Contribution to 2D/3D face recognition/authentification

Walid Hariri

ENSEA, Université Cergy Pontoise/
Université Badji Mokhtar- Annaba
Laboratoires ETIS, LABGED

Thèse

présentée en première version en vu d'obtenir le grade de Docteur,
spécialité « STIC »

par

Walid HARIRI

Contribution à la Reconnaissance/Authentification de Visages 2D/3D

Thèse soutenue le 05 Octobre 2017 devant le jury composé de :

M.	Med Tarek Khadir	Université d'Annaba, Algérie	(Président)
M.	Nadir Farah	Université d'Annaba, Algérie	(Directeur)
M.	Dan Vodislav	Université de Cergy Pontoise, France	(Co-directeur)
M.	Hamid Seridi	Université de Guelma, Algérie	(Rapporteur)
M.	Chaabane Djeraba	Université de Lille1, France	(Rapporteur)
M.	Hedi Tabia	ENSEA, France	(Examinateur)
M.	Mahmoud Boufaida	Université Constantine 2, Algérie	(Examinateur)
M.	Aladine Chetouani	Université d'Orléans, France	(Examinateur)

This thesis is dedicated to

my family, my friends and all who contributed to realize this work.

ACKNOWLEDGMENTS

First of all, I would like to thank my supervisors Pr. Nadir Farah and Pr. Dan Vodislav for giving me the opportunity to work in collaboration between LABGED and ETIS laboratories through the co-tutorship convention between UBMA and Cergy-Pontoise university. Their continuous guidance and support were always a source of inspiration for me. They created a motivating, enthusiastic and constructive environment. Their critical and insightful feedback greatly improved my work and quality of presentation.

I would like to express my gratitude to my thesis co-advisor Dr. Hedi Tabia who was very generous with his time and knowledge and consistently assisted me at every stage of my PhD. Thanks go also to and Pr. David Declercq an Dr. Abdallah Benouareth for their help.

Many thanks go to the people who I have been worked with starting by my office-mate Franklin and all my friends in ETIS laboratory for the insightful discussion and suggestions.

Finally, I am thankful to to my mother, father and sisters for their support and love, without them, this work will never be accomplished.

Contents

List of figures

Abstract

Title Contribution to 2D/3D Face Recognition/Authentication .

Abstract 3D face analysis including 3D face recognition and 3D Facial expression recognition has become a very active area of research in recent years. Various methods using 2D image analysis have been presented to tackle these problems. 2D image-based methods are inherently limited by variability in imaging factors such as illumination and pose. The recent development of 3D acquisition sensors has made 3D data more and more available. Such data is relatively invariant to illumination and pose, but it is still sensitive to expression variation. The principal objective of this thesis is to propose efficient methods for 3D face recognition/verification and 3D facial expression recognition. First, a new covariance based method for 3D face recognition is presented. Our method includes the following steps : first 3D facial surface is preprocessed and aligned. A uniform sampling is then applied on the face surface to localize a set of feature points, around each point, we extract a matrix as local region descriptor. Two matching strategies are then proposed, and various distances (geodesic and non-geodesic) are applied to compare faces. The proposed method is assessed on three datasets including GAVAB, FRGCv2 and BU-3DFE. In the second part of this thesis, we present an efficient approach for 3D facial expression recognition using kernel methods with covariance matrices. In this contribution, we propose to use Gaussian kernel which maps covariance matrices into a high dimensional Hilbert space. This enables to use conventional algorithms developed for Euclidean valued data such as SVM on such non-linear valued data. The proposed method have been as-

sessed on two known datasets including BU-3DFE and Bosphorus datasets to recognize the six prototypical expressions.

Keywords Face analysis, Identification, Verification, Covariance matrix, Geodesic distances, Face matching, Facial expression, Kernel-SVM, Classification.

Titre Contribution à la Reconnaissance/Authentification de Visages 2D/3D.

Résumé L'analyse de visages 3D y compris la reconnaissance des visages et des expressions faciales 3D est devenue un domaine actif de recherche ces dernières années. Plusieurs méthodes ont été développées en utilisant des images 2D pour traiter ces problèmes. Cependant, ces méthodes présentent un certain nombre de limitations dépendantes à l'orientation du visage, à l'éclairage, à l'expression faciale, et aux occultations. Récemment, le développement des capteurs d'acquisition 3D a fait que les données 3D deviennent de plus en plus disponibles. Ces données 3D sont relativement invariables à l'illumination et à la pose, mais elles restent sensibles à la variation de l'expression. L'objectif principal de cette thèse est de proposer de nouvelles techniques de reconnaissance/vérification de visages 3D et de reconnaissance d'expressions faciales 3D. Tout d'abord, une méthode de reconnaissance de visages en utilisant des matrices de covariance comme des descripteurs de régions de visages est proposée. Notre méthode comprend les étapes suivantes : le prétraitement et l'alignement de visages, un échantillonnage uniforme est ensuite appliqué sur la surface faciale pour localiser un ensemble de points de caractéristiques. Autours de chaque point, nous extrayons une matrice de covariance comme un descripteur de région du visage. Deux méthodes d'appariement sont ainsi proposées, et différentes distances (géodésiques / non-géodésique) sont appliquées pour comparer les visages. La méthode proposée est évaluée sur trois bases de visages GAVAB, FRGCv2 et BU-3DFE. La deuxième partie de cette thèse porte sur la reconnaissance des expressions faciales 3D. Pour ce faire, nous avons proposé d'utiliser les matrices de covariances avec les méthodes noyau. Dans cette contribution, nous avons appliqué le noyau de Gauss pour transformer les matrices de covariances en espace d'Hilbert. Cela permet d'utiliser les algorithmes qui sont déjà implémentés pour l'espace Euclidean (i.e. SVM) dans cet espace non-linéaire. Des

expérimentations sont alors entreprises sur deux bases d'expressions faciales 3D (BU-3DFE et Bosphorus) pour reconnaitre les six expressions faciales prototypiques.

Mots-clés Analyse de visage, Identification, Vérification, Matrice de covariance, distance géodésique, Appariement de visage, Expression faciale, Noyau, SVM, Classification.

Notations

Symbol	Definition/Explanation
SPD	Symmetric Positive Definite matrix
Sym_d^+	Manifold of SPD Matrices
$\mathbb{R}^d$	ensemble des vecteurs réels à d dimensions
$\mathcal{M}$	Manifold space
T_X	Tangent space on the Manifold point X
SVM	Support Vector Machines
ICP	Iterative Closest Point
d_g^2	Geodesic distance between two points on Sym_d^+
d_{ainv}^2	Affine-Invariant distance
d_{ld}	Log Determinant distance
$mlog$	Inverse-exponential map
d_α	The Alpha Divergence distance
d_{ot}	Optimal Transportation distance
$k1$	Min principal curvature
$k2$	Max principal curvature
r	Patch radius
P_i	Patch around the feature point p_i
$\mathcal{P}$	Set of patches P_i
f_j	Feature vector on the point p_j
exp	Ordinary matrix exponential
log	Ordinary logarithm exponential
$exp_X(y)$	Map the point of tangent vector y to the manifold point Y
$log_X(Y)$	Tangent vector y of the Manifold point Y on X
$\mathcal{H}$	Hilbert space
RKHS	Reproducing kernel Hilbert space
RBF	Radial basis function
$\mathcal{K}$	Kernel function
ROC	Receiver operating characteristic
CMC	Cumulative match characteristic

Author's publications

Journals

— **Hariri, W.**, Tabia, H., Farah, N., Benouareth, A., Declercq, D, 3D face recognition using covariance based descriptors. **Pattern Recognition Letters**, volume 78, pages 1-7, 2016.

— **Hariri, W.**, Tabia, H., Farah, N., Benouareth, A., Declercq, D, 3D Facial expression recognition using kernel methods on Riemannian manifold. **Engineering Applications of Artificial Intelligence**, Volume 64, Pages 25-32, 2017.

International conferences

— **Hariri, W.**, Tabia, H., Farah, N., Declercq, D., and Benouareth, A. Hierarchical covariance description for 3D face matching and recognition under expression variation. In International Conference on 3D Imaging (**IC3D**), pages 1-7. IEEE, 2016.

