3 Experimental Results

Figure 3.2: De-identified images at different levels of the DCT-domain based foveation method.

both, a face recognizer based on PCA at different obfuscation levels to evaluate the performance of privacy; and the corresponding preservation of data is experimentally compared and measured using a support vector machine classifier. Two classification tasks are performed: gender and expression classification. Experimental results are presented in a privacy-awareness map for both face recognition and face classification accuracies. Such a map allows policy makers to find the appropriate operating point for surveillance systems. The experimental results have shown that the DCT-domain based foveation method can be used efficiently compared to other methods for privacy protection since it provides a good balance between privacy and information preservation.

3.3.1. Test Data

Experimental results are carried out on the grayscale FERET database [42]. In this database, each person has two images ($fa$ and $fb$) which are obtained at different times and with different facial expressions. A small subset containing 548 subjects (267 female and 281 male), showing both neutral and smile expressions, were randomly chosen, as shown in Figure 3.3. The set of all neutral face images, $fa$, is considered as the gallery, and the set of face images with other facial expression, $fb$ (the probe), is used to test the algorithms. The performance of face recognition is reported based on the rank-1 recognition rates between a test $p \in P$ and the images in the gallery.

For PCA recognition tasks, training images used in our experiments are roughly 80% from the $fa$ partition and 20% from $Dup I$ images (subjects taken in a later
time). Face images are matched by projecting them into facespace via the projection $W_{\text{opt}}$ defined in section 2.2.6. The Cosine Mahalanobis distance or Whitened Cosine is used as a similarity measurement.

All the images were previously aligned based on the eye coordinates, so that the face images were rotated, cropped, and scaled to have the same interocular distance [43]. Each resulting face image of this normalization process has a fixed size of $64 \times 64$ pixels. Since all images are geometrically normalized, the foveation point $x^f = [32, 32]$ is chosen as the center of the face images. The macro-block size, $b$, is selected to be equal to 16 pixels and the observed image width is $N = 64$ pixels.

Since the de-identification methods have a different intensity scale to obfuscate, as shown in Figure 3.4, all methods are normalized to a common scale which ranges from 0% to 100%. That is, we choose the first 15 levels which indicate the maximum distortion range allowed by each method.

### 3.3.2. Experiment 1: Performance Analysis of Privacy Protection

To measure the performance of the privacy protection, a closet-set model is used, that is, every individual depicted in the test set is also present in the gallery set. The PCA [38] method is evaluated using the original images in the gallery set and the images de-identified by the ad-hoc methods (pixelation, blurring, randomizing, Gaussian noise, threshold and eigenfeatures) as well as foveation in the probe set. A set of de-identification levels to defeat the face recognition software is defined in order to rank each recognition accuracy as shown in Figure 3.5.

The first 7 levels of de-identification have to be chosen to provide privacy protection for both foveation and pixelation methods. In general, when applying eigenfe-
3.3 Experimental Results

Figure 3.4: De-identified images at different levels. a) Original, b) Pixelation, c) Gaussian Noise, d) Randomizing, e) Threshold, f) Blurring, g) Eigenfeatures and h) DCT-domain foveation. The parameter values are $N = 64$ pixels and the macro-block size is $b = 16$, the fixation point is found at the $x^f = 32, 32$ pixels.
3. Foveation for Face De-Identification

Figure 3.5: Face recognition rates for original images in the gallery and de-identified images in the probe. It is evident that the declining of recognition accuracies for each of these algorithms differ at different levels of de-identifying. For both foveation and pixelation recognition accuracies decrease more 20\% at the level 7 than blurring. It can be seen that the foveation curve falls more steeply. The Random noise, Gaussian noise and eigenfeatures methods decrease slowly up to 15 levels, and therefore need much more levels for privacy protection.

It can be observed that for both foveation and pixelation methods, the face recognition performance is below 20\% when the de-identification level is equal 8. For blurring, ten de-identification levels are needed to conceal information that identifies the individuals. In the case of randomizing recognition accuracies stay high (> 65\%) for up to 15 levels of de-identification, meaning that this method is not suitable for anonymization.

As it can be seen from figure, despite using a de-identification level of 46\%, foveation is able to preserve many facial characteristics for classification tasks such as gender and expression characterization. Foveation has proven to provide a good trade-off between privacy and awareness and these results will be discussed in the next section, where it is simultaneously examined both face recognition and face classification accuracies.
3.3 Experimental Results

![Graph showing classification rates for different de-identification methods.](image)

Figure 3.6: Data preservation assessment is achieved by measuring classification accuracy across all levels of de-identification. For both gender and facial expression classification tasks, foveation performs comparably over blurring at all levels of de-identification, with higher classification rates at lower face recognition rates.

3.3.3. Experiment 2: Performance Analysis of Information Preservation

In order to quantify the preserved information (gender and expression) of de-identified faces cross all levels of de-identification, a support vector machine (SVM) classifier with a radial basis function kernel, to perform face classification, is used to
quantify the de-identified images intelligibility. Specifically, a 5-fold cross-validation scheme which involves partitioning the dataset in 5 equally sized subsets is used (training on four subsets and testing on the remaining fifth), and the average accuracy is reported.

In the experiment, each de-identification level is used to measure the preserved information in de-identified images. Figure 3.6 shows resulting accuracies and their rank for both gender and expression classification accuracies, accomplished by de-identification methods. The classification accuracy on the original data is plotted as baseline.

It can be observed a 50% decrease of classification accuracy at higher levels of de-identification for both pixelation and thresholding methods, while other methods such as blurring, randomizing, gaussian noise, eigenfeatures and foveation can actually preserve more information at higher obfuscation levels (more than 80%, near the baseline) for up to 14 levels. That is, the information is slowly decreasing at the higher levels of de-identification. Nevertheless, such results do not imply that the subject’s identity is being protected as it will be discussed in the next section, but they do indicate that the performance of the DCT-domain based foveation technique behaves well since the information is better preserved, and further foveation is more efficient in terms of computational overhead [10].

3.3.4. Experiment 3: Privacy and Awareness Map

In this section, the trade-off between privacy and awareness for each of these de-identification methods is presented by introducing a privacy-awareness map. This map is obtained by plotting both gender and expression classification rates along with the recognition rates across all different levels of de-identification. Thus, the trade-off is achieved by finding a quantitative measure maximizing the classification accuracy for gender and expression while minimizing the recognition rate. Note that the trade-off between privacy and preserved information is different for each method since each of those algorithms differently de-identifies at the expense of losing classification accuracy.

In these experiments, for each of de-identification levels for a given de-identification algorithm, we plot the privacy-awareness map using the data points indicating privacy (i.e. recognition rate) and preserved data (i.e. expression classification...
3.3 Experimental Results

Figure 3.7: The privacy-awareness map. Each data point represents privacy (i.e. recognition accuracy) and awareness (i.e. classification accuracy) for a given de-identification level in each of these algorithms.

It can be observed that for gender classification as a data preservation function, the DCT-domain based foveation technique performs comparably over blurring. However, comparative results show that the trade-off between privacy and awareness,
compared with other ad-hoc methods, achieve a 88\% classification accuracy, especially at lower face recognition rates (< 30\%). While, on the other hand, for pixelation, the data utility is lower than the threshold method where the classification accuracy for pixelation is below 70\%. Other methods such as randomizing, Gaussian noise and thresholding hold a data utility as high as 80\% when the face recognition performance is higher than 80\%.

For expression classification as data preservation function, the DCT-domain based foveation technique overcomes the other methods even at face recognition rates lower than 20\%, as shown in Figure 3.7 b). That is, decline in classification accuracy, while increasing for up to 15 levels, is constant. Decline in accuracy for the other methods such as blurring and pixelation is also slowly decreasing.

Although foveation performs comparably over blurring for both gender and expression, and better than pixelation for expression, at higher levels of de-identification; its performance on this trade-off between privacy and information preservation in comparison with the other methods is more effective, specifically at face recognition rates lower than 20\%.

3.3.5. Discussion

The privacy-awareness map allows to analyze the privacy-awareness trade-off. From this map, we can validate both face recognition and face classification accuracies for each method. Thus, the performance of the ad-hoc methods such as pixelation, Gaussian noise, thresholding, randomizing, pixelation is overtaken by foveation, but blurring is comparable with foveation in performance. However, the foveation-based privacy approach is significantly faster than its spatial domain counterpart, so it requires much lower computation overhead since the cut-off frequency is constant over each of these foveation regions.

In other words, DCT-domain based foveation for face de-identification can be incorporated into both DCT-based video codification and standard motion compensation techniques for low bit rate coding such as H.263 or MPEG4 video coding. Furthermore, the main advantage of DCT-domain based foveation is that in a macro-block based compression scheme, the disjoint foveation regions can coincide with macro-blocks without requiring an extra processing at the coder. That way, the cut-off frequency is only computed for each of these disjoint foveation regions. Also,
DCT-domain based foveation using triangular-transition weights speeds up the encoder [10, 44].

Therefore, the real-time foveation technique proposed as an alternative method for obfuscation image defeats face recognition, while still maintaining its information in such a way that facial expression and gender can be classified effectively. Furthermore, in the privacy-awareness map for facial expression, the foveation method is slowly decreasing in classification accuracy with regard to the baseline.

3.3.6. Experiment 4: Additional Experiments using $(2D)^2$PCA

Additional experiments were performed using a new subset of the FERET database. In these experiments, the face images were cropped to a fixed size of $64 \times 64$ pixels and the number of projection vectors used for the $(2D)^2$PCA [45] method is controlled by the threshold, $\theta$, set to 0.95. The Frobenius norm is used as the similarity measure.

In this case, a new macro-block size, $b$, was set to $8 \times 8$. A small subset containing 200 subjects, 99 female and 101 male were randomly chosen.

Experimental results show that the new technique achieves a trade-off between privacy and awareness of a 30% of recognition rate and a classification accuracy as high as 80% (near the baseline) obtained by the privacy-awareness map. Thus, such an experiment reinforces the previous results obtained by experiments performed in above subsections and will be discussed below.

Performance Analysis of Privacy Protection

The performance of the privacy protection methods was analyzed using in this case a $(2D)^2$PCA-based algorithm [45] for face recognition. The recognition accuracies for every distortion method at different de-identification levels are shown in Figure 3.8. In this case, both pixelation and foveation recognition accuracies are below 30% at level 8 of de-identification. In contrast, blurring and eigenfeatures need up to 12 and 20 levels of de-identification, respectively. In the case of randomizing, recognition accuracies stays high at 80% for up to 20 levels of de-identification, meaning that this method is not suitable for anonymization. Despite using a de-identification level of 40%, foveation is able to preserve many facial characteristics for classification tasks such as gender and expression characterization, providing a good trade-off between
privacy and awareness.

![Figure 3.8: Face recognition rates for original unfiltered images in the gallery and de-identified images in the probe. It is evident declining of recognition accuracies each of these algorithms differ at different levels of de-identifying. For both foveation and pixelation recognition accuracies decrease up to 30% at the level 8. It can be seen that the Foveation curve falls more steeply. Randomizing is not able to decline in accuracy even with the highest levels of de-identification.](image)

**Performance Analysis of Awareness**

The de-identified images intelligibility is also quantified through a SVM classifier, as explained in section 3.3.3. According to these experiments, it is evident that the ad-doc de-identification methods such as foveation, blurring, eigenfeatures and random noise can actually preserve information at high obfuscation levels, as shown in Figure 3.9, where the preserved information is constant (greater than 80% near the baseline) for up to 18 levels of de-identification. The accuracy on original data is plotted as baseline. Therefore, foveation is comparable in performance over these algorithms but DCT-domain foveation is very efficient in terms of computational overhead [10]. Nevertheless, this does not imply that the subject’s identity is being protected as it will be discussed in the next section. From this insights, an analysis is done since there exist ad-hoc methods that can preserve information for various purposes at high obfuscation levels, whereas fail to prevent recognition. Data preserved for pixelation is considerably below, since information decreases from the first five levels.
3.3 Experimental Results

Figure 3.9: Data preservation assessment is achieved by measuring classification accuracy across all levels of de-identification. For both gender and facial expression classification tasks, foveation performs comparably over blurring at all different levels of de-identification, with higher classification rates at lower face recognition rates.

Privacy and Awareness Map

The awareness-privacy map is also obtained by plotting both gender and expression rates and recognition rates. Therefore, for gender and expression classification as data preservation function, DCT-domain foveation performs comparably over blurring, specially at lower face recognition rates than 30\%, as shown in Figure 3.10. Both
techniques result in the trade-off between privacy and awareness, but the foveation-based privacy approach is significantly faster than its spatial domain counterpart, so it requires much lower computation overhead since the cut-off frequency is constant over each of the foveation regions. Thus, the proper cut-off frequency is computed for each macro-block according to the viewing distance, and as a result, a large amount of operations required per pixel is saved.