— **Hariri, W.**, Tabia, H., Farah, N., Declercq, D., and Benouareth, A. Geometrical and visual feature quantization for 3D face recognition. In **VISAPP**, 12th Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2017.

Local communication

— **Hariri, W.**, Tabia, H., Farah, N., Benouareth, A., Declercq, D, La Reconnaissance de visages 3D en utilisant des descripteurs de covariance. Journée de l'AS Visage, geste, action et comportement, Paris, Gdr-Isis 2016.

1 Introduction

Sommaire

Face analysis including face recognition and facial expression recognition finds applications in various fields such as human-computer interaction, security systems, etc. Pioneer researchers focused on 2D methods to handle the problem of face recognition and facial expression recognition. These methods have many limitations especially with the presence of pose and illumination variations which deteriorate the recognition performance. Since 3D data become more and more available with the development of the acquisition systems, face analysis systems using 3D date become feasible. In this thesis, we handle the problem of 3D face recognition and 3D facial expression recognition.

1.1 Face recognition

With the increasing need for efficient security systems, many research investments have been made for recognizing and authenticating

human subjects. Traditionally, there are two different methods to recognize/authenticate human subjects. The first method depends on what the user knows such as passwords, phone number, birth dates, etc. The problem with this method of authentication is that these information can easily be stolen or guessed.

The second method is based on the possession of a physical devise (e.g. RFID card). This device may be swiped or entered to allow access to a resource. The problem with this kind of authentication is that the device could be easily lost, stolen, or broken. These two authentication methods can be used in a complementary manner in order to obtain a high security as in the case of the credit card.

The biometric characteristics are an alternative solution to the two previous authentication methods. They enable reliable and efficient identity management systems since they are permanent, universal and easy to access. The use of biometric traits to control a subject's identity rather than passwords and tokens is more reliable to improve the security systems than controlling what he possesses or what he knows. Additionally, biometry-based procedures obviate to remember a PIN number or carry a badge.

Among the various human characteristics used in biometric systems, we can find iris, face, fingerprint, gait or DNA. The system constraints and requirements should be taken into account as well as the technical, social and ethical factors (Introna et Nissenbaum 2010). For instance, while fingerprint is the most wide-spread biometric technique, it requires strong user collaboration. Similarly, iris recognition, although being very accurate, highly depends on the image quality and also requires active participation of the subjects.

Face recognition can be a good alternative compared to other biometric techniques. It became an active research area due to its favorable compromise between accessibility and reliability. It allows identification at relatively high distances of unaware subjects that do not have to cooperate.

Given a 2D or 3D image or a video sequence, the face recognition problem can be briefly interpreted as an identification or a verification of one or more persons.

A typical face recognition system is composed of four modules :

1. **Sensor module :** which captures the face data (2D/3D) ;
2. **Feature extraction module :** which processes the output of the sensor to extract a discriminatory feature set ;
3. **Matching module :** in which the extracted features are compared against the features of the stored scans, and matching scores between faces are computed ;
4. **Decision making module :** the recognition is based on the matching scores between the face to be recognized and the stored face in the database.

Every face recognition system has two phases of operation. The first phase is called enrollment in which an individual uses the system for the first time as presented in Figure 1.1. At the enrollment phase, facial information of the user is stored by the system, which forms a database so-called *Gallery*. In the second phase, given a face to be recognized called *probe*, the system sorts all face gallery according to their similarity to the probe.

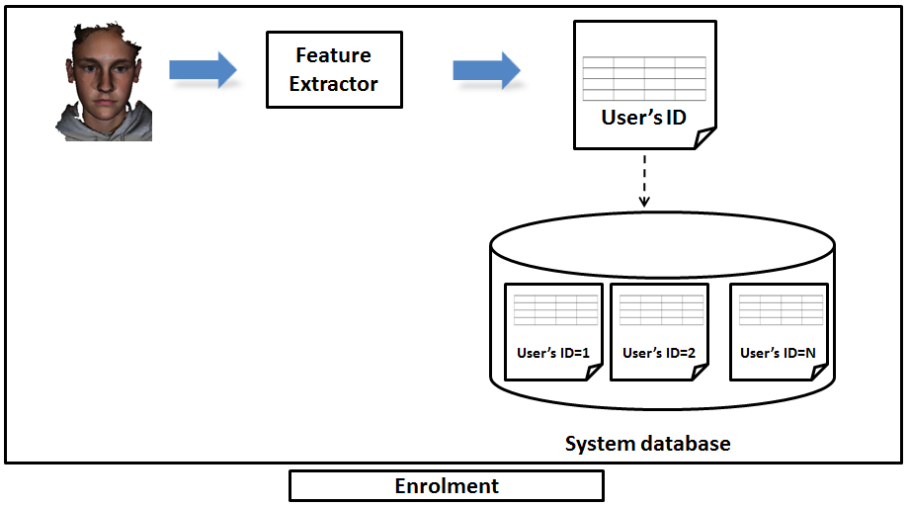


FIGURE 1.1 – *Enrolment process.*

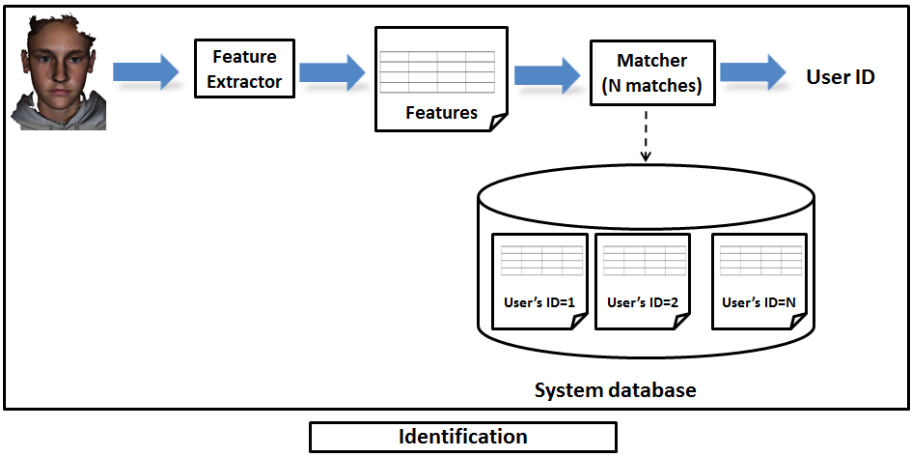


FIGURE 1.2 – *Identification process.*

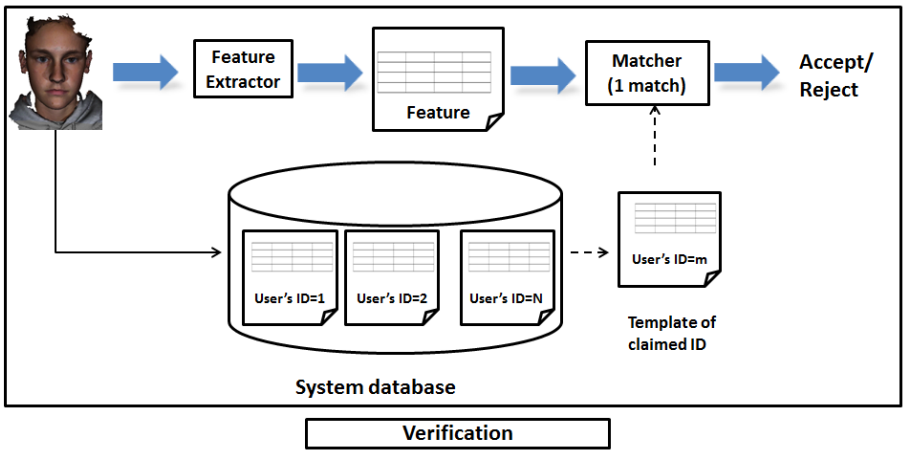


FIGURE 1.3 – *Verification process.*

1.1.1 Face recognition system design

Face recognition systems may process the probe sample either in "verification" or "identification" scenario. In the verification scenario, the probe is compared to the claimed template (one to one comparison) for validation and it is either accepted or rejected. In the identification scenario, it is compared to all reference faces in the gallery (one to many comparison) to answer the question of whom this face belongs to. Figure 1.2 and Figure 1.3 introduce the identification and the verification scenarios respectively. If the face in the probe is already registered in the reference database, it is called **closed-set identification**. Otherwise, it is **open-set identification**.

1.1.2 Performance evaluation of face recognition systems

Each scenario has its own set of performance metrics. In face verification, if the match score meets the threshold criteria, the identity claim is accepted, otherwise, it is rejected. This setting leads to four possible outcomes (see Figure1.4) :

1. **True accept :** The person is who he claims to be (genuine) and his claim is proved.
2. **True reject :** The person is not who he claims to be (imposter) and his claim is disproved.
3. **False accept :** The person is not who he claims to be and his claim is proved.
4. **False reject :** The person is who he claims to be and his claim is disproved.

Threshold based decisions always introduce a tradeoff to be considered. In the case of face verification, if the threshold is too high, **False reject rate** (FRR) might increase since more legitimate claims would be rejected. If the threshold is too low, acceptance of false claims will be more likely, increasing **False acceptance rate** (FAR). This relationship is shown with **Receiver operating characteristic** (ROC) graph which represents the probability of true acceptance versus probability of false acceptance. ROC curves are also used to measure performances of open-set identification systems. Instead of true acceptance rate, detection and identification rate is calculated and plotted against FAR. Detection and identification rate is the percentage of the probe samples represented in the gallery that are correctly accepted and identified.

With face identification scenario, the **rank** performance measure is used. In rank-1 case, a probe face is identified as the first identity in the list of subjects sorted in decreasing order of computed similarity scores. Correspondingly, rank-n systems examine the top n matches. Thereby, identification rate is a function of rank.

On the other hand, in closed-set face identification, the performance is computed as a function of rate only and reported on **Cumulative match characteristic** (CMC) curves. CMC plots the size of the rank order list against the identification rate.

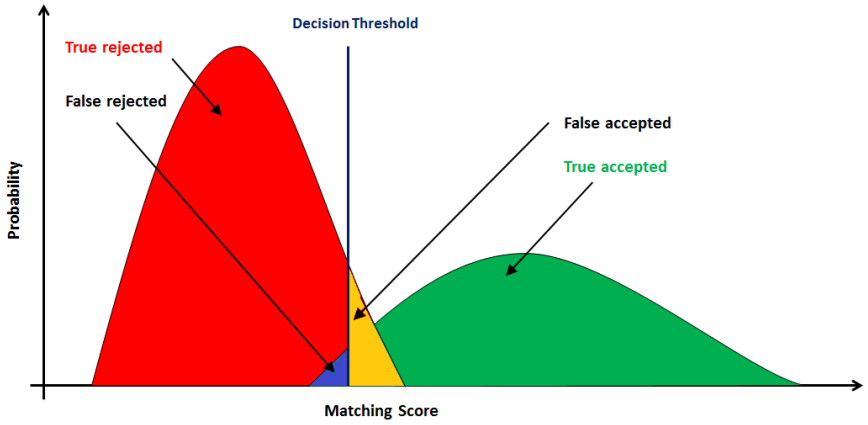


FIGURE 1.4 – *Evaluation metrics of face recognition systems.*

Despite the the recent advances achieved in face recognition in the last decades, it still suffers from intra-class variations (variations between scans of the same subject) due to various factors in real-world scenarios (e.g. illumination variations, pose changes, expression variations, occlusion, age, poor image quality), and inter-class variations (similarity between scans of different subjects). In the Face Recognition Vendor Test 2002 (Phillips et al. 2003), it was demonstrated that using 2D intensity or color images, a recognition rate higher than 90% could be achieved under controlled variations. However, with the introduction of aforementioned variations, the performances deteriorated (Al-Osaimi et al. 2012, Ocegueda et al. 2013, Tang et al. 2013, Petrovska-Delacrétaz et al. 2008).

After the availability of 3D scanners and 3D face databases, many researchers focused their energies toward 3D face recognition in order to utilize the more precise information associated with a 3D facial shape (Bowyer et al. 2006, Mohammadzade et Hatzinakos 2013, Zhao et al. 2011). Indeed, 3D model retains all the information about the face geometry. Moreover, 3D face recognition also growths to be a further evolution of 2D recognition problem, because a more accurate representation of the facial

features leads to a potentially higher discriminating power (Abate et al. 2007). In a 3D face model, facial features are represented by local and global curvatures that can be considered as the real signature identifying persons. The 3D facial representation seems to be a promising tool to deal with many of the human face variations such as illumination (Al-Osaimi et al. 2012, Smeets et al. 2010) and pose changes (Ocegueda et al. 2013).

Various 3D based methods have been proposed in the literature to tackle the problem of 3D face recognition. Indeed, some methods perform very well only on faces with controlled environment where pose, illumination and other factors are controlled. Some others try also to deal with the different face variations.

The first aim of this thesis is to propose a 3D face recognition method that is robust under the different variations.

1.2 Facial expression recognition

Closely related topic to face recognition is the automatic facial expression recognition which has attracted much attention from behavioral scientists since the work of Darwin et al. (1998). In his book *The Expression of the Emotions in Man and Animals*, Darwin asserted that facial expressions were universal and innate characteristics. The known psychologist Mehrabian (1972) has studied the effect of the verbal and non-verbal messages, which reported that face to face communication is governed by :

1. 7% (verbal : words account)
2. 55% (facial : expression, posture, gesture)
3. 38% (vocal : tone of the voice accounts)

In the 20th century, two other well-known psychologists Paul Ekman and Wallace Friesen have developed in (Ekman et Friesen 1971) the Facial Action Coding System (FACS) which is a detailed technical guide that explains how to categorize behaviors based on muscles. In 2003, Fasel et Luettin (2003) presented the sources of facial expressions which include

mental states, physiological activities and verbal/non-verbal communications, as shown in Figure 1.5. Mental state is one of the main sources, including felt emotions, conviction and cogitation. Physiological states such as pain, tiredness also influence unconscious face muscle activities appearing in forms of expressions. Verbal communication such as illustrators, listener responses ; and nonverbal communication such as unfelt emotion and emblems can also cause facial expressions. Different groups of primary emotions have been presented by psychologists as presented in Table1.1.

Psychologist	**Categories of emotions**
Ekman et Friesen (1971)	anger, fear, disgust, sadness, happiness, surprise.
Plutchik (2003)	acceptation, anger, anticipation, disgust, fear, happy, sad, surprise.
Tomkins (1962)	anger, interest, contempt, disgust, anxiety, fear, happy, shame, surprise.
James (1884)	fear, grief, love, rage.

TABLE 1.1 – *Emotion categories.*

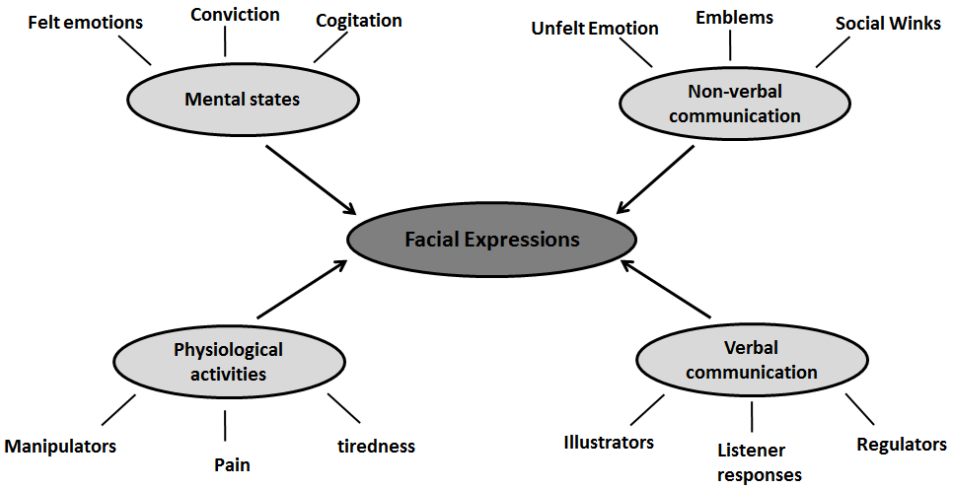


FIGURE 1.5 – *Sources of Facial Expressions (Fasel et Luettin 2003).*

Based on this finding, facial expressions have been studied by clinical and social psychologists, medical practitioners. Recently, with the advances in computer graphics, computer vision and robotics, it has become a worthwhile topic to study in computer sciences especially for human computer interaction (HCI) systems (Liu et al. 2009). Among the aforementioned groups of emotions, human computer interaction research community is focusing on recognizing the six basic expressions defined by (Ekman et Friesen 1971).