Although foveation performs comparably over blurring at all different levels of de-identification for gender and expression recognition tasks, its performance of preserved data utility has proven to be more efficient than other methods, specifically at lower face recognition rates than 30%.

Discussion

The real-time foveation technique proposed as an alternative method for obfuscation image defeats face recognition while still maintains its information in such a way that facial expression and gender can be classified effectively. Notice that in the case of the privacy-awareness map for facial expression, foveation slightly increased the classification accuracy above the baseline. Furthermore, as it was previously mentioned, the DCT-domain foveation for privacy protection provides a computational speed-up because it can be incorporated into both DCT-based video codification and standard motion compensation techniques without requiring any modification at the decoder.

This section introduced a new DCT-domain foveation based method for face de-identification, which not only protects the individual’s identity but also preserves information which is used by a classifier for both gender and facial classification tasks. Also, we introduced a privacy-awareness map, which measures simultaneously the privacy protection and preserved information. Privacy protection is quantified by building both a PCA and (2D)^2PCA based standard face recognizers, while preserved information was measured by a SVM classifier.
Figure 3.10: The privacy-awareness map. Each data point represents privacy (i.e. recognition accuracy) and awareness (i.e. classification accuracy) for a given de-identification level in each of these algorithms.
4.1. Overview of Face Recognition Methods

Automated methods for face recognition starting in the 1960s with the pioneering work of Bledsoe [46]. The first fully functional automated face recognition system was not developed until 1977 by Kanade [47]. Since then, it has been established as a specific research area. Due to the relatively small amount of publications, it appears that during the 1980s the work in face recognition remained largely dormant.

On the one hand, the earliest methods [48–51], geometric facial features such as eyes, nose, mouth, and chin were used. This features and their relations, such as areas, distances and angles between these landmarks, were used as face descriptors for face recognition. Advantages of this method include the use of small amount of data and insensitivity to variations in illumination and viewpoint. However, facial features detection and measurement techniques developed to date are not sufficiently reliable for the geometric feature-based recognition [52].

On the other hand, statistical learning methods have been used in building current face recognition software, which use a training procedure based on appearance images. The appearance-based approach, such as PCA [4], and linear discriminant analysis (LDA) [7], has significantly advanced face recognition technology. Such approaches generally operate directly on the array of pixels with their corresponding light intensities (image-based representation) and extract features in a subspace derived from training images. PCA constructs an optimal subspace to represent the face object, whereas LDA constructs a discriminant subspace to distinguish faces of different person. LDA yields better results than PCA-based methods but, in scenarios where very few data are available, PCA has reported better results than LDA [53]. The
appearance-based methods require that the face images be properly aligned, typically based on the eyes locations.

In 1986 Sirovich and Kirby [54] proposed a method that became dominant in the image processing area in the following years, which is a technique based on the Principal Component Analysis (PCA). In their work, they achieve to reconstruct an image using a lower dimension than the original and yet, without losing essential information. The first application of PCA to face recognition appears in 1991 developed by Turk and Pentland and the technique is known as Eigenfaces or Karhunen-Loeve (KL) transform [4]. They demonstrated that the residual error when coding using the PCA could be used both to detect faces and to determine the precise localization and scale of faces in an image. This method not only became one of the most known methods for face recognition, but also it is the current baseline face recognition algorithm against which others are measured. Since then, face recognition has received plenty of attention marked by a noticeable increase in the number of publications and methods [55].

Due to the large appearance changes in human face images, the linear subspace methods may not capture the non-linearity in facial images representation, then as a result, the PCA and LDA algorithms have been extended to represent non-linear mappings in a higher-dimensional space [56]. Unfortunately, the use of a high-dimensional space increases computational costs. To make more efficient calculations, instead of finding the necessary features maps in high dimensions, kernels are used directly in the face high dimension spaces to calculate the needed inner products. All previous experiences have produced an improvement in performance over linear face recognition approaches when combining linear subspace methods with the kernel techniques [57–61]: the kernel PCA (KPCA), the kernel discriminant analysis (KDA), and the kernel ICA.

Another way to deal with the characteristic non-linearity, of the faces space, is to construct a local appearance space using suitable filters in order to avoid affecting the face appearance, despite differences in facial expression and head orientation. One of the most successful system in this category is the graph matching system [62] based on the Dynamic Link Architecture (DLA) [63]. An image graph representing a face image is a geometrical structure consisting of various nodes connected by edges. The nodes are located at facial landmarks such as the pupils and the corners of the mouth. A set of training images is represented by the corresponding bunch of image graphs.
of those images. A set of complex Gabor wavelet coefficients (or Gabor jets) are used as local features at each node. These Gabor jets contain information of multiple orientations and frequencies for each node. These locally estimated coefficients are robust to illumination change, translation, distortion, rotation, and scaling. When performing face recognition on a new facial image, each graph in the training set is matched to the image and the best match indicates the identity of the person. Systems based on the EBGM approach have been applied to face detection and extraction, pose estimation, gender classification, sketch-image-based recognition, and general object recognition. The success of the EBGM system may be due to its resemblance to the human visual system [64]. Their method is also fully automatic, including face localization, landmark detection, and flexible graph matching. The drawback of this method is its requirement for accurate landmark localization, which is not an easy task, especially when illumination variations are present.

Another actively researched approach to face recognition is that of Advance Correlation Filters (ACFs), which initially was developed and applied in the field of Automatic Target Recognition (ATR). The ACF approach processes images in the frequency domain using correlation filter solutions designed for specific optimization criteria [65, 66]. Despite their capabilities, ACFs are still less well known than the above mentioned algorithms in the field of biometrics [3].

The simplest correlation filter, commonly is used in applications such as communication channels and radar receivers [67], is the Matched Filter (MF), where the goal is detecting a reference pattern or signal corrupted by additive white noise. MF is characterized by maximizing the Signal-to-Noise Ratio (SNR) of the response of the desired signal and minimizing the effects of noise. But MF is optimal only if the reference pattern and the input pattern are identical except for the additive white noise and translation. In current applications for face recognition, the test pattern will differ from the reference pattern in many ways, e.g., rotations, scale changes, illumination, and hence the MF does not perform well [68].

In face recognition, the test face image from a subject is bound to differ from the reference face image of the same subject due to variations induced by expression changes, illumination differences, pose variations and aging. It is evident that the computational cost is impractical or prohibitive because of the number of filters used when we consider all possible factors that cause the face appearance to change.

One of the first CFs to incorporate such a composite design, of handing pattern
variability using fewer correlation filters, is the *Synthetic Discriminant Function Filter* [69], which is able to recognize different patterns or classes using a single filter as opposed to using a single filter for each class or image sample (as with the case of MFs). This SDF filter is a weighted sum of MFs where the weights are chosen so that the correlation outputs corresponding to the training images would yield prespecified values at the origin. For example, the correlation values (at the origin) corresponding to the training face images of authentic subjects can be set to one, and the origin values due to the impostor training images can be set to zero. The resulting correlation filter is expected to yield correlation peak values close to one for nontraining face images from the authentic class and correlation peak values close to zero for images from the impostor class.

Although the original SDF filter produces prespecified correlation peak values, it only controls the output at the origin for centered training images. Since the test patterns are not necessary centered, it is nearly impossible to know where these controlled values in the output are, unless the rest of the correlation plane is controlled to take on smaller values. Thus, one of the most often used correlation filters is the minimum average correlation energy (MACE) filter [70], which minimizes the average correlation plane energy resulting from training images, while constraining the correlation peak value at the origin to prespecific values. Correlations outputs from MACE filters typically exhibit sharp peaks, making the peak detection and location relatively easy and robust. The optimal-trade-off synthetic discriminant function (OTSDSF) filter [71] is an extension of the MACE filter to achieve robustness to additive noise. MACE and OTSDF are constrained to have a certain value for the inner product between the training image and the filter. This inner or dot product value is referred to as the value at the origin (for centered images) or the correlation peak in the correlation plane.

Another advance in CF designs was removing the correlation peak constraints. Removing these constraints increases the solution space and may improve the chances of finding a filter with better recognition performance. The family of unconstrained CFs include the maximum average correlation height (MACH) filter [72], the unconstrained MACE (UMACE) filter [72]; the minimum output sum of squared error (MOSSE) filter [73] and the quadratic correlation filters (QCFs) [74] that determine and use a quadratic nonlinearity to maximize a metric of separation between authentic and impostor classes. Another advance were optimal tradeoff circular harmonic
function filters (OTCHF) [75] that can handle in-plane rotation. Since the face images are mapped into polar coordinates, in-plane rotations appear as shifts in the polar domain and shifts can be handled by correlation filters, which produce large correlation peaks for the target rotated between $-\pi/4$ and $\pi/4$. A deeper study of these correlation filters can be found in [76].

Whereas the conventional correlation filters perform directly on image pixels, other methods, such as features correlation filter (FCF) [77] are applied into feature spaces (PCA and Gabor wavelets) while preserving the benefits of conventional correlation filters, i.e., shift-invariant and occlusion-insensitive. Another variation of MACE filters is the introduction of a linear subspace learning method called class-dependence feature analysis (CFA) [78–80]. Different from traditional linear subspace learning methods, the projection axis obtained by CFA tries to discriminate one specific class from all other class. Different projection axes concern different classes. CFA is based on the design of an advanced correlation filter technique which emphasizes the outputs of one face class and suppresses the outputs of other faces classes. Different correlation filters can be designed according to different criterions [81].

It can be noted from the literature review that several different types of correlation filters have been proposed for face recognition when illumination variations are present [77,82–85], including the hybrid shift-invariant PCA-correlation filter approach, termed Corefaces [86], that uses the subspace representation of PCA and shift-invariant and discrimination properties of advanced correlation filters. Other ACFs are also applied in general face recognition [68, 87], large scale face recognition [79, 88, 89], multi-modal face recognition [90], to PDA/cell-phone based face recognition [91, 92].

Several appearance-based face recognition methods, such as Principal Component Analysis (PCA), two-dimensional PCA (2DPCA) and Linear Discriminant Analysis (LDA) used in the frequency domain, have proven to be far more robust than their spatial domain counterparts. By using the PCA subspace of the phase-only spectra, Savvides et al [86] proposed a hybrid PCA-correlation filter for illumination-invariant face recognition from still images. In contrast, Bhagavatula et al [93] proposed to perform PCA and FLDA using the Eigen and Fisher-Fourier Magnitudes, the resulting FM-PCA and FM-FLDA subspaces are shift-invariant and are not prone to registration errors of the input image. To achieve a higher recognition accuracy of these algorithms, 2DPCA based algorithms such as FM-2DPCA, FM-(2D)^2PCA and
FM-DiaPCA were proposed by Zeytunlu et al [94]. Ribarić et al [95] presented four phase-information extraction approaches where the MagUn approach achieves the best recognition when illumination variation is present in the dataset. However, this is not the case when pose and expression variations were present.

Li et al [96] proposed a method which involves using a hybrid-PCA subspace together with illumination tolerant correlation filters to reconstruct and recognize images, respectively, using a different representation of the face. In their method, they synthesize a realistict face image from the composite sketch using this hybrid subspace. Then, the face synthesized is used in real-time applications for law-enforcement officers looking for criminals. Sao et al [97] proposed to perform eigenanalysis using the cosine and sine functions of both, the phase spectra of the Fourier transform and of the corresponding analytic image to avoid the phase wrapping problem. Benitez-Garcia et al [98] proposed a sub-block based eigenphases algorithm where the face image under analysis is divided into optimal sub-blocks, and then all phase spectra are computed. The resulting hybrid-PCA subspace is obtained by concatenating the phase spectra of all blocks. Inspired by CFA [68,78,99], an effective feature extraction method also based on multi-subregion, called Multi-Subregion based correlation filter bank (MS-CFB), is proposed by Yan [100], which combines the benefits of global based and local-based feature extraction algorithms. It performs multiple correlation filters corresponding to different face subregions and are jointly designed to optimize the overall correlation outputs. MS-CFB not only takes the differences among face subregions into account, but also effectively exploits the discriminative information in face subregions. However, the multi-block strategy can not handle face recognition with large pose variations; i) some face subregions contain undesired background mainly caused by pose changes, which degrades the discriminability of features extracted by MS-CFB, since it is based on the sum of the correlation outputs from all face subregions), and ii) the mismatching of face subregions between training samples and test samples can occur when dealing with large pose variations.