Facial expression data as well as face data exhibit large inter-class and intra-class variations which make the FER a difficult task. The second aim of this thesis is to propose a 3D FER system that efficiently classifies the six prototypical expressions regardless to their identity. Main contributions of the thesis are summarized below.

1.3 Original contribution

The two main contributions conducted in this PhD thesis involve :

3D face recognition using covariance based descriptors in this contribution, we handle the problem of 3D face recognition/verification under expression, pose and partial occlusion variations. To this end, we proposed a new covariance based method and demonstrate its efficiency as local descriptors for 3D face recognition. Since covariance matrices are not elements of an Euclidean space, they are elements of a Lie group, which has a Riemannian structure. Therefore, matching with covariance matrices requires the computation of geodesic distances on the manifold using a proper metric. In this contribution, we have compared the performance of our recognition system using several distances for covariance matrices including geodesic distances. Finally, for the recognition step, we considered the two following solutions : (i) the Hungarian solution for matching unordered sets of covariance descriptors from two 3D faces, and (ii) the mean of distances between each pair of homologous patches in the two 3D faces. The proposed face recognition method has been assessed on three challenging datasets including Gavab, FRGCv2 and BU-3DFE.

3D facial expression recognition on Riemannian manifold in order to accurately recognize the six prototypical expressions (i.e. Happiness, Angry, Disgust, Sadness, Surprise and fear) from 3D facial data regardless to the face identity, we proposed a new kernel based method on Riemannian manifold. Inspired by the successful use of kernel methods on Riemanniam manifold in order to embed the non-linear manifold into a Hilbert space, which make the conventional computer vision and machine learning algorithms applicable, we applied Gaussian kernel on Sym_d^+ by replacing Euclidean distance by a proper geodesic distance on manifold. Since in our framework, covariance matrices are considered as unordered set of descriptors, therefore we used a kernel on sets rather than directly

using Gaussian kernel in the SVM classifier. For this end, we build a global kernel function so that one can compare two facial expressions by using the covariance descriptors. The proposed 3D FER method have been assessed on two challenging datasets including Bosphorus and BU-3DFE datasets.

The rest of this manuscript is laid out as follows :

In chapter 2, we first present the state-of-the-art of 3D face analysis including the existing methods proposed to tackle the the problem of face recognition under different variations (e.g. expression, pose, occlusions). Second, we review the state-of-the-art methods for 3D facial expression recognition.

In chapter 3 we present the covariance based method proposed to tackle the problem of 3D face recognition under different variations.

Chapter 4 is dedicated to the proposed 3D facial expression recognition method. This method uses kernel methods on Riemannian manifold in order to classify the six prototypical facial expressions.

Chapter 5 summarizes the main contribution results of our thesis, and gives possible future directions.

2 3D Face analysis- State of the art

Sommaire

2.1 Introduction

Face recognition is the most commonly used biometric technique, it has moved to the forefront by offering a good compromise between effectiveness and acceptability. Besides being non-invasive and natural, it holds the promise of recognition at a distance without the cooperation or knowledge of the subject. In a typical face recognition system, given an image,

firstly, face is detected and segmented from the background. Next, several landmarks are localized and used in the next steps. This is followed by features extraction and finally the matching process to compare faces. Figure 2.1 presents an overview of the face recognition system. Since the first face recognition system has been presented in 1973 by Kanade (1973), it has become a very popular area in computer vision on account of rapid advancements in image capture devices, increased computational power and large variety of its applications. With the increasing popularity, face recognition systems reached recognition rates greater than 80% in constrained situations even exceeding human performance, especially in the case of very large galleries. However, Face Recognition Vendor Test (Phillips et al. 2003) found that in real world scenarios where face images can include a wide range of variations, performances degrade very rapidly. As of those variations, we find principally illumination, pose, expression, occlusion and age. Pose and illumination variations are two major sources of degradation in recognition performances. As the pose of a subject or direction of illumination deviates from the frontal, it often causes face image differences that are larger than what conventional face recognition can tolerate. Extensive efforts have been put to achieve pose/illumination invariant face recognition. Facial shape deformation caused by expression variation is also a grand challenge in face recognition systems.

Facial expressions form a significant part of our nonverbal communications, and understanding them is essential for many applications such as human-computer interaction (HCI), interactive games, computer-based learning, entertainment, etc. The most used expressions are : Happiness, Angry, Disgust, Sadness, Surprise and Fear. In order to build human-like emotionally intelligent HCI devices, researchers are trying to include emotional state in such systems.

The recent development of 3D acquisition sensors has made 3D data more available, and this kind of data comes to alleviate the problems inherent in 2D data such as illumination, pose and scale variations as well

as low resolution. 3D face datasets become more and more available, providing the worldwide researchers of face and facial expression recognition community a large scale data for training and evaluating their approaches.

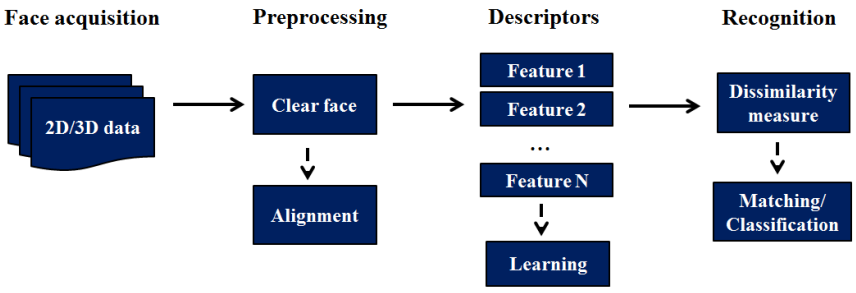


FIGURE 2.1 – *Face recognition system.*

In the following, we present the two modes of 3D face data acquisition and different 3D face representations. Next, we review some well-know 3D face datasets and the 3D face preprocessing and alignment steps.

3D face acquisition Different techniques for capturing 3D face/facial expression data have been employed including the use of single image reconstruction (Kemelmacher-Shlizerman et Basri 2011, Wang et Lai 2011), structured light technologies (Jarvis 1993, Huang et al. 2003). Two methods for stereo reconstruction algorithms have been also used : photometric stereo (Woodham 1980), and multiview stereo (Seitz et al. 2006, Beeler et al. 2010, Yin et al. 2008, Benedikt et al. 2010). Each technique has its own advantages, limitations and cost. These techniques are used to create 3D point clouds, sampled from the surface of the subject and they can be classified mainly into two types : non-contact scanners which don't require the physical participation of the subject, they can be further divided as passive and active sensors. Contact scanners on the other hand detect the range through physical touch. These scanners are out of our scope due to the inability of utilization in face recognition field. Passive and active range scanning technologies will be described in the following.

— **Passive sensing :** inspired by human visual system, a number of passive cues have long been known to contain information on 3D face such as shading, perspective, focus, stereo, motion parallax,

occluding contours, etc. Assuming that the sensor simply records light that already exists in the scene, 3D scene description is derived from 2D images by analyzing the reflectance proprieties. Main advantages of passive techniques include their ability to recognize non-cooperative persons at a distance. 3D face model is generated from sequence of images and utilized in person identification at 3, 6 meters by Medioni et al. (2007). In (Rara et al. 2009), authors have proposed a framework for face recognition at 15 meters based on sparse-stereo reconstruction. The setup consists of a stereo pair of high resolution cameras with adjustable baseline where user can pan, tilt, zoom and focus the cameras to converge the center of the cameras field of views on the subject's nose tip.

Although these advantages, these techniques have a limitation of uniform appearance resulting in low-accuracy reconstruction in their application to faces. In addition, the ambient light (i.e. combination of light reflections from various surfaces to produce a uniform illumination) affects significantly the ability of the system to successfully extract all the desired features unless controlled lighting is used. To overcome the limitations on accuracy, model-based approaches have been proposed in (Zhang et al. 2004, Fua 2000) to constrain face reconstruction in regions of uniform appearance.

Moreover, in order to overcome limitations of uniform appearance of the face allowing accurate reconstruction, active sensing has become widely used for acquisition of face shape due to the increases in digital camera resolution (Kittler et al. 2005).

— **Active sensing :** various technologies have been proposed which utilize some kind of emission such as laser, infrared structured, modulated light. Active systems can easily measure surfaces in most environments due to their own illumination. These systems work on two principles : time-of-flight ; and triangulation. Time-of-flight sensors measure the time taken for the projected illumination

pattern to return from the object surface. This sensor technology requires nano-second timing to resolve surface measurements to millimetre accuracy. A number of commercial time-of-flight systems are available based on scanning a point or stripe across the scene. Triangulation systems, on the other hand, use a focused beam of light to probe the environment and by tracing a line of sight through illuminated pixel.

Under uncontrolled illumination conditions (i.e. outdoors), active sensor systems suffer from the background lighting problem which can make the projected pattern less visible and cause loss of accuracy. Nevertheless, these systems are more robust when dealing with controlled illumination conditions (i.e. indoors). Moreover, for environments without direct sunlight, 3D data construction of the scene can be achieved within seconds with very high reliability and accuracy (El-Hakim et al. 1995).

3D face representations In this section, we present the main surface representations for 3D face which could be inter-convertible.

— Point cloud representation is the simplest form to represent 3D facial surface. It contains a collection of unstructured coordinates denoted by x, y, z (Pears et al. 2012). Most of the scanners use this representation in order to store the captured 3D facial information. Sometimes texture attributes are also concatenated to the shape information. In this case, the representation simply becomes x, y, z, p, q, where p and q are spatial coordinates. The disadvantage of this representation is that the neighborhood information is not available as each point is simply expressed as three/five attribute coordinates vector (Gokberk et al. 2008). Often the point cloud data is fitted to a smooth surface to avoid drastic variations due to noise. Figure 2.2(b) presents the point cloud representation for a sample face from the BU-3DFE dataset.

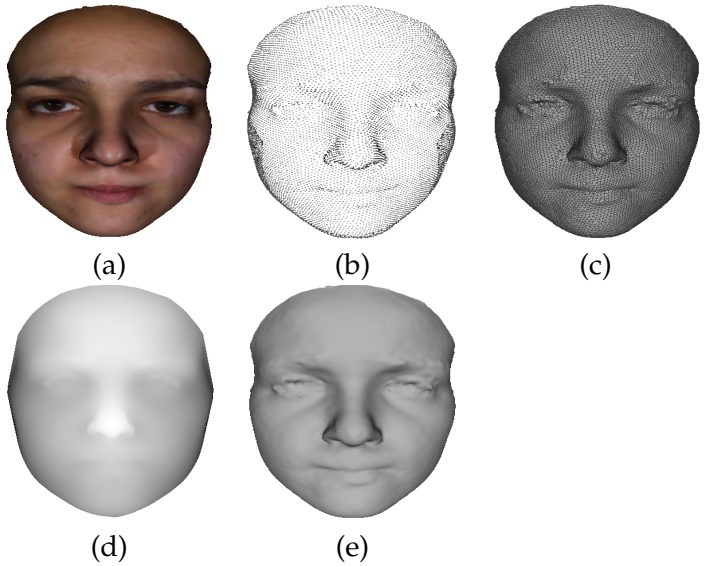


FIGURE 2.2 – *a) Texture map, b) Point cloud, c) Triangular mesh, d) Depth map, e) 3d rendered as shaded model.*

— 3D Mesh representation uses pre-computed and indexed local information about the 3D surface. It requires more memory and storage space than point cloud representation, but it is more preferred as it is flexible and more suitable for 3D geometric transformations such as translation, rotations and scaling. Each 3D polygonal mesh is expressed as a collection of mesh elements : vertices (points), edges (connectors between vertices) and polygons (shapes formed by edges and vertices). Different methods have been introduced to build a polygonal mesh from point cloud using PCA in (Xu et al. 2004), and medial axis transform approximation (Amenta et al. 2001). Figure 2.2(c) presents the mesh representation for a sample face from the BU-3DFE dataset.

— Depth image representation is also called 2.5D or range image representations Bowyer et al. (2006). 2.5D images are a conventional 2D representation, nevertheless, pixels represent the distance between the camera and the observed object. Since it is a 2D representation, many existing 2D image processing approaches can readily be applied to this representation. Figure 2.2(d) presents the point cloud representation for a sample face from the BU-3DFE dataset.

Datasets description Most of the well-known 3D face datasets were collected using laser-based sensors.

— The BU-3DFE dataset (Yin et al. 2006) contains 3D face scans and associated texture images of 100 subjects, displaying six prototypical expressions (anger (AN), disgust (DI), fear (FE), happiness (HA), sadness (SA), and surprise (SU)) at four different intensity levels. A neutral expressive face for each subject is also provided in the dataset. Thus there are a total of 2500 3D faces. The resolution of the 3D models is comprised between 20,000 and 35,000 polygons, depending on the size of the subject's face. Moreover, scans were accompanied by a set of metadata including the position of 83 facial feature points on each facial model, as depicted in Figure 2.3. Examples of 3D faces from this dataset can be seen in Figure 2.5.

— The GAVAB dataset (Moreno et Sanchez 2004) contains 549 three-dimensional facial surface images corresponding to 61 individuals (45 male and 16 female). It includes many variations with respect to the pose of each individual. Each subject in the GAVAB dataset was scanned 9 times for different poses and facial expressions. The whole set of individuals are Caucasian and most of them are aged between 18 to 40 years. There are systematic variations over the pose and facial expression of each person. In particular, 2 frontal and 4 rotated images without any facial expressions. There are also 3 frontal images in which the subject presents different and accentuated facial expressions (laugh, smile and a random expression chosen by the user). Figure 2.6 shows an example of faces taken from this dataset. In this experiment, we will deal only with expressive faces to assess the performance of our proposed method under this facial deformation.

— The FRGCv2 database (Phillips et al. 2005) is one of the most comprehensive and popular datasets, containing 4007 3D face scans of 466 different persons, the data were acquired using a minolta

910 laser scanner that produces range images with a resolution of 640×480. The scans were acquired in a controlled environment and exhibit large variations in facial expression and illumination conditions but limited pose variations. The subjects are 57% male and 43% female, with the following age distribution : 65% 18-22 years old, 18% 23-27and 17% 28 years or over. The database contains annotation information, such as gender and type of facial expression. Figure 2.7 presents faces for the same subject from FRGCv2 dataset.

— The Bosphorus database (Savran et al. 2008) is a multi-expression, multi-pose 3D face database enriched with realistic occlusions. The database consists of 4666 scans from 105 subjects and is acquired with the Inspeck Mega Capturor II 3D scanner (structured-light technique) leading to 3D point clouds of approximately 35 000 points. Each scan has been manually labelled for 24 facial landmark points such as nose tip, inner eye corners, right nose peak, left nose peak, etc. Subjects are aged between 25 and 35 in various poses, expressions and occlusion conditions. Only 65 subjects posed all the six prototypical facial expressions (i.e., angry, happiness, fear, sadness, surprise, and disgust) and neutral. Figure 2.8 presents seven expressive faces for the same subject from Bosphorus dataset (6 expressions + 1 neutral scan).

We present in Table 2.1 a description of the aforementioned datasets and their characteristics.