Thus, if the subspace of the phase and magnitude spectrums are separately used, then an increase in recognition accuracy can be accomplished over the corresponding subspace in the spatial domain.
4.2. Face Recognition in Subspaces

Face images, represented as high-dimensional pixel arrays, often belong to a manifold of intrinsically low dimension. This observation has allowed to apply algebraic and statistical tools for extraction and analysis of the underlying manifold. In this section, we describe some of techniques that analyze linear and nonlinear subspaces from the original Eigenfaces technique. We also discuss a comparative experimental evaluation of some of these techniques.

Face images may be represented as high-dimensional pixel arrays where the pixels may encode color, only intensity or both. In this research, we assume the latter case (gray-level imagery). After proper normalization and resizing to a fixed \(m\)-by-\(n\) size, the pixel array can be represented as a point (i.e., vector) in an \(mn\)-dimensional image space by simply writing its pixel values (typically raster) in a fixed order. The original space of an image is just one of infinitely many spaces in which the image can be examined. A critical issue in the analysis of such multidimensional data is the dimensionality, that is, the high number of coordinates necessary to specify a data point, even for images of modest size.

Recognition methods that operate on the multidimensional vector space representation suffer from a number of potential disadvantages, most of them rooted in the so-called curse of dimensionality:

- Handling high-dimensional examples, especially in the context of similarity- and matching-based recognition, is computationally expensive.

- The number of parameters to estimate, for parametric methods, typically grows exponentially with the dimensionality. Often such number is much higher than the number of images available for training, making the estimation task in the image space ill-posed.

- Similarly, for nonparametric methods, the sample complexity - the number of examples needed for an efficient representation of the underlying distribution of the data efficiently - is prohibitively high.

The value of a pixel is typically highly correlated with the values of the surrounding pixels. Moreover, the appearance of faces is highly constrained; for example, any frontal view of a face is roughly symmetrical, has eyes on the sides, nose in the
middle, and so on. A vast proportion of the points in the image space does not represent physically possible faces. Thus, the natural constraints dictate that the face images are in fact confined to a subspace referred to as the face subspace.

4.2.1. Principal Component Analysis

Principal Component Analysis (PCA) is a dimensionality reduction technique based on extracting the desired number of principal components of the multidimensional data. The first principal component is the linear combination of the original dimensions that has the maximum variance, then the $n$th principal component is the linear combination with the next highest variance, subject to requirement of being orthogonal to the $n-1$ first principal components.

The idea of PCA is illustrated in Figure 4.1, where the axis labeled $\phi_1$ corresponds to the direction of maximum variance and is chosen as the first principal component. In a two-dimensional case, the second principal component is then determined unequivocally by using the orthogonality constraints; in a higher-dimensional space the selection process would continue, guided by the variance of the projections.

The PCA, also known Karhunen-Loève Transform (KLT), with the basis $\phi = [\phi_1, \ldots, \phi_N]^T$ that for any $k \leq N$ minimizes the average $L_2$ reconstruction error for data points $x$

$$\varepsilon(x) = \left| x - \sum_{i=1}^{k} (\phi_i^T x) \phi_i \right|.$$  \hspace{1cm} (4.2.1)

Without loss of generality, we hereafter assume that the data are indeed zero-mean; that is, the mean face $\bar{x}$ is always subtracted from the data.

The basis vectors in KLT can be calculated in the following way. Let $X$ be the $N \times M$ data matrix whose columns $x_1, \ldots, x_M$ are observations of a signal embedded in $\mathbb{R}^N$; in the context of face recognition, $M$ is the number of available face images, and $N = mn$ is the number of pixels in an image. The KLT basis $\phi$ is obtained by solving the eigenvalue problem $\lambda = \phi^T \Sigma \phi$, where $\Sigma$ is the covariance matrix of the data

$$\Sigma = \frac{1}{M} \sum_{i=1}^{M} x_i x_i^T$$  \hspace{1cm} (4.2.2)

and $\phi = [\phi_1, \ldots, \phi_N]^T$ is the eigenvector matrix of $\Sigma$, $\lambda$ is the diagonal matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$ of $\Sigma$ on its diagonal; so $\phi_j$ is the eigenvector corresponding
to the $j$th largest eigenvalue. Then it can be shown that the eigenvalue $\lambda_i$ is the variance of the data projected on $\phi_i$.

Thus, to perform PCA and extract $k$ principal components of the data, one must project the data onto $\phi$, the first $k$ columns of KLT basis $\phi$, which correspond to the $k$ highest eigenvalues of $\Sigma$. This can be seen as a linear projection $\mathbb{R}^N \rightarrow \mathbb{R}^k$, which retains the maximum energy (i.e. the variance) of the signal. Another important property of PCA is that it decorrelates the data: the covariance matrix of $\phi_k^T X$ is always diagonal.

The main properties of PCA are the approximate reconstruction $x \approx \phi_k y$, orthonormality of the basis $\phi_k$, and decorrelated principal components $y_i = \phi_i^T x$. This properties of PCA are illustrated in Figure 4.1, where PCA successfully finds the principal manifold, but in contrast it becomes less successful due to nonlinearity of the principal manifold [1], as shown in Figure 4.2.

A brief overview of the concept of PCA-based face recognition is as follows; once a face subspace is created, all training images are projected into this subspace. Next each test image is projected into this subspace. Later, each test image is compared to all the training images by a similarity measuring distance. Finally, the most similar training image found (or the closest to the test image) is used to identify the test image.

Projecting images into subspaces has been studied for many years, and as result has helped to revolutionize algorithms for face recognition. An in-depth analysis of
techniques applied in PCA can be found in [101], where as a result of the study, it specifies the conditions to find and select a particular subspace. It indicates how the selection may improve performance in face recognition as well as it discusses on how some variations within the subspace may affect such performance; and presents the different distance measurements used.

Figure 4.3: Visualization of eigenfaces. The average face on the left followed by eleven top eigenfaces from the ORL face database.

4.2.2. Eigenfaces and Related Techniques

The main idea behind of eigenfaces arises from the work by Kirby and Sirovich [54] for analysis and representing of human faces into a low-dimensional space by using PCA. Their work was follow by Turk and Pentland [54] for face recognition through PCA. Because the basis vectors constructed by PCA had the same dimension as the input face image, they were named “eigenfaces”.
The procedure to obtain eigenfaces starts with finding the mean face, then subtracting it from the whole data set of faces. Later, each face was projected into the principal subspace, and the coefficients of PCA expansion were averaged for each subject. The result was a single \textit{k-dimensional} representation of a given particular subject. Figure 4.3 shows an example of the mean face and few of the top “eigenfaces”.

When a test image is projected into the subspace, Euclidean distances between its vector coefficients and those representing each subject were computed. Then each test image can be classified as belonging to one of the familiar subjects, as a new face, or nonface by evaluating the similarity or closeness to the training image and the PCA reconstruction error (Eq. 4.2.1). Moreover, when the appearance of an object class (e.g., faces) is modeled by a subspace, the distance from this subspace can serve to classify an object as a member or a nonmember of the class (subspace techniques for detection).

### 4.2.3. Linear Discriminant: Fisherfaces

PCA has been intensively exploited in face recognition methods, but many other linear projection methods have been studied too. It well-known that variation in data is due to presented changes in illumination and facial expression. The PCA technique selects a subspace that retains most of that variation, and consequently the similarity in the face subspace is not necessary determined by the identity.

Belhumeur \textit{et al} [7] proposed to solve this problem with “Fisherfaces”, which provides discrimination among the classes while PCA deals with the input data in their entirety, without paying any attention for the underlying structure [102]. Fisher’s linear discriminant (FLD) selects the linear subspace \( \phi \), which maximizes the ratio

\[
\phi_{opt} = \arg \max \frac{|\phi^T S_b \phi|}{|\phi^T S_w \phi|} \tag{4.2.3}
\]

where

\[
S_b = \sum_{i=1}^{m} N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \tag{4.2.4}
\]

is the \textit{between-class} scatter matrix, and

\[
S_w = \sum_{i=1}^{m} \sum_{\mathbf{x} \in X_i} (\bar{x} - \bar{x}_i)(\bar{x} - \bar{x}_i)^T \tag{4.2.5}
\]
is the within-class scatter matrix; \( m \) is the number of subjects (classes) in the database. Thus, in essence, the procedure of FLD consists on finding the base of vectors proving the best discrimination among the classes, while maximizing between-class differences and minimizing the within-class ones. That is, FLD finds the projection of the data in which the classes are most linearly separable. The dimension of \( \phi \) is at most \( m - 1 \).

![Figure 4.4: Visualization of fisherfaces. Twelve top fisherfaces from the ORL face database.](image)

Because in practice \( S_w \) is usually singular, the FLD algorithm first reduces the dimensionality of the data using PCA, and then FLD is applied to the resulting space to further reduce the dimensionality to \( m - 1 \). Figure 4.4 shows the Fisherfaces corresponding to the projection matrix \( \phi \).

Recognition is then accomplished by a NN classifier in this final subspace. Experiments reported by Belhumeur et al [7] were performed on data sets containing frontal face images of 5 people with drastic lighting variations and another set with faces of 16 people with varying expressions and again drastic illumination changes. In all the reported experiments Fisherfaces achieve a lower error rate than eigenfaces. Indeed, FLD provides better classification performances only when a wide training set is available [103]. However, it has shown in [104, 105] that, combing PCA and FLD, discriminant information together with redundant one is discarded. In these cases, a Linear Discriminant Analysis (LDA) [103] is directly applied on the input space [104, 105].

A variant of LDA, such as D-LDA (Direct LDA) and F-LDA (Fractional LDA), uses weighed functions to avoid that output classes, which are too close, can induce misclassification of the input. However, the main disadvantage of the PCA and LDA are their linearity. The PCA method extracts a low-dimensional representation of the
input data only exploiting the covariance matrix, so that no more that first- and second order statistics are used [102]. Bartlett Marian et al [6] showed that first- and second order statistic hold information only about the amplitude spectrum of an image, discarding the phase-spectrum; while some experiments bring out the human capability to recognize objects is mainly driven by the spectrum-phase [102].

**Independent Component Analysis**

Independent Component Analysis (ICA) can be considered as a generalization of the PCA, minimizing the sample covariance (second-order dependence) of the data and minimizing higher-order dependencies as well [106]. Thus, ICA provides three main advantages over PCA:

1. It allows a better characterization of data in an $n$-dimensional space.
2. The vectors found by the ICA are not necessarily orthogonals, making possible to reduce the reconstruction error.
3. The captured discriminant features not only exploit the covariance matrix, but also consider higher-order statistics.

ICA yields a linear projection $\mathbb{R}^N \rightarrow \mathbb{R}^M$ in a similar way as PCA does, but with properties such as approximated reconstruction $x \approx Ay$, nonorthogonality of the basis of $A$, that is, $A^T A \neq I$, and the near-factorization of the joint distribution $P(y)$ into distributions of the (non-Gaussian) ICs.

An example of ICA basis is shown in Figure 4.5 for a set of 3D points. It can be observed that the 2D subspace recovered by ICA appears to reflect the distribution of the data much better than the subspace obtained with PCA [1]. However, despite that ICA perform slightly better than PCA [38], it is questioned whether it is worth using ICA given its greater computational complexity.

**4.2.4. Two-Dimensional Two-Directional PCA**

When PCA is used for face recognition, the 2D image matrices must be previously transformed into 1D image vectors by vectorization, which usually lead to a high-dimensional image vector space. This makes difficult to evaluate the covariance matrix accurately due its large size and the relatively small number of training samples.
Furthermore, two difficulties are present: computing the eigenvectors of a large size covariance matrix is very time-consuming, and when computing the covariance matrix, it will be well estimated if and only if the number of available training samples is not far smaller than the dimension of this matrix. In fact, in real applications, it is too hard to collect enough number of samples, making the estimate of the covariance matrix, in 1D subspace analysis, not well estimated and not full rank.

Although eigenfaces can be calculated efficiently using the SVD techniques, while avoiding the process of generating the covariance matrix, this does not imply that the eigenvectors can be evaluated accurately in this way since the eigenvectors are statistically determined by the covariance matrix, no matter what method is adopted for obtaining them [107].