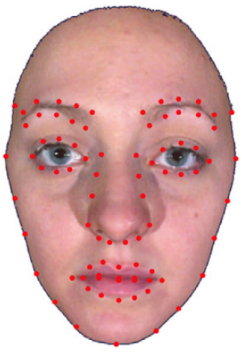
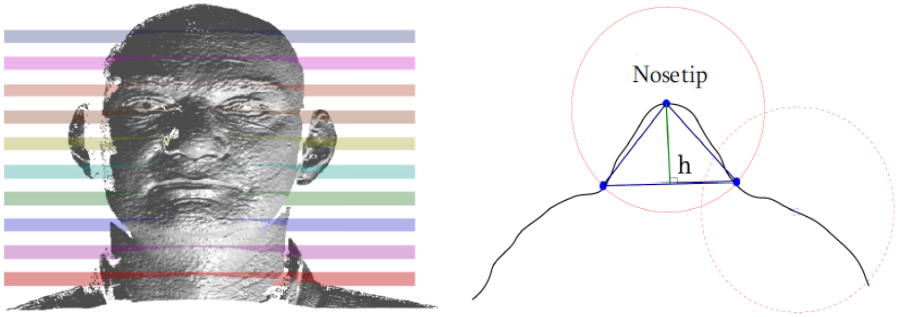


FIGURE 2.3 – *The 83 facial points given in the BU-3DFE database (Sandbach et al. 2012).*



(a) (b)

FIGURE 2.4 – *An illustration of 3D facial nosetip detection. Horizontal planes for 3D facial scan slicing (a). Horizontal facial profile (b).Guo et al. (2016)*

3D face preprocessing and alignment 3D face data acquired by a 3D sensor needs to be processed before feature extraction stage. Generally, these data include non-facial regions (e.g. hair, clothing, shoulders, etc) which should be removed. To do so, landmark points (e.g. nose tip, eye corners) are often localized to detect the facial surface. Authors in Mian et al. (2006) detected nose tip and cropped face via sphere centered at the nose tip. Guo et al. (2016) have detected the nosetip to remove undesired points outside the 3D facial region. First, a set of horizontal planes are used to slice the 3D facial scan, resulting in a set of horizontal profiles of the 3D face, as presented in Figure2.4 (Left). Then, a set of probe points are located on each profile and a circle is placed at each point, resulting in two intersection points with the horizontal profile, as shown in Figure2.4 (Right). A triangle is formed by the probe point and the two intersection points. The probe point with the largest altitude h of its associated triangle along the profile is considered to be a nosetip candidate. Other nose tip detection methods can be found in (Chew et al. 2009, Segundo et al. 2007, Colombo et al. 2006, Xu et al. 2006).

Holes are parts of missing data which the sensor couldn't capture because of undesired objects such as eyebrows, hand, hair, etc. In order to recover these holes, two solutions can be applied, interpolation techniques or using facial symmetry. To deal with small holes, linear interpolation can be sufficient, otherwise, cubic interpolation is more accurate.

Dataset	Sensor	Number of subjects	Male	Female	Total scans	Expressions	Missing data	Occlusion
BU-3DFE (Yin et al. 2006)	Laser sensor	100	44	56	2500	Yes	No	No
GAVAB (Moreno et Sanchez 2004)	Laser sensor	61	45	16	549	Yes	Yes	Yes
FRGCv2 (Phillips et al. 2005)	Laser sensor	466	264	202	4007	Yes	No	No
Bosphorus (Savran et al. 2008)	Structured light	105	60	45	4666	Yes	Yes	Yes

TABLE 2.1 – *Description of the 3D face datasets.*

Facial symmetry is also used to recover large holes using the non-occluded side of the face. For example, Colombo et al. (2011) proposed a detection and restoration strategy for the recognition of three dimensional faces partially occluded by external objects. They considered any part of the acquired 3D scene that does not look like part of a face and lies between the acquisition device and the acquired face to be a generic occlusion (i.e. hole). Restoration of occluded regions exploits the information provided by the non-occluded part of the face to recover the whole face, using an appropriate basis for the space in which non-occluded faces lie.

The cropped data can also be affected by noise caused by imaging conditions such as illumination and surface texture. This noise can be removed using median filter. The basic idea of the median filtering consists of simultaneous replacing every pixel of an image with the median of the pixels contained in a window around the pixel (Yagou et al. 2002).

Once 3D face has been preprocessed, the next step is to deal with pose differences. Since faces can be captured with different poses, their comparison become difficult and thus, it needs alignment step.

Iterative Closest Point (ICP) algorithm (Besl et McKay 1992, Chen et Medioni 1991, Zhang 1994) is mostly used to find correspondences between two 3D shapes and align them. However, when the pose difference is high, a good initialization is required to avoid local minimum.

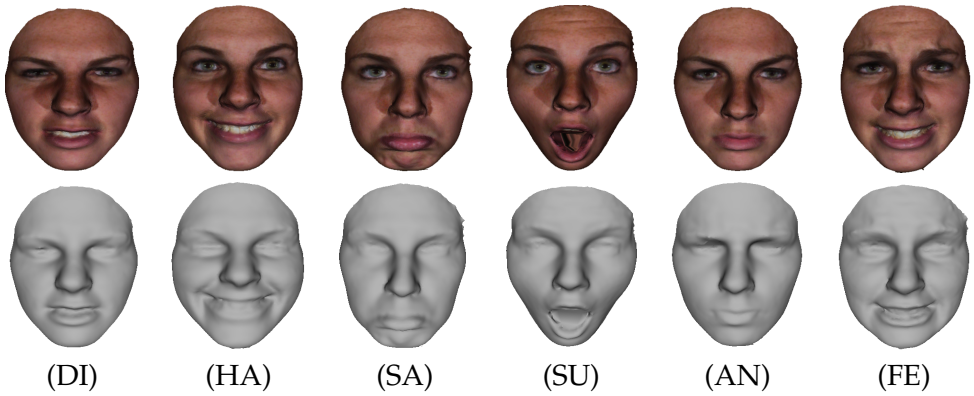


FIGURE 2.5 – *Face expressions from BU-3DFE dataset for the same subject. 3D textured models in first row and 3D shape models in second row.*

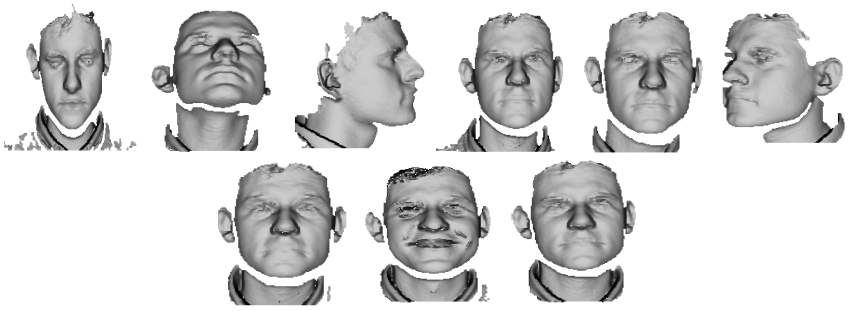


FIGURE 2.6 – *3D scans of the same subject from the GAVAB dataset.*

FIGURE 2.7 – *3D scans of the same subject from the FRGCv2 dataset..*

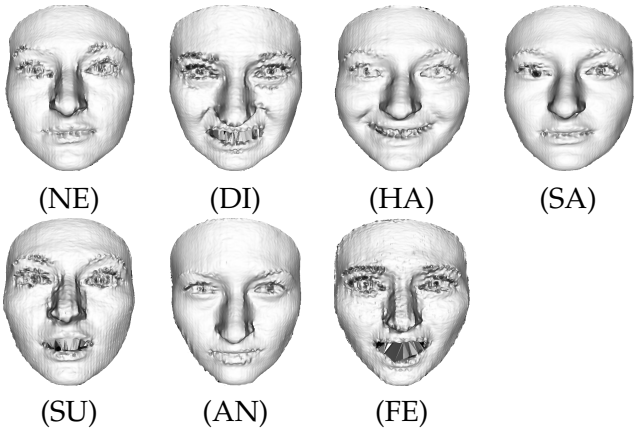


FIGURE 2.8 – *Face expressions from Bosphorus dataset for the same subject.*

2.2 3D face recognition- State of the art

In the following, we introduce 3D face challenges and the existing methods proposed to address these issues.

2.2.1 Challenges of 3D face recognition

When acquired in non-controlled conditions, scan data are often affected by many factors : pose expression, occlusion, illumination, weather and so on. In the following we briefly describe these variations.

Expression challenge

Facial expression variations are reported as one of the main challenges of face recognition, since it is generated by facial muscle contractions which result in temporary facial deformations in both facial geometry and texture. Thus, expressive faces complicate the face recognition by creating higher intra-class variance than inter-class variance which can dramatically deteriorate the recognition performance as illustrated in Figure 2.8.

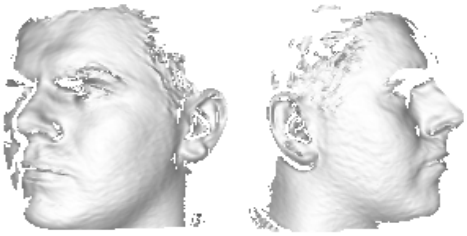


Figure 2.9 – *Profile faces from GAVAB dataset.*

Figure 2.10 – *Profile faces from Bosphorus dataset.*

Partial occlusion and pose challenges

Occlusion Variation means that only partial faces are available which also degrade the face recognition performance. These occlusions can be classified into two categories :

— **External occlusions** : caused by the non-availability of 3D facial data due to external objects such as sunglasses, hats, eyeglasses, or face may be partially covered by hair, or parts of cheeks due to a bad angle for laser reflection and other undesired regions. Figure 2.11 presents partial occlusions for 3D faces from Bosphorus dataset.

— **Internal occlusion** : for a non-frontal pose of the subject, some parts of the face may not be captured during the scan. These results in missing data are referred to as internal occlusion. Figure 2.9 presents right and left profile scan from GAVAB dataset. Figure 2.10 presents six profile faces from Bosphorus dataset.

Although many researchers dealt with expression variations, very few have attempted occlusion variation problem for 3D face recognition. The first explorations for partial occlusion challenge was in Brunelli et Poggio (1993), Pentland et al. (1994), Beymer (1994). Nevertheless, the substantial facial appearance change caused by pose variation continues to challenge the state-of-the-art face recognition systems. Essentially, it results from the complex 3D structure of the human head.

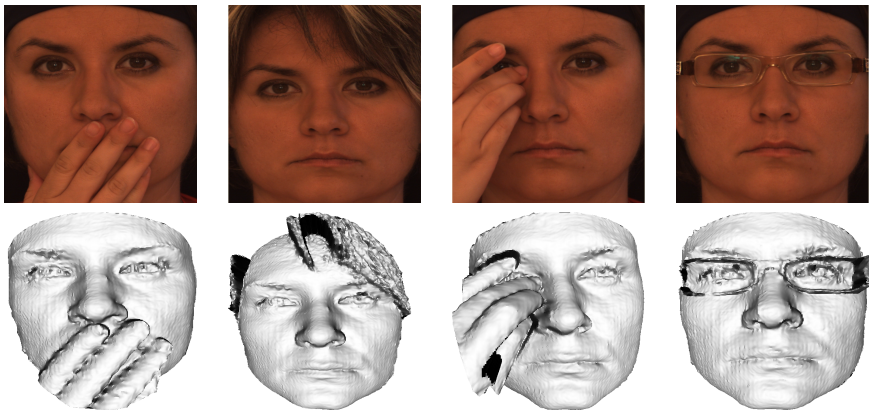


FIGURE 2.11 – *Partial occluded faces from Bosphorus dataset. Texture in first row and 3D shape in second row.*

2.2.2 3D face features

In order to address the aforementioned challenges, various methods have been proposed in the literature to efficiently describe faces. A 3D face model captures geometrical information of the facial surface. Various shape related features can thus be extracted from 3D face models such as : normals, binormals, tangent vectors or curvatures, all of which describe shape variations over local patches inspired by differential geometry of 3D surfaces. Therefore, different features have been extracted to deal with the facial variations. We classify them according to the challenge to be handled.

Expression-invariant features

Various methods have been proposed to handle the problem of expression variation. Statistical models have been widely used where the most popular is the Principal component Analysis (PCA) model. Al-Osaimi et al. (2009) employed the PCA to learn and model expression deformations. The generic PCA deformation model is built using non-neutral faces of distinct persons. The expression deformation templates are used to eliminate the expressions from non-neutral face scan. A multi-resolution PCA model has been proposed by ter Haar et Veltkamp (2010), they used a limited collection of facial landmarks along with neutral and expression scans. A single morphable identity model and seven isolated morphable expression models per subject are then built. Expression is then neutralized and coefficients of identity model are utilized for face recognition. In (Russ et al. 2006), a statistical model is built using the correspondence information. A PCA shape model can deal with expressions by including faces with expression in the training data.

Other methods based on the assumption that deformation caused by expression variation is isometric, meaning that the deformation preserves lengths along the surface (i.e. surface distances are relatively invariant to small changes in facial expressions), and therefore help generate features

that are robust to facial expressions. Drira et al. (2013) proposed a geometric framework for analyzing 3D faces recognition under different variations. They proposed to represent facial surfaces by radial curves emanating from the nose tips and use elastic shape analysis of these curves to compare faces. They used the elastic Riemannian metric to measure facial deformations to handle the large facial expression variation.

Similar approach have been proposed by Lee et Krim (2017) using deformed circular curves. The shortest geodesic distances between the reference point (e.g. nosetip) and a point on the curve is computed to generate a matrix or in one-dimensional function. The functions are compared to each other to measure the similarity between faces. Experiments have shown that there is little difference in the geodesic distance between the same face with different expressions (intra-class difference). Whereas, different face models of different people have shown a low similarity due to the shape of facial curves (inter-class difference).

Another algorithm is proposed by Sun (2015) which measures the minimum possible distortion when trying to isometrically embed one facial surface into another. A geodesic polar parametrization of the facial surface is proposed in (Mpiperis et al. 2007) in which authors have studied local geometric attributes under this parameterization. They assumed with this parameterization that the intrinsic surface attributes do not change under isometric deformations. To deal with the open mouth problem, they modified the parametrization by disconnecting the top and bottom lips. Berretti et al. (2010a) have encoded the geometric information of the 3D face surface in the form of a graph. Nodes of the graph represent equal width iso-geodesic facial stripes which provide a representation of local morphology of the face.

Moreover, various methods use only regions that are not or not much affected by expressions. Guo et al. (2016) presented 3D face by a set of keypoints and their associated local feature descriptors to achieve robustness to expression variations. To measure the dissimilarity between faces,

authors have computed the number of feature matches, the average distance of matched features, and the number of closest point pairs after registration. The global dissimilarity is then obtained by the fusion of the similarity metrics. Maiti et al. (2014) have extracted the T-region from the face to get the facial region having minimum variation with expression. For each region, they have extracted the wavelet coefficients and a dictionary learning using K-SVD. Moeini et al. (2014) have extracted rigid parts of the face from both the texture and depth image based on 2D facial landmarks. Gabor filter was then applied to the extracted feature vectors from texture depth images. Finally, classification is applied using the Support Vector Machine. Shape index and spherical bands on the human face are used in (Ming 2015) to segment a group of regions on each 3D facial point cloud. Then the corresponding facial areas are projected to regional bounding spheres to obtain regional descriptor.

Lei et al. (2013) divided the 3D facial into three regions according to their deformations that are caused by facial expressions as follows : rigid (i.e. nose region), semi-rigid (i.e. eyes-forehead region) and non-rigid (i.e. mouth region). Only regions which are relatively less influenced by the deformations caused by facial expressions (i.e. rigid and semi-rigid regions) are considered for features extraction and classification.

Li et Da (2012) split the face surface into six regions which are forehead, left mouth, right mouth, nose, left cheek and right cheek. In order to choose the regions to use for matching, authors extracted facial curves in these regions to map facial deformation.

Pose-invariant and partial occlusion invariant features

In order to handle partial occlusion problem, authors in (Drira et al. 2013) detected the external object parts by comparing the given scan with a template scan. The template scan is developed using an average of training scans that are complete, frontal and have neutral expressions. Next, the basic matching procedure between a template and a given scan is car-

ried out using ICP algorithm. The broken curves, caused by the removal of occluding object are completed with the help of PCA based statistical model. This model is used to complete the incomplete curves using training data.

Bellil et al. (2016) proposed a method based on Gappy Wavelet Neural Network. Occluded regions are then refined by removing wavelet coefficient above a certain threshold. Alyuz et al. (2013) addressed the problem of external occlusions. They proposed a registration framework in which a possible non-occluded model is adaptively selected for each probe face, by employing non-occluded facial parts. Figure 2.12 presents highly occluded and correctly classified areas from Bosphorus dataset. For the detection of distinctive facial features, such as eyes and mouth, they employed the relative geometry information of these features.