To tackle the previously described problems, a straightforward image projection technique called two-dimensional principal component analysis (2DPCA) was proposed by Yang et al [107] for image feature extraction. As opposed to PCA, 2DPCA is based on 2D matrices rather than 1D vectors. That is, the matrix image is not previously transformed into a vector. Instead, an image covariance matrix is directly built using the original matrices in such a way that the spatial structure information can be preserved. Because the size of the image covariance matrix is equal to the width of images, which is quite small compared to the covariance matrix size in PCA, 2DPCA evaluates the image covariance matrix more accurately and computes the corresponding eigenvectors more efficiently than PCA. Consequently, the image covariance matrix will usually be full rank, and thus the curse of dimensionality and the Small Sample Size (SSS) problem can be avoided. Evidently, the experimental results on several face databases reported in [9, 45, 107] have shown the improvement.
of 2DPCA over PCA.

In the next section we review the details of 2DPCA [107] and its extensions to a \((2D)^2PCA [45]\).

**Two-Dimensional Principal Component 2DPCA**

Given an \(m \times n\) image matrix, let us consider a linear transformation that maps original matrix \(A\) into an \(m\)-dimensional projected feature vector called principal component vector, then the new matrix \(Y \in \mathbb{R}^{m \times d}\) is:

\[
Y = AV
\]  

(4.2.6)

where \(V \in \mathbb{R}^{n \times d}\) is a matrix with orthonormal columns and \(n \geq d\).

The 2DPCA method, likewise PCA, searches for the optimal projection by maximizing the total scatter of the projected data. In 2DPCA, the total scatter of the projected samples were used to determine a good projection matrix and it can be characterized by the trace of the covariance matrix of the projected features vectors. That is, the following criterion was adopted:

\[
J(V) = \text{trace}\{E[(Y - EY)(Y - EY)^T]\}
\]

\[
= \text{trace}\{E[(AV - E(AY))(AV - E(AV))^T]\}
\]

\[
= \text{trace}\{E[((A - EA)V)((A - EA)V)^T]\}
\]

\[
= \text{trace}\{V^T E[(A - EA)^T(A - EA)]V\}, \quad (4.2.7)
\]

where the last term in Eq. 4.2.7 results in fact that \(\text{trace}(AB) = \text{trace}(BA)\) for any two matrices [108]. Thus, the total power equals to the sum of the diagonal elements or trace of the covariance matrix.

Let us define the image covariance matrix \(C_r = E[(A - EA)^T(A - EA)]\), which is an \(n \times n\) nonnegative definite matrix. Suppose that there are \(M\) training face images, denoted by \(m \times n\) matrices \(A_k = 1, 2, \ldots, M\), and the mean face of all training samples is defined as follows:

\[
\bar{A} = \frac{1}{M} \sum_k A_k
\]  

(4.2.8)
Then $C_r$ can be evaluated by

$$C_r = \frac{1}{M} \sum_{k=1}^{M} (A_k - \bar{A})^T (A_k - \bar{A}). \quad (4.2.9)$$

Now, the alternative criterion in 4.2.7 can then be expressed by

$$J(V) = \text{trace}(V^T C_r V) \quad (4.2.10)$$

where $V$ is a unitary column vector. This alternative criterion is called the generalized total scatter criterion. The unitary vector $V$ that maximizes the criterion is called the optimal projection axis. This means that the total scatter of the projected samples is maximized after the projection of an image matrix onto $V$.

The optimal projection axis $V_{opt}$ is the unitary vector that maximize $J(V)$ is defined as follows:

$$V_{opt} = \{V_1, \ldots, V_d\} = \arg \max \text{tr} (V^T C_r V) \quad (4.2.11)$$

where the optimal projection axes $\{V_1, \ldots, V_d\}$ are the orthonormal eigenvectors of $C_r$ corresponding to the first $d$ largest eigenvalues. In general, it is not enough to have only one optimal projection axis, and therefore a set of projection axes are needed. Because the size of $C_r$ is only $n$ by $n$, computing its eigenvectors is very efficient. Like PCA, the number of eigenvectors can be controlled by

$$\sum_{i=1}^{d} \lambda_i \geq \theta. \quad (4.2.12)$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the $n$ largest eigenvalues of $C_r$ and $\theta$ is a pre-set threshold value. Each image $A$ is projected onto these $d$ dimensional subspace according to Eq. (4.2.6) yielding an $m \times d$ matrix $Y = AV$.

Although 2DPCA obtains higher recognition accuracy than PCA, a vital unresolved problem is that it needs many more coefficients for image representation than PCA. For example, let us consider images to a fixed size of $100 \times 100$ pixels, then the number of coefficients of 2DPCA is $100 \times d$, where $d$ is set to no less than 5 for efficient face representation and recognition. Although this problem can be alleviated by using PCA after 2DPCA for further dimensional reduction, it is still unclear how
the dimension of 2DPCA could be reduced directly [107].

Alternative 2DPCA

Equation 4.2.9 is deemed as the original 2DPCA which is working in the row direction of images. An alternative way of 2DPCA can be obtained by applying 2DPCA on the column direction of images [45] as follows:

\[
C_c = \frac{1}{M} \sum_{k=1}^{M} (A_k - \bar{A})(A_k - \bar{A})^T.
\]

(4.2.13)

Similarly to Eq. 4.2.6, projecting the original matrix \(A\) onto \(U\) yields a \(q \times n\) matrix \(B \in \mathbb{R}^{q \times n}\) as

\[
B = U^T A
\]

(4.2.14)

where \(U \in \mathbb{R}^{m \times q}\) is a matrix with orthonormal columns and \(m \geq q\).

The same criterion is adopted to find the optimal projection matrix \(U\):

\[
J(U) = \text{trace} \left\{ E[(B - EB)(B - EB)^T] \right\}
\]

\[
= \text{trace} \left\{ E[(U^T A - E(U^T A))(U^T A - E(U^T A))^T] \right\}
\]

\[
= \text{trace} \left\{ U^T E[(A - EA)(A - EA)^T] U \right\},
\]

(4.2.15)

Similarly, the optimal projection axis \(U_{opt}\) is chosen to maximize the trace of the scatter matrix \(C_c\):

\[
U_{opt} = \{U_1, \ldots, U_d\} = \text{arg max} \text{ tr} (U^T C_c U)
\]

(4.2.16)

where the optimal projection axes \(\{U_1, \ldots, U_d\}\) are the orthonormal eigenvectors of \(C_c\) corresponding to the first \(q\) largest eigenvalues. The eigenvectors can also be controlled by setting a \(\theta\) threshold values using Eq. (4.2.12). Because the eigenvectors of Eq. (4.2.13) only reflect the information between columns of image, the alternative 2DPCA is said to be working in the column direction of images.

Finally, by simultaneously projecting an \(m \times n\) image \(A\) onto the projection matrices \(V\) and \(U\), an \(q \times d\) feature matrix \(C\) is yielded

\[
C = U^T A V
\]

(4.2.17)
where \( C \) is also called the coefficient matrix in image representation, and it can be used to reconstruct the original image \( A \), as

\[
\tilde{A} = UCV^T \quad (4.2.18)
\]

Figure 4.6 shows one reconstructed image at different dimensions using the FERET database. It can be observed that, when selecting a set of projection axes for representing the image sample (i.e., \( 55 \times 54 \)), the quality of the face image is improved.

Therefore, bilateral projection methods use two projection matrices for both rows and columns. The former computes these projections separately while the latter computes them simultaneously via iterative process. A deeper study of these techniques (i.e., non-iterative and iterative) and their variations (i.e., 2DLDA, B2DPCA, GLRAM, BDPCA, CSA, K2DPCA), can be found in [109].

![Figure 4.6: Visualization of reconstructed training images gotten by (2D)^2 PCA under different dimensions.](image)

### 4.2.5. Kernel-PCA Method

Since much of the important information may be contained in the high order dependences among the pixels of a face image, Bernhard et al [110] proposed a method, performing a nonlinear form of principal component analysis, that not only extracts higher order statistics of samples as features, but also maximizes the class separation when these features are projected to a lower dimensional space for efficient recognition.

The basic methodology of KPCA is to apply a nonlinear mapping to the input \( \psi(x) : \mathbb{R}^N \rightarrow \mathbb{R}^L \) and then solve for a linear PCA in the resulting feature space \( \mathbb{R}^L \), where \( L >> N \) and possibly infinite. Because of this increase in dimensionality,
the mapping $\psi(x)$ is made implicit (and economical) by the use of kernel functions satisfying Mercer’s theorem [1].

$$k(x_i, x_j) = [\psi(x_i) \cdot \psi(x_j)] \quad (4.2.19)$$

where kernel evaluations $k(x_i, x_j)$ in the input space correspond to dot-products in the higher dimensional feature space. Since computing covariance is based on dot-products, performing a PCA in the feature space can be formulated with kernels in the input space without the explicit (and possibly expensive) direct computation of $\psi$. The covariance (assuming that the projected data into feature space has zero-mean) is given by

$$\Sigma_K = \left(\psi(x_i), \psi(x_j)^T\right) \quad (4.2.20)$$

with the corresponding eigenvalue problem $\lambda V = \Sigma_K V$. Since the eigenvectors (columns of $V$) must lie in the span of the training data $\psi$, it must be true that for each training point

$$\lambda(\psi(x_i)V) = (\psi(x_i) \cdot \Sigma_K V) \text{ for } i = 1, \ldots, T \quad (4.2.21)$$

and, there exist coefficients $w_i$ such that,

$$V = \sum_{i=1}^{T} w_i \psi(x_i). \quad (4.2.22)$$

Substituting Eq. 4.2.22 into 4.2.21 and defining an $T \times T$ matrix $A$ by $A_{i,j} = [\psi(x_i) \cdot \psi(x_j)]$ leads to the equivalent eigenvalue problem formulated in terms of kernels in the input space

$$T\lambda w = Kw \quad (4.2.23)$$

where $w = (w_1, \ldots, w_T)^T$ is the vector of expansion coefficients of a given eigenvectors $V$ as defined in Eq. 4.2.22. The kernel matrix $K_{i,j} = k(x_i, x_j)$ is diagonalized with a standard PCA. Orthonormality of the eigenvectors, $(V^n \cdot V^n) = 1$, leads to the equivalent normalization of their respective expansion coefficients, $\lambda_n(w^n \cdot w^n)$.

The KPCA principal components of any input vector can be efficiently computed with simple kernel evaluations against the dataset. The $n$th principal component $y_n$ of $x$ is given by

$$y_n = (V_n \cdot \psi(x)) = \sum_{i=1}^{T} w_i^n k(x, x_i) \quad (4.2.24)$$
where $V_n$ is the $n$th eigenvector of the feature space defined by $\psi$. Like a PCA, the eigenvectors $V_n$ can be ranked by decreasing order of their eigenvalues $\lambda_n$ and a $d$-dimensional manifold projection of $x$ is $y = [y_1, \ldots, y_d]^T$, with components defined in Eq. 4.2.24. In KPCA we can use more eigenvectors projections than the input dimensionality of the data, since KPCA is based on the matrix $K$, the number of eigenvectors or features available is $T$. Typical kernels include Gaussian, $exp(||x_i - x_j||^2/\sigma^2)$, polynomials $(x_i \cdot x_j)^d$, and sigmoids $\tanh(a(x_i \cdot x_j) + b)$, all of which satisfy Mercer’s theorem [1]. The kernel Fisherfaces algorithm can be derived in a similar way like KPCA [], which maximizes the between-scatter to within-scatter ratio in the feature space through the use of the kernel matrix $K$. This algorithm, using face images with varying pose, scale, and illumination, showed performance superior to that of ICA, PCA, KPCA, and Fisherfaces.

This section has showed some of subspace methods to represent humane faces onto a low-dimensional subspace for face recognition. Thus, this low dimensionality allows a face recognition system to simplify computations and to focus the attention on the features of the data relevant to identify a person.
4.3. Correlation Filters

Initially, Correlation Filters (CFs) were applied in the field of automatic target recognition and pattern recognition in general [65]. Later, CFs were also applied in biometric authentication and in more specifically in the field of face recognition [68,77,79,82–92].

Correlation, also known as sliding dot product or sliding inner-product, is a metric to measure the similarity between a reference pattern \( r(x,y) \) and a test pattern \( t(x,y) \). The cross-correlation \( c(\tau_x, \tau_y) \) between the two patterns for various possible shifts is given by:

\[
C(\tau_x, \tau_y) = \int \int t(x,y)r(x-\tau_x,y-\tau_y)dxdy
\]  

(4.3.1)

where the limits of integration are based on the support of \( t(x,y) \) and the parameters \( \tau_x \) and \( \tau_y \) called lag indicate the time-shift between the reference and the test patterns.

Specifically, if the input \( t(x,y) \) begins compared contains a shift version \( r(x-\tau_x, y-\tau_y) \) of the reference signal, the correlator will exhibit a peak at \( x = x_0 \) and \( y = y_0 \). If the input does not contain the reference pattern \( r(x,y) \), the correlator output will be low.

When CFs are applied in live image recognition applications, the reference pattern may be the face image stored on a smart card while the test images may be the set of face images that are captured by a CCTV system.

In the pattern recognition area, a CF tries to obtain the similarity of a test object against a set of training objects and in fact, it works fine when the test object almost matches with one in the training set. Some advanced techniques of CF (ACF) were considered in order to improve the use of the basic correlation, involving extra attributes such as tolerance to real-world differences, distortions (such as image rotations, scale changes, illumination variations, etc.), and to achieve discrimination from sets of other classes [111]. One of the fundamental differences between CFs and ACF is the ability to synthesize ACF using multiple instances of training data (multiple facial images for the case of face recognition). This quality is extremely desirable in face recognition because the human face is subject to numerous variations, both intrinsic and extrinsic. Thus, when ACFs allow variations or, at least to be partially represented through the use of proper training data, face recognition systems can increase performance and robustness.
4. Face Recognition

Also, ACFs have significant tolerance to changes in illumination even if the test images have different level of illumination than the training images, and even further, they can successfully achieve high recognition rates without the need to re-train the classifiers. The performance of ACF in face recognition is reported to be significantly much higher compared to traditional approaches such as nearest-neighbor methods (1-NNM) [67].

4.3.1. Correlation pattern recognition for biometrics

The use of CFs starts by defining a reference image \( h(m,n) \) and a test image \( x(m,n) \) and proceed to calculate the 2D discrete correlation \( c(m,n) \), resulting in a 2D correlation plane \( c(m,n) \):

\[
c(m,n) = x(m,n) \otimes h(m,n) \tag{4.3.2}
\]

\[
= \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} x(u,v)h(u,v)
\]

\[
= \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} x(m+n,v)h(u-v)
\]

The advantage of adding all elements in Eq. (4.3.2) is that no single pixel in the test image by itself is critical to forming the correlation output, and further it results in a graceful degradation [111] that can be thought as smoothing filter whose impulse response is precisely \( h(m,n) \).

Using the Fourier transform and its properties we can express Eq. (4.3.2) in the frequency domain as

\[
F^{-1} = \{F(u,v).H^*(u,v)\} \tag{4.3.3}
\]

where \( F(u,v) \) and \( H(u,v) \) are the 2D Fourier transform of \( f(m,n) \) and \( h(m,n) \) respectively. The symbols \( F^{-1} \), \( . \), and \( * \) represent the inverse Fourier transform, the element by element multiplication of the 2D 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 of floating point operations required.

Thus, correlation pattern recognition (CPR) is based on the standard 2D Discrete Fourier Transform (DFT) pair. This fundamental concept starts from both, a 2D
4.3 Correlation Filters

discrete input signal $f(m, n)$ of size $M \times N$ and its corresponding Fourier Transform as $F(u, v)$. Then the Fourier transform pair $f[m, n] \xleftrightarrow{F} F[u, v]$ is defined as:

$$
F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n)e^{-j2\pi um/M}e^{-j2\pi vn/N} \quad (4.3.4)
$$

$$
f(m, n) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(u, v)e^{j2\pi um/M}e^{j2\pi vn/N} \quad (4.3.5)
$$

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

Figure 4.7: Block diagram showing the correlation process. Face recognition is performed by filtering the input face image with a synthesized correlation filter and processing the resulting correlation output.

Figure 4.7 shows schematically how the cross-correlation via Fourier transform for face verification is performed. During enrollment, a few face images of the authentic user are acquired. Afterwards, the FTs of all these images are then used to generate a 2D filter. Then, this correlation filter bank is stored (perhaps on a smart card) to identify that authentic user. In recognition mode, when this authentic user presents his face image to the live CCTV during verification, the 2D FFT of his live face image is multiplied pixel-wise by the stored filter bank. The correlation output array is obtained through the use of the 2D inverse FFT, then a search for peaks is performed,
4. Face Recognition

and the relative heights of these peaks are used to determine whether the object of interest is present or not.

When a test image perfectly matches with the filter, the correlation output exhibits a sharp peak for the authentic user and no such discernible correlation peak for an impostor, as illustrated in Figure 4.7. Specifically, when the target scene matches the reference image exactly, the output is the autocorrelation of the reference image.

The shifting of the image also shifts the correlation peak to a new position without affecting the peak sharpness (graceful degradation). Therefore, advanced correlation filters offer many attractive qualities such as shift invariance, normalized outputs, and noise tolerance. ACFs when correlated with an image result in a correlation plane $C$ which measures the correlation between the filter and the image. Correlation of a class-specific filter with authentic and impostor data yields very different correlation planes.

In this section, we briefly show how correlation filters can be used for face recognition. In addition, correlation filter methods also offer benefits to work in the spatial frequency domain, (i.e., the 2D Fourier transforms of the images) i) Graceful degradation in the matching operation refers to if some of the pixels are occluded, they simply do not contribute to the correlation peak. ii) Shift-Invariance refers to the resulting automatic shift invariance. Since correlation filters are linear shift invariant filters, any translation in the test input image will result in correlation output being shifted by exactly the same amount. Instead, we implicitly center the test image by locating the correlation peak. In practice, even with the best algorithm, centering test images is only approximate and can be time consuming. Thus, methods that avoid the need for explicit test image centering are attractive. iii) Closed-form solutions refers to correlation filters can be designed using closed-form expressions.
Chapter 5

Hybrid Two-directional Two-dimensional-PCA Correlation Filter for Face Recognition

5.1. Introduction

It is well known that when identifying faces in an automatic way, there exist great difficulties when the images to be processed are coming from a surveillance video system in which there are uncontrolled factors such as severe variations of lighting, expression, pose, occlusion and resolution [2]. Furthermore, an automatic face detection process does not provide registration accuracy neither does guarantee that the location of the face is perfect, and as a result, there are misalignment errors [2,86,112]. The recognition performance is affected by undesirable scene background and faces partially occluded as well [84].

This chapter presents a novel hybrid two-directional two-dimensional Principal Component Analysis ((2D)$^2$PCA) correlation filter for face recognition that allows to obtain a far better representation of the phase spectrum than its hybrid PCA-correlation filter counterpart. This hybrid (2D)$^2$PCA-correlation filter is capable of simultaneously dealing with several uncontrolled factors in face recognition. Such factors are addressed by combining two major approaches: (2D)$^2$PCA in the Fourier domain and the advance correlation filters (ACFs). The former method helps to extract and to represent more efficiently the facial features using the original image matrices, while the last method is used to simultaneously handle illumination variations, expression, partial occlusions and spatial shifts. The comparative results using
the Yale-B, ORL and FERET face databases show that this new proposed method produces a higher margin of separability between genuine and impostor classes than the advanced correlation filter designs based just on PCA, despite of using a subspace of smaller dimensionality. Additionally, it achieves a higher recognition accuracy that other methods in the frequency domain (Eigenphases and FM-$(2D)^2$PCA) and in the space domain (PCA and $(2D)^2$PCA).

**Advantages of Hybrid $(2D)^2$PCA-Correlation Filter**

This new hybrid $(2D)^2$PCA-correlation filter in addition to further enhancing the quality of the reconstructed phase spectrum, it models the phase spectrum variations of face images in the frequency domain. The main advantages of this novel method are:

**i)** Performing a two-directional two-dimensional Principal Component Analysis $(2D)^2$PCA [45] in the frequency domain on the phase-only spectrum of the face images; this procedure not only can represent more efficiently the face images but it is also both, computationally more efficient than the PCA alone and it increases the recognition rate.

**ii)** Extending further the hybrid-$(2D)^2$PCA subspace to achieve shift-invariance. This is done by developing a hybrid $(2D)^2$PCA-correlation filter that performs phase matching with a built-in shift-invariance function: if the test input image is shifted, probably due to registration errors, then the correlation output is also shifted by the same amount of pixels. This function avoids changing the peak sharpness on the correlation plane, and therefore the classification decision is not affected by this shifting.

### 5.2. CSS-Based Hybrid $(2D)^2$PCA-Correlation Filter Bank

An overview of the class-specific subspace-based hybrid $(2D)^2PCA$ correlation filter bank for face recognition is shown in Fig. 5.1. The hybrid-$(2D)^2PCA$ subspaces, based on an individual subspace, are first used to yield a filter bank to make them shift invariant, and thus tolerant to face image registration errors, resulting in
a css-based hybrid \((2D)^2\)PCA-correlation filter. During the training stage, a phase spectrum subspace, for every person, is built from the phase-only training images (see Section 5.3), and used to extract useful discriminatory information for each subject. Each hybrid-(2D)^2 PCA subspace is represented by two projection matrices (see Section 5.4) of the phase-only spectra, and then used to design a filter bank for all classes, that is, by projecting the test image phase spectrum onto hybrid-(2D)^2 PCA subspace, and reconstructing it for each phase spectrum subspace. Thus, we automatically turn each of the reconstructed phase-only test image into a correlation filter (see Section 5.5). Additionally, to each correlation filter designed, we add the average training phase derived from the phase-only training images corresponding to each subject.

During the test stage, the actual test, once designed the correlation filter bank, is then cross-correlated with every stored correlation filter derived in the target set, and the resulting product is input to a 2D inverse FFT to produce a correlation output. The PSR value is determined for each correlation output, and selected the one with the largest PSR value to label the test image. As illustrated in Fig. 5.1, a sharp correlation peak is produced when the input image is from an authentic, and no such discernible peak exist if the input is from an impostor. The location of the correlation peak will depend on the input image, if the input image is translated with respect to training images, then the output peak will be also shifted by the same amount without affecting the peak sharpness, neither the recognition performance.

![Figure 5.1: Schematic representation of the specific class subspace-based hybrid \((2D)^2\) PCA-correlation filter bank. The \(\otimes\), \(\odot\) and \(\oplus\) operators represent the correlation, projection and addition respectively.](image-url)
5.3. Phase-Only Fourier Synthesis

It has been shown (Oppenheim et al [113]) that the phase information of an image in the Fourier domain is more important than its magnitude. This statement can be illustrated as shown in Figure 5.2, where the Fourier transform for two subjects is computed, and the respective phase and magnitude spectrum are extracted, as shown in Figures 5.2 b) and c), respectively. Afterwards, a new face image can be synthesized from the first subject’s phase spectrum using the second subject’s magnitude spectrum, and vice versa, as shown in Figure 5.2 f), where the new synthesized face images clearly resemble the face image from which the phase spectrum was originally extracted from.

Mathematically, let \( f(x, y) \) be the brightness of an image at a spatial point with coordinates \((x, y)\), and if \( F(u, v) = |F(u, v)|e^{j\theta(u,v)} \) represents its Fourier transform at the respective spatial frequencies \((u, v)\), with spectral magnitude \(|F(u, v)|\) and spectral phase \(\theta(u,v)\), then it is possible to define the magnitude-only Fourier synthesis as

\[
f_m(x, y) = F^{-1}\{|F(u, v)|\} \tag{5.3.1}
\]

as well as to define its corresponding phase-only Fourier synthesis

\[
f_\theta(x, y) = F^{-1}\{M(u, v)e^{j\theta(u,v)}\} \tag{5.3.2}
\]

where \(M(u, v)\) is either unit, average or a spectral magnitude representing the class of signals.

When calculating the magnitude-only Fourier synthesis (see Figure 5.2 d)), \(f_m(x, y)\) is not resembled to \(f(x, y)\), meanwhile the phase-only synthesis does retain all the necessary information of the original image \(f(x, y)\), such as the edge information, as shown in Figure 5.2 e). Thus, a 2D signal constructed from only the phase is intelligible and retains many of the important features of the original. Moreover, by using only the phase information, the original 2D signal can be reconstructed up to a scale factor [114].

In practice, when using the phase-only spectra with the purpose of phase matching, it is necessary to previously perform a pre-whitening step, which correspondingly produces a unity magnitude for all the spatial frequencies of the training and test images.
5.3 Phase-Only Fourier Synthesis

Figure 5.2: Importance of phase information in images. a) Original images of first and second subjects, b) Phase spectrum of subject 1 and 2, c) Magnitude spectrum of subject 1 and 2, d) Magnitude-only Fourier synthesis of subject 1 and 2, e) Phase-only Fourier synthesis of subject 1 and 2, and f) Inverse Fourier transform from combination of the phase spectrum of the first subject with the magnitude spectrum of second subject, and vice versa.

The previous idea was exploited in [115], where it was proven that the eigenvectors obtained by performing PCA in the frequency domain alone were simply the same principal components resulting in its spatial domain counterpart and only differ by a sign change. Moreover, by modelling the subspace of the phase-only spectra of the training images, it yields what is known as eigenphases, and it simultaneously gives robustness against occlusion and illumination variations. This method assumes that most of illumination variations are in the lower frequency spectrum.

Even though this subspace of eigenphases seems good to represent face images in the minimum mean squared error sense [96–98, 115], PCA may not capture the variability in a set of training images discriminating one person’s face from another, unless this information is explicitly given in this dataset [107]. To overcome such drawbacks, the recognition performance can be improved by building a hybrid-(2D)$^2$PCA subspace, which not only increases the margin of separability between classes, but also keeps the advantage of robustness and efficiency in representing face images of the phase-only synthesis in conditions of extreme illumination and partially occluded faces.

The next two sections show how such advantages can be achieved.
5.4. Using (2D)$^2$ PCA in the Frequency Domain

Given an $m \times n$ image matrix $^1A^w$ in the Fourier domain of the phase-only spectra, and a linear transformation that maps the original phase-only matrix $A^w$ into an $m$-dimensional phase-only projected vector, then the new phase-only spectra matrix $Y \in \mathbb{C}^{m \times d}$ is:

$$Y = A^w V$$ \hspace{1cm} (5.4.1)

where $V \in \mathbb{C}^{n \times d}$ is the phase-only projection matrix with orthonormal columns and $n \geq d$. This other description space is called phase-only feature space.

Now, consider a set of N image samples where the $k$th training sample is an $m \times n$ space domain matrix $A^s_k$, chosen from $k = 1 \ldots N$, and consider the total scatter matrix $C^r_f$ of the projected samples defined as:

$$C^r_f = \frac{1}{N} \sum_{k=1}^{N} \{ T_{DFT_w}(A^s_k - \mu) \}^+ \{ T_{DFT_w}(A^s_k - \mu) \}$$ \hspace{1cm} (5.4.2)

$$= T_{DFT_w} X^+ X T_{DFT_w}^{-1}$$

where $T_{DFT_w}$ is the discrete Fourier transform matrix of the phase-only spectra, $\mu \in \mathbb{C}^{m \times n}$ is the mean image of all the samples and the superscript $^+$ represents the conjugate transpose, then the scatter matrix is in fact characterized by the trace of the covariance matrix of the projected feature vectors in the frequency domain. Since Eq.(5.4.2) is an $n \times n$ Hermitian nonnegative definite matrix, their eigenvalues will be real and positive.

Note that $X^+ X$ is the respective spatial covariance matrix $C_s$ defined as:

$$C^r_s = X^+ X = \sum_{k=1}^{N} (A^s_k - \mu)^+ (A^s_k - \mu)$$ \hspace{1cm} (5.4.3)

Thus, after applying the previously defined linear transformation, the scatter of the projected feature vectors $\{Y_1, Y_2, \ldots, Y_d\}$ is $V_f^+ C^r_f V_f$. Then, the projection optimal $V_{opt_f}$ is defined as:

$$V_{opt_f} = \{V_{1_f}, V_{2_f}, \ldots, V_{d_f}\} = \text{arg max } \text{tr} (V_f^+ C^r_f V_f)$$ \hspace{1cm} (5.4.4)

$^1$A pre-whitening step is performed to yield a phase-only spectrum with a unity magnitude for all spatial frequencies $(u, v)$. 

1. Department of Electronics
2. National Institute of Astrophysics, Optics and Electronics
where \( \{V_1, V_2, \ldots, V_d\} \) are the orthonormal eigenvectors of \( C_f^r \) corresponding to the first \( d \) largest eigenvalues. However, one optimal projection axis is not enough to accurately represent the data [107], and therefore a set of projection axes are needed.

Note that the size of image covariance matrix \( C_f^r \) is \( n \times n \) which allows us to efficiently compute its eigenvectors [45, 107]. Likewise as PCA, the number of eigenvectors of \( d \) is chosen by setting a threshold:

\[
\sum_{i=1}^{d} \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq \theta.
\] (5.4.5)

where \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are the \( n \) largest eigenvalues of \( C_f^r \) and \( \theta \) is a pre-set threshold value.

Eq.(5.4.2) can be considered as the original 2DPCA which is working in the images row direction [107]. An alternative way of 2DPCA can be obtained by applying 2DPCA on the image columns [45] as follows:

\[
C_c^r = \frac{1}{N} \sum_{k=1}^{N} \{T_{DFT_w}(A^s_k - \mu)\} \{T_{DFT_w}(A^s_k - \mu)\}^+ = T_{DFT_w}XX^+T_{DFT_w}^{-1}
\] (5.4.6)

where \( XX^+ \) is the space domain covariance matrix \( C_c^s \) defined as:

\[
C_c^s = XX^+ = \sum_{k=1}^{N} (A^s_k - \mu)(A^s_k - \mu)^+
\] (5.4.7)

Similarly, the projection \( U_{opt,f} \) is chosen to maximize the trace of the scatter matrix:

\[
U_{opt,f} = \{U_1, U_2, \ldots, U_q\} = \arg \max \text{ tr } (U_f^+C_f^cU_f)
\] (5.4.8)

where \( \{U_1, U_2, \ldots, U_q\} \) are the orthonormal eigenvectors of \( C_f^c \) corresponding to the first \( q \) largest eigenvalues. The number of eigenvectors of \( q \) can also be selected by setting a \( \theta \) threshold value using Eq. (5.4.5).

Likewise, as in Eq. (5.4.1), the new phase-only feature matrix \( B \in \mathbb{C}^{q \times n} \) is defined by the following linear transformation:

\[
B = U^+A^w
\] (5.4.9)
where $U \in \mathbb{C}^{m \times q}$ is the phase-only projection matrix with orthonormal columns, with $m \geq q$. In the above Eqs. (5.4.1) and (5.4.9), the phase-only projection matrices are computed off-line.

### 5.5. Hybrid (2D)$^2$PCA-Correlation Filter

Once both row-based and column-based optimal linear subspaces have been computed for the phase-only spectra of the training images, then the reconstructed phase spectrum $e^{\phi_R(u,v)}$ is computed as:

$$ e^{\phi_R(u,v)} = U_{f_j}(u,v)e^{\phi_T(u,v)}V_{f_i}^+(u,v) $$

(5.5.1)

where $e^{\phi_T(u,v)}$ is the test image phase spectrum projected onto the frequency domain eigenvectors $U_{f_j}$ and $V_{f_i}$. Note that the subscripts $i$ and $j$ denote the $i$th and $j$th projection vectors.

![Figure 5.3: Phase-only synthesis from the reconstructed phase spectrum of a training face image by using the inverse Fourier transform under similar compression ratios. a) Original and synthesized face, b) 2D$^2$PCA based eigenvectors of the phase-only spectra (phase angles between the imaginary and real part of each frequency), c) Inverse Fourier transform of the phase-only synthesis based on 2D$^2$PCA and d) Inverse Fourier transform of the phase-only synthesis based on PCA.](image-url)
This $(2D)^2$PCA based subspace of the phase-only spectra not only improves the reconstruction accuracy but also increases the recognition performance, as it is shown in the experiments described in the next section. Figure 5.3 shows a training face image synthesized from the Fourier transform phase with unity magnitude, where the phase spectrum was reconstructed under similar compression ratios [45], by choosing the appropriated number of eigenvectors in PCA according to Eq.5.4.5. Given this proper dimensionality, the compression ratios for both PCA and $(2D)^2$PCA based correlation filters was found out. It can be shown that the $(2D)^2$PCA based phase-only Fourier synthesis yields face images with higher quality while preserving the edge information.

On the contrary, when synthesizing a test face image at different dimensions by using PCA-based phase-only spectra, it does not preserve many edge information or structure of a face image as the synthesized original face image, as shown in Figure 5.4 d). However, the PCA approach leads to the idea of basing face recognition on a small set of image features that best approximate the set of known face images, without requiring that they correspond to our intuitive notions of facial parts and features [4]. By using $(2D)^2$PCA-based phase-only spectra, test face images is not only preserved the edge information but also they corresponds to our intuitive notions of facial parts and features, as shown in Figure 5.4 e). That way, we may claim this reconstruction accuracy is reflected in the recognition performance.

In spite of yielding illumination-tolerant images synthesized from Fourier transform discriminating one person of each other, reconstruction errors in the frequency domain can be produced because if the input image is shifted then a linear term is added to the phase [86]. Thus, by using a similarity measure such as the Frobenius norm between the reconstructed image and the test image, the recognition performance will degrade [94,95].

However, in order to achieve shift-invariance along with tolerance to illumination and occlusion, as it was done in [86], a hybrid $(2D)^2$PCA-correlation filter was developed from the reconstructed phase spectrum. To examine how well this phase spectrum is reconstructed, i.e., if the reconstructed phase spectrum is identical to the test phase spectrum, the phase-only correlation between the conjugate of the reconstructed phase spectrum and the phase spectrum of the test image at all possible shifts
Figure 5.4: Phase-only synthesis from the reconstructed phase spectrum of a test face image by using the inverse Fourier transform under similar compression ratios. a) Original and synthesized face, b) 2D^2PCA based eigenvectors of the phase-only spectra (phase angles between the imaginary and real part of each frequency), c) Inverse Fourier transform of the phase-only synthesis based on 2D^2PCA and d) Inverse Fourier transform of the phase-only synthesis based on PCA.

Figure 5.5: Correlation plane corresponding to the genuine and impostor class. a) Sharp correlation peak belonging to a face from the same person and b) Correlation output belonging to a face from a different person.

is computed as follows:

\[ C(u, v) = e^{-\phi_R(u,v)}e^{\phi_T(u,v)} = e^{\phi_T(u,v)-\phi_R(u,v)} \] (5.5.2)
\[ C_T(x, y) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} e^{i\phi_T(u,v) - \phi_R(u,v)} e^{i \frac{2\pi}{m} ux} e^{i \frac{2\pi}{n} vy} \]  
(5.5.3)

thus, as expected, a constant flat spectrum is obtained cancelling out all complex phases \( \phi(u, v) \) by adding their opposites. Such a constant frequency spectrum represents a sharp peak (as delta-function type output in the ideal case) in the correlation output.

Figure 5.6: Correlation output plane. a) Face image centered in the scene, b) The same face image shifted up 40 pixels. The PSR remains exactly the same as the unshifted face and the peak is shifted up by the same amount.

Thus, by using the inverse Fourier transform, a sharp peak appeared in the correlation output plane \( C_T \), whose peak location will be shifted depending on the shifting of the input image. Figure 5.5 a) shows a sharp correlation peak resulting from the (2D)\(^2\)PCA filter cross-correlated with a face from the same person whose filter was made for. Meanwhile, Figure 5.5 b) shows a shapeless correlation output resulting from the filter cross-correlated with a face image from a different person to whom the filter was made. The shift-invariance property is illustrated in Figure 5.6. Part a) shows the output correlation of a face image centered in scene of Figure 5.7 a), giving a very large PRS of 101.4988, whereas in Figure 5.6 b) depicts the resulting output correlation of a shifted and occluded version (see Figure 5.7 b)) of the reference image.

\(^2\)When frequency spectra are multiplied, their phases are added.
This face image was shifted up to 40 pixels where the PSR remains exactly the same but the peak is shifted by exactly the same amount showing the location of the face in the scene. Thus there is no need to go through the trouble of centering the input image prior to recognizing it.

An appropriate measure to compute the peak sharpness and its location along with the shift in the input image is the Peak-to-Sidelobe Ratio (PSR) metric, which allows us to decide whether a test image belongs to the authentic class. Contrarily, when dealing with impostor face images, the correlation plane should not exhibit such sharp correlation peaks. The Peak-to-Sidelobe Ratio (PSR) is defined as follows:

\[
PSR = \frac{\text{peak} - \mu_{\text{area}}}{\sigma_{\text{area}}} \tag{5.5.4}
\]

where \(\mu_{\text{area}}\) and \(\sigma_{\text{area}}\) are the mean and standard deviation, respectively, of an small area around, but excluding the peak. The PSR is computed following [86]. By calculating the correlation output plane with its corresponding peak using Eq. (5.5.3), a smaller region centered at the peak is selected (in the example of Figure 5.5 a 20 \(\times\) 20 wide area), then even a smaller region around the peak (in the example a 5 \(\times\) 5 region) is excluded and finally, the remaining region defined as the sidelobe region is used to compute the mean and standard deviation of the sidelobes. In fact, the larger the PSR value the more likely the test images belongs to the original class. Otherwise, the correlation output exhibits a very low PSR.

Note that when a shifted test image is projected into the phase-only subspace,
5.6. Experimental Results

In this section the performance for both the hybrid $(2D)^2$PCA-correlation filter bank and the hybrid PCA-correlation filter bank is evaluated experimentally, and for both methods, we have used the Logarithm transformation [116] of the intensity values for face image enhancement. Additionally, other state-of-the-art algorithms such as IPCA [117], PCA [38], FM-$(2D)^2$PCA [94] and $(2D)^2$PCA [45] are used as baseline. With the purpose of exploring the face identification performance of this technique, the experiments were carried out on two well-known face databases: Yale-B and AR. The Yale-B [118] face database is used to examine the performance under 64 different lighting conditions and various facial expressions. The AR face database [103] is employed to test performance when there are different lighting conditions, partial occlusions, and facial expressions.

In the experiments, the face images used for the Yale-B database are of a fixed size $64 \times 64$ pixels while for the AR database are of a fixed size $32 \times 32$. The number of projection coefficients used in PCA or its case IPCA (see Eq. 5.4.5) is controlled by the value $\theta$, set to 0.95 for all the datasets. Given this dimension, the number of projection vectors is computed for both PCA and $(2D)^2$PCA based correlation filters for representing face images under similar compression ratios [45]. The computational time for each of these methods, which includes both training and recognition phases, is measured for all face databases. Additionally, all the experiments are based on the rank-1 recognition rates and tested on a Core i5 2.5Ghz CPU, 8 Gb memory computer.
Figure 5.8: Face images exhibiting little or extreme illumination from Yale-B face databases.

Table 5.1: Face recognition performance comparison of PCA, (2D)²PCA, FM-(2D)²PCA, IPCA, the hybrid PCA-correlation filter and the hybrid (2D)²PCA-correlation filter for the Yale-B database, where training images for both Set 1 and Set 2 were captured under illumination directions in both negative and positive azimuth, respectively.

<table>
<thead>
<tr>
<th>Training images</th>
<th>PCA</th>
<th>(2D)²PCA</th>
<th>FM-(2D)²PCA</th>
<th>IPCA</th>
<th>H-PCA</th>
<th>H-(2D)²PCA</th>
<th>Method [97]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 1</td>
<td>47.3%</td>
<td>50.3%</td>
<td>61.0%</td>
<td>67.6%</td>
<td>100%</td>
<td>100%</td>
<td>95.0%</td>
</tr>
<tr>
<td>Set 2</td>
<td>51.1%</td>
<td>52.5%</td>
<td>59.1%</td>
<td>78.0%</td>
<td>98.6%</td>
<td>99.5%</td>
<td>97.0%</td>
</tr>
<tr>
<td>Average time(s)</td>
<td>5.28</td>
<td>2.82</td>
<td>3.17</td>
<td>3.54</td>
<td>8.02</td>
<td>4.75</td>
<td>-</td>
</tr>
</tbody>
</table>

5.6.1. Experiments on the Yale-B database

The performance of the proposed approach was evaluated using 64 frontal face images under different lighting conditions, from negative azimuth to positive azimuth, of 10 people. Examples of cropped images of one subject are shown in Fig. 5.8. In this case, two different experiments were carried out by forming two set of training images. The first set contains 350 images of negative azimuth (with 35 images per person), and the second set contains 290 images of positive azimuth (with 29 images per person). Table 5.1 shows the performance as well as average running times of the proposed approach along with other standard methods using Set 1 and Set 2 as training and testing sets, respectively. A significant improvement on the recognition accuracy and robustness of the hybrid (2D)²PCA-correlation filter against the other methods can be easily noticed. In both cases correlation filter methods overcome other methods, however the hybrid (2D)²PCA-correlation filter slightly improved the performance, but the running time used (4.75s) is still lower that its PCA-based correlation filter counterpart.

5.6.2. Experiments on the AR database

In this section, we evaluate the performance of the css-based hybrid (2D)²PCA-correlation filter method using a subset of the AR face database [103]. This subset
5.6 Experimental Results

Figure 5.9: Cropped face images exhibiting different illumination conditions, facial expressions and partial occlusions included in a) Set 1 and b) Set 2.

Figure 5.10: Example of face images used for training of a subject on the AR database.

consists of 1150 images of 50 different subjects (25 males and 25 females), and each subject has 23 different face images, with different illumination conditions; facial expressions and partial occlusions. Three different experiments were performed by dividing the dataset into two different sets, as shown in Fig. 5.9.

In the first experiment, we evaluated the performance of all 6 methods against illumination variations and facial expressions applying the Set 1, where they were trained using seven different images without occlusions, as shown in Fig. 5.10.

In the second experiment, the performance was evaluated using the Set 2 along with another set chosen arbitrarily. We use a total of 20 different images including lighting variations; facial expressions and partial occlusions (sunglasses and scarf). We trained the face subspace with only six face images per person (i.e. four persons with neutral expression, one with illumination variation, one with sunglasses, and one using scarf).

In the last experiment, we trained just as the second experiment, however in the testing stage we used another subset of eight images per person, which includes images with partial occlusions. Table 5.9 shows that the proposed method presents an improvement in recognition accuracy over the PCA-based correlation filter and the other methods, and shows that is still more tolerant to lighting variations, facial expressions, and partial occlusions. Also, the average time of recognition is significantly
Table 5.2: Face recognition performance comparison of PCA, (2D)\(^2\)PCA, FM-(2D)\(^2\)PCA, IPCA, the hybrid PCA-correlation filter and the hybrid (2D)\(^2\)PCA-correlation filter for the AR database.

<table>
<thead>
<tr>
<th></th>
<th>PCA</th>
<th>(2D)(^2)PCA</th>
<th>FM-(2D)(^2)PCA</th>
<th>IPCA</th>
<th>H-PCA</th>
<th>H-(2D)(^2)PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training images</td>
<td>% Rec (%)</td>
<td>% Rec (%)</td>
<td>% Rec (%)</td>
<td>% Rec (%)</td>
<td>% Rec (%)</td>
<td>% Rec (%)</td>
</tr>
<tr>
<td>Set 1</td>
<td>87.7%</td>
<td>88.0%</td>
<td>87.1%</td>
<td>95.7%</td>
<td>93.1%</td>
<td>95.0%</td>
</tr>
<tr>
<td>Set 2</td>
<td>38.3%</td>
<td>39.2%</td>
<td>41.0%</td>
<td>54.6%</td>
<td>88.3%</td>
<td>91.0%</td>
</tr>
<tr>
<td>Set 3</td>
<td>27.0%</td>
<td>30.0%</td>
<td>33.0%</td>
<td>44.0%</td>
<td>77.0%</td>
<td>80.5%</td>
</tr>
<tr>
<td>Average time(s)</td>
<td>6.9</td>
<td>5.6</td>
<td>5.2</td>
<td>7.2</td>
<td>12.6</td>
<td>10.7</td>
</tr>
</tbody>
</table>

lower that its PCA-based correlation filter counterpart. By training the subspace algorithms with the Set 1, IPCA performs better than most of presented methods, however, its performance was severely degraded when test images with partial occlusions and facial expressions were present.

5.6.3. Discussion

Observations from the performed evaluations support that the hybrid (2D)\(^2\)PCA-correlation filter improves recognition accuracies in less time, and can deal with faces with strong illumination variations, partial concussions, and facial expressions better than its hybrid PCA-based correlation filter counterpart.

Based on recognition rates, if dimensionality is reduced, the recognition accuracy of the hybrid PCA-correlation filter can be degraded, while the hybrid (2D)\(^2\)PCA-correlation filter performance may be increased. However, by increasing the dimensionality beyond from the similar compression ratios, the resulting hybrid-(2D)\(^2\)PCA subspace could not represent faces, decreasing the performance. Furthermore, by increasing both the number of eigenvectors and the training images, it is evident that recognition rates may be improved, but it may incur in an increase of processing time because the training, projection, reconstruction and face matching stages are time consuming.
6.1. Face Recognition

This thesis introduced a new hybrid $\text{(2D)}^2\text{PCA}$-correlation filter which allows to get a far better representation than its PCA counterpart of the reconstructed phase spectrum for face images. Our method is based on the best of two approaches for representing face images: the two-directional two-dimensional PCA ($\text{(2D)}^2\text{PCA}$) in the Fourier transform domain for computing the eigenvectors of the phase-only spectra and advanced correlation filters for simultaneously handling variability such as illumination variations, spatial shifts, poses (within $\pm20^\circ$ in yaw and tilt), facial expressions and occlusions. This hybrid-$\text{(2D)}^2\text{PCA}$ subspace produces a higher margin of separability between genuine and impostor classes because the preserved edge information allows to efficiently represent face images of the phase-only synthesis under uncontrolled conditions.

Afterwards, a shift-invariant hybrid $\text{(2D)}^2\text{PCA}$-correlation filter bank is designed to accomplish tolerance to image registration errors by reconstructing the phase-only image at each applied shift once the face image is projected into the hybrid-$\text{(2D)}^2\text{PCA}$ subspace. The peak location into the correlation output plane, which contains the phase-only cross-correlation of the reconstructed phase-only image and the test image at all possible shifts in the scene image, will depend on the input image spatial shift. Thus, this method models the phase variations for a particular person enabling to later perform a classification using ACFs. In fact, our hybrid $\text{(2D)}^2\text{PCA}$-correlation filter bank inherits and further improves the attractive qualities from the hybrid PCA-correlation filter bank. Moreover, the $\text{(2D)}^2\text{PCA}$ based phase-only Fourier synthesis not only preserves better the edge information and yields face images with higher...
quality, but also enhances the quality of the reconstructed phase spectrum for test images by using the same dimensions of feature vectors.

Observations from performed evaluations support that the hybrid \((2D)^{2}\)PCA-correlation filter improves recognition accuracies in less time, and can deal with faces with strong illumination variations, partial concussions, and facial expressions better than its hybrid PCA-based correlation filter counterpart.

Based on recognition rates, if dimensionality is reduced, the recognition accuracy of the hybrid PCA-correlation filter can be degraded, while the hybrid \((2D)^{2}\)PCA-correlation filter performance may be increased. However, by increasing the dimensionality beyond from the similar compression ratios, the resulting hybrid-(2D)\(^2\)PCA subspace could not represent faces, decreasing the performance. Furthermore, by increasing both the number of eigenvectors and the training images, it is evident that recognition rates may be improved, but it may incur in an increase of processing time because the training, projection, reconstruction and face matching stages are time consuming.

Experimental results show that the proposed method clearly improves significantly the face recognition accuracy that other related algorithms in the frequency domain (Hybrid PCA-Correlation Filter, FM-(2D)\(^2\)PCA), and overcomes algorithms in the space domain (PCA, IPCA, 2D\(^2\)PCA). This results were supported by the similar compression ratios on face images from AR, Yale-B face databases.

Future research work will be to incorporate face super-resolution into our hybrid \((2D)^{2}\)PCA-correlation filter for its application in video surveillance system, and handle pose variations. Compared to still images face recognition, there exist several disadvantages of video sequences. First, images captured by CCTV cameras are generally of poor quality and in addition to noise level is higher, images may be blurred due to movement or the subject being out focus. Second, image resolution is normally lower for video sequences and if the subject is very far from the camera, the face image resolution can be as low as \(64 \times 64\). Another measure to guarantee the robustness and efficiency of the recognition accuracy is that of recognizing faces in arbitrary in-depth rotations as well as experimenting of face recognition across pose in several face databases which contains larger pose variations and vertical in-depth such as CMU-PIE, FERET, XM2VTS and LFW databases.
6.2. Face De-Identification

We also introduced a new DCT-domain foveation based method for face de-identification, which not only protects the individual’s identity but also preserves information which it is used by a classifier for both gender and facial classification tasks. This model, by applying DCT-domain foveation, can simplify the cut-off frequency and contrast sensitivity computations by introducing disjoint DCT-domain foveation regions. The proposed face de-identification technique based on DCT-domain foveation is competitive to the other methods tested here in balancing privacy protection and utility data preservation. Moreover, we also performed a exhaustive research which analyses different face de-identification algorithms, and seeks to thwart face recognition while preserving awareness sufficient on both gender and facial expression classification tasks.

The privacy-awareness trade-off is measured by plotting both the classification accuracy and the recognition rate quantifiers, which yield a privacy-awareness map. The recognition accuracy of the de-identified faces was used to measure the privacy protection by building a PCA standard and a (2D)^2PCA method in the top-rank, meanwhile the classification rate for gender and expression of the de-identified faces was used to quantify the awareness data by applying a support vector machine classifier.

Experimental results show that the proposed foveation-based face de-identification method clearly simultaneously preserves privacy and information in de-identified face images.

Future research work associated to foveation for preserving privacy and information in de-identified face images is that of incorporating a formal privacy model to prevent attacks applied to gallery images [18], who attempts to reidentify (by face recognition software) a de-identified face set. There are tree kind of attacks to recognize face images from the gallery set by using the probe face set. 1) Original images are matched to altered faces, where the de-identified images are just run through the face recognition software. 2) Altered images as gallery are matched to original images as probe. It assumes that attacker already has an original face images set that were de-identified. 3) Altered images are matched to altered images. The attacker can invoke the same de-identification technique, so the training and probe face sets are de-identified, and thus a de-identified gallery set is matched to a de-identified probe

Face De-Identification Algorithms
Also, we consider that such an approach could be highly effective if we use multiple fixation points in an image, for example to place fixation points into nose, mouth and eye for preserving more facial details in a de-identified face. Also, it is necessary to add or integrate more attributes to the data utility functions, such race, age, hair, eyeglass, etc.


[81] B. Kumar and A. Mahalanobis, “Recent advances in composite correlation filter designs.”


Appendix

MATLAB Code

A.1. Hybrid (2D)$^2$ PCA Correlation Filter

A.1.1. YALE-B face database

Two sets of training data are formed, is to say, the set 1 contains 350 images of negative azimuth (with 35 images for each person), and the set 2 contains 290 images of positive azimuth (with 29 images for each person).

```matlab
function Eij = itwopca2DDFT_ver2_yale (d1, nd2)
% The main function for Hybrid (2D)$^2$ PCA Correlation Filter
% YALE-B FACE DATABASE
% clear all;
clear;
oclock; % position
path = '../CroppedYaleSize64';
class_num = 10;
% In this experiment, we select the images of 35 negative azimuths for training
% and use the remaining images of 29 positive azimuths for testing.
% The performance is evaluated by interchanging the training and testing data sets
% Note that a positive azimuth implies that the light source was to the right
% of the subject while negative means it was to the left.
% Negative azimuths (it includes images with sign A+, left shadow,
% light source :right)
train_num = 35;
% Positive azimuths (it includes images with sign A-, right shadow)
test_num = 29;
Test_Image ={};
FSS_U1 = {};
FSS_V1 = {};
for i=1:class_num
pos = 1; % position
pos2 = 1;
```
A. MATLAB Code

```matlab
name = ['path'/yaleB'num2str(i,'%02d')'/yaleB'num2str(i,'%02d')'_P00.info'];
 fid = fopen(name, 'r');
 s = fgetl(fid);
 while ischar (s) % 1:class_sample = 1:64;
    name2 = ['path '/yaleB' num2str(i,'%02d') '/' s '.pgm'];
    azimuths = s(13);
    X = double(imread(name2));
    X = log(X+1);
    imgdft = fft2(X)/sqrt(64*64);
    phase = atan2(imag(imgdft), real(imgdft));
    phase_spec = exp(1i*phase);
    if azimuths == '+' %Negative azimuth
        Train_Image (:,:,pos) = phase_spec;
        pos = pos+1;
    else
        Test_Image {i,pos2} = phase_spec;
        pos2 = pos2+1;
    end; %if
    s = fgetl(fid);
end; %While

% Built a specific face subspace (FSS)
[v_sort, vj_sort, mean_Image] = iFSS(Train_Image, 1, train_num, d, nd);
FSS_U1{i} = vj_sort; %Left projection matrix
FSS_V1{i} = v_sort; %Right projection matrix
meanImage(:,:, i) = mean_Image;
Train_Image = [];
fclose(fid);
end; %for

[size_y, size_x] = size(Test_Image{1});
ac9 = 0;
for i = 1 : class_num
    for j = 1 : test_num
        for k = 1 : class_num
            % Project it
            CoeTest0 = FSS_U1{k}'*Test_Image{i,j}*FSS_V1{k};
            % Reconstruct it
            Filter2D0 = FSS_U1{k}.*CoeTest0.*FSS_V1{k}';
            corr = ifft2(Test_Image{i,j}.*conj(Filter2D0+meanImage(:,:,k)));
            PSR(k) = computePSRs(corr, size_y, size_x);
            % Compute PSR by using a different sidelobe region
            PSR(k) = computePSRs2(corr, size_y, size_x);
        end;
        [max_d9, max_index9] = max (PSR);
        if i == max_index9
            ac9 = ac9 + 1;
        end
    end
end
```

Department of Electronics
National Institute of Astrophysics, Optics and Electronics
A.1 Hybrid (2D)$^2$ PCA Correlation Filter

end;
Rank = (ac9 / (class_num*test_num)) * 100;
time1 = etime(clock,t0)
Errors = (class_num*test_num) - ac9;

function [Ev_sort, Evj_sort, mean_Image] = iFSS(Train_Image, class_num, train_num)
% Computing of eigenphases for Hybrid (2D)$^2$ PCA Correlation Filter

ImageSize = size(Train_Image(:,:,1));
Covi = zeros(ImageSize(2),ImageSize(2)); %Row direction
Covj = zeros(ImageSize(1),ImageSize(1)); %Column direction
mean_Image = mean(Train_Image,3);
Train_ImageSub = zeros(ImageSize(1),ImageSize(2),class_num*train_num);
for i=1:class_num*train_num
    Train_ImageSub(:,:,i) = Train_Image(:,:,i) - mean_Image(:,:);
    Covi = Covi + Train_ImageSub(:,:,i)'*Train_ImageSub(:,:,i);
    Covj = Covj + Train_ImageSub(:,:,i)*Train_ImageSub(:,:,i)';
end;
Covi = Covi/(class_num*train_num);
Covj = Covj/(class_num*train_num);
[V,D] = eig(Covi);% find eigenvalues and eigenvectors
% Sort eigenvalues in descending order
[d2,index] = sort(d1, 'descend');
% Normalized eigenvalues
normalised_evalues = d2 ./ sum(d2);
threshold = 0; i = 1;
% Dimension is chosen according to the threshold, eg., 0.95
% eg., d = 28;
% eg., nd = 27;
while (threshold <= 0.85) && (i <= length(normalised_evalues))
    threshold = sum(normalised_evalues(1:i))/sum(normalised_evalues); i = i+1;
end
d = i - 1;
v_sort = zeros(ImageSize(2),d);
Ev_sort = zeros(ImageSize(2),d);

% Orthogonal eigenvectors corresponding to the first d largest eigenvalues
for i=1:d
    v_sort(:,i) = V(:,index(i));
end;
% Normalize eigenvectors
for i = 1 : d
    v_sort(:,i) = v_sort(:,i) / norm(v_sort(:,i));
end;
[V,D] = eig(Covj);
di = diag(D);
[d2,index] = sort(d1, 'descend');
% Normalized eigenvalues
normalised_evalues = d2 ./ sum(d2);
threshold = 0; i = 1;
while (threshold <= 0.85) && (i <= length(normalised_evalues))
    threshold = sum(normalised_evalues(1:i))/sum(normalised_evalues); i = i+1;
end
nd = i - 1;
vj_sort = zeros(ImageSize(1),nd);
Evj_sort = zeros(ImageSize(2),nd);
% Orthogonal eigenvctors corresponding to the first nd largest eigenvalues
for i=1:nd
    vj_sort(:,i) = V(:,index(i));
end;
% Normalize
for i = 1 : nd
    vj_sort(:,i) = vj_sort(:,i) / norm(vj_sort(:,i));
end;

function [PSR] = computePSRs(corr, size_y, size_x)
% Peak-Sidelobe Ratio
% To compute the PSR's values, the mean and standard deviation of the
% sidelobe is computed from the correlation energy plane (C).
% PSR = (max(C) - mean(C))/std(C)
% To do this, an $5 \times 5$ area of the correlation centered on
% the maximum value was masked and the rest of the values of the
% correlation plane was used for the sidelobe.
%
% Calculating PSR
    corr=real(corr);
    corrN = corr./max(max(corr)); %Normalize
    corrMax = max(max(corrN));
    cS = fftshift(corrN);
% Find max in each col, and then find max in each row
    [maxY, yLoc] = max(cS,[],1); % yLoc stores indices corresponding to column vector
    [maxX, xLoc] = max(cS,[],2); % xLoc stores indices corresponding to row vector
    yVal = 1;
    xVal = 1;
% Find the correlation peak on columns and rows
    for coord = 2:size_x
        if maxY(coord) == 1 % Columns
            yVal = yLoc(coord);
        end
    end
    for coord = 2:size_y
        if maxX(coord) == 1 % Rows
            xVal = xLoc(coord);
        end
    end
% A $5x5$ area around of the peak is extracted and the remaining region
% defined as the sidelobe regions is used to compute the mean and standard
% deviation of the sidelobes
% Note: you can limit the sidelobe region by adding another rectangular region
% eg., $20\times20$, refer to function \([\text{PSR}] = \text{computePSRs2}(\text{corr}, \text{size}_y, \text{size}_x)\)

region = 3;
yValP = yVal + region; % Col Right
yValN = yVal - region; % Col Left
xValP = xVal + region; % Row Bottom
xValN = xVal - region; % Row Top
if yValP > size_x
    yValP = size_x;
end
if yValN < 1
    yValN = 1;
end
if xValP > size_y
    xValP = size_y;
end
if xValN < 1
    xValN = 1;
end
if yValN > size_x
    yValN = size_x;
end
cStop = cS(1:xValN,1:size_x);
cSleft = cS(xValN:xValP,1:yValN);
cSright = cS(xValN:xValP,yValP:size_x);
cSbottom = cS(xValP:size_y,1:size_x);
% Concatenate and vectorize
cStot = [cStop(:); cSleft(:); cSright(:); cSbottom(:)];
cSMean = mean(cStot(:));
PSR = abs((corrMax - cSMean)/(std(cStot(:))));

\begin{function}
\begin{verbatim}
%  \text{computePSRs2}(cS, \text{size}_y, \text{size}_x)
%  \text{corr}=real(cS);
%  corrN = corr./max(max(corr)); % Normalize
%  corrMax = max(max(corrN));
%  cS = fftshift(corrN);
%  Find first Max in each column and then in each row
%  [maxY, yLoc] = max(cS,[],1); % Find maximum in column vector
%  [maxX, xLoc] = max(cS,[],2); % Find maximum in row vector
%  yVal = 1;
%  xVal = 1;
\end{verbatim}
\end{function}
% Find the correlation peak at columns and rows

for coord = 2:size_x
    if maxY(coord) == 1 % Columns
        yVal = yLoc(coord);
    end
end

for coord = 2:size_y
    if maxX(coord) == 1 % Rows
        xVal = xLoc(coord);
    end
end

region = 3;
yValP = yVal + region; %Col Der
yValN = yVal - region; %Col Izq
xValP = xVal + region; %Row Baja
xValN = xVal - region; %Row Sube

if yValP > size_x
    yValP = size_x;
end

if yValN < 1
    yValN = 1;
end

if xValP > size_y
    xValP = size_y;
end

if xValN < 1
    xValN = 1;
end

if yValN > size_x
    yValN = size_x;
end

slope_reg = 11;
yValPsl = yVal + slope_reg; %Column Right
yValNsl = yVal - slope_reg; %Column Left
xValPsl = xVal + slope_reg; %Row Bottom
xValNsl = xVal - slope_reg; %Row Top

if yValPsl > size_x
    yValPsl = size_x;
end

if yValNsl < 1
    yValNsl = 1;
end

if xValPsl > size_y
    xValPsl = size_y;
end

if xValNsl < 1
    xValNsl = 1;
end

if yValNsl > size_x
    yValNsl = size_x;
end

cStop = cS( xValNsl:xValN-1, yValNsl:yValPsl); %cStop = cS(1:xValN,1:size_x);
A.1 Hybrid (2D)^2 PCA Correlation Filter

\[ cS_{left} = cS(xValN:xValP, yValN:yValN-1); \]
\[ cS_{right} = cS(xValN:xValP, yValP+1:yValPs1); \]
\[ cS_{bottom} = cS(xValP+1:xValPs1, yValNs1:yValPs1); \]
% Concatenate and vectorize
\[ cS_{total} = [cS_{top}(:); cS_{left}(:); cS_{right}(:); cS_{bottom}(:)]; \]
\[ cS_{mean} = mean(cS_{total}(:)); \]
\[ PSR = abs((corrMax - cS_{mean})/(std(cS_{total}(:)))); \]
Appendix

Published Works

B.1. Conference Proceedings


B.2. International Journals