Smeets et al. (2013) proposed the local feature based MeshSIFT algorithm to deal with missing 3D data for 3D face recognition. They first detected salient points on 3D facial surface as mean curvature extrema in scale space. Next, orientations are assigned to each of these salient points as presented in Figure 2.13. The neighborhood of each salient point is then described in a feature vector consisting of concatenated histograms of shape index and slant angles. Finally, the feature vectors of two 3D facial surfaces are reliably matched by comparing the angles in feature space.

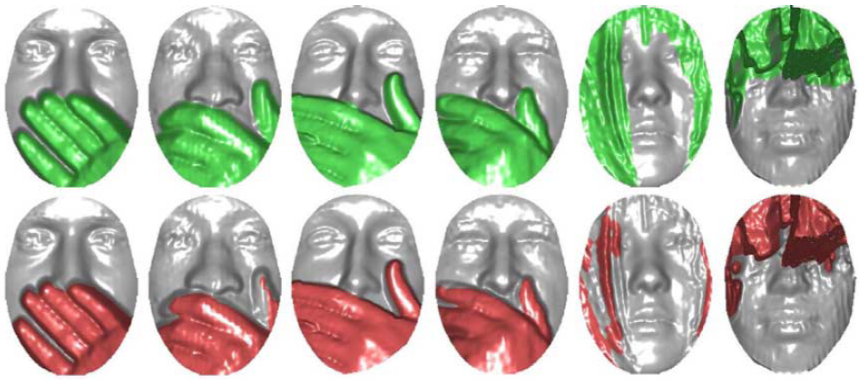


FIGURE 2.12 – *Highly occluded sample images that are correctly classified by the proposed masked Fisherfaces approach from the Bosphorus dataset. Top and bottom rows show the corresponding manually labeled (in green) and automatically detected (in red) occlusion masks. Alyuz et al. (2013)*

Berretti et al. (2013) handled 3D face recognition issue when just parts of probe scans are available. They used Scale Invariant Feature Transform (SIFT) to locate keypoints on depth image along with facial curves that connected those key-points.

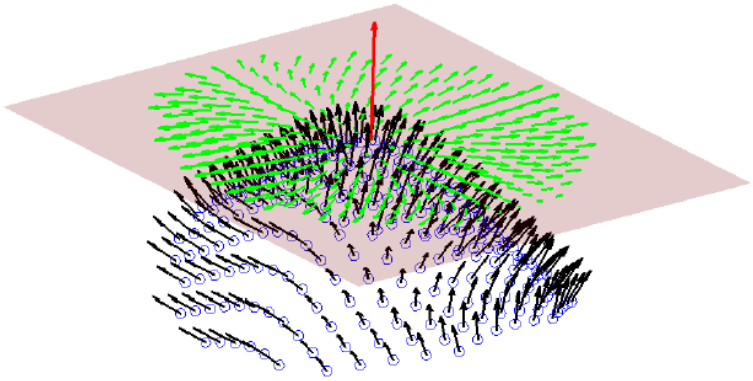


FIGURE 2.13 – *The neighbourhood of a scale space extremum with normals and their projection onto the tangent plane. Smeets et al. (2013)*

To handle the problem of pose variation, various methods applied a registration step to accurately compares probe and gallery faces. Ratyal et al. (2015) applied registration to correct the pose of 3D faces using vertical symmetry plane and horizontal nose plane.

Other methods applied landmarking on the face surface to detect the pose of the face, Perakis et al. (2009) proposed a method that treats the partial matching problem using a 3D landmark detector to detect the pose of the facial scan. This information is used to mark areas of missing data and to roughly register the facial scan with an annotated face model which exploits the facial symmetry where data are missing. Authors in (Passalis et al. 2011) have also used facial symmetry to handle the problem of missing data. Whereas, automatic landmarking has been applied to estimate the pose and to detect occluded areas.

Model-based methods have been applied to estimate the pose variation of 3D face as introduced in (Lee et Ranganath 2003). The proposed method consists of three parts ; an edge model, a color region model, and a wireframe model. The first two submodels are used for image analysis and the third mainly for face synthesis. In order to match the model to face images

in arbitrary poses, the 3D model can be projected onto different 2D view planes based on rotation, translation and scale parameters. Therefore, the pose of an unknown probe face is estimated by model matching, and the system synthesizes face images of known subjects in the same pose.

2.2.3 Similarity comparison

In order to measure the dissimilarity between 3D faces, various methods apply matching between their probe and gallery features using different metrics. Authors in (Zhang et al. 2014) proposed to use multiple keypoint descriptors (MKD) and the sparse representation-based classification (SRC). Each 3D face scan is represented as a set of descriptor vectors extracted from keypoints by meshSIFT. Descriptor vectors of gallery samples form the gallery dictionary. Given a probe 3D face scan, its descriptors are extracted at first and then its identity can be determined by using a multitask SRC. The proposed approach does not require a prealignment between two face scans and is quite robust to the problems of missing data, occlusions and expressions.

In order to reduce large storage space and expensive computational cost in developing 3D face matching, Yu et al. (2016) proposed a 3D directional vertices approach to represent and match 3D face surfaces by much fewer sparsely distributed vertices. To do so, authors extracted ridge and valley curves on a 3D surface along which the surface bends sharply. The recognition accuracy of the proposed method gives higher recognition performance compared to benchmark method presented in Mahoor et Abdel-Mottaleb (2009).

Other methods extract landmarks from the face surface, which are less sensitive to expression variation. To compute the similarity between faces, they apply matching between the extracted landmarks. For instance, Salazar et al. (2014) proposed an approach which learns the locations of a set of landmarks present in a database to automatically predict the locations of these landmarks on a newly available scan. The predicted landmarks are

then used to compute point-to-point correspondences between a template model and the newly available scan. Elaiwat et al. (2014) have applied Curvelet transform to detect salient points on the face that can capture invariant local features around the detected keypoints. Mian et al. (2008) present a feature-based algorithm for the recognition of textured 3D faces. They proposed a keypoint detection technique to detect where the shape variation is high in 3D faces. Next features from a probe and gallery face are projected to the PCA subspace and matched. The set of matching features are used to construct two graphs. The similarity between two faces is measured as the similarity between their graphs. Guo et al. (2016) applied The Nearest Neighbor Distance Ratio (NNDR) approach to perform feature matching as presented in Figure 2.14.

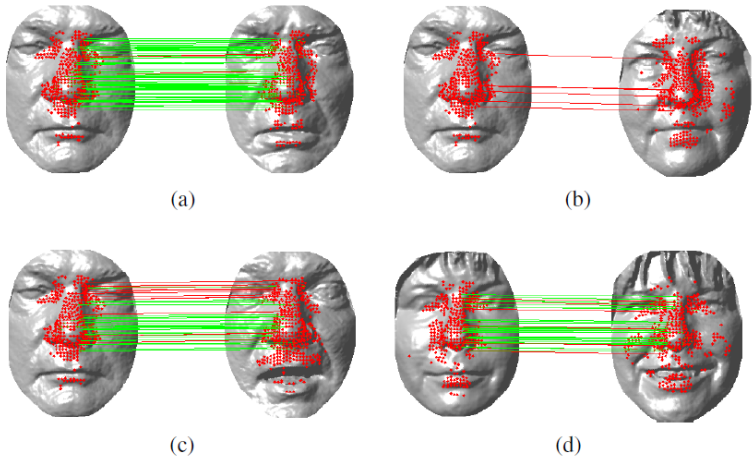


FIGURE 2.14 – *Feature matching results. (a) Faces of an individual with a neutral expression. (feature matches : 85, false matches : 4) ; (b) Faces of two individuals with a neutral expression. (feature matches : 6, false matches : 6) ; (c) Faces of an individual with different expressions. (feature matches : 42, false matches : 13) ; (d) Faces of an individual with different expressions and hair occlusions.(Guo et al. 2016)*

Registration based approaches have been also used to align probe and gallery faces. The most used is ICP which computes the distance between the aligned face surfaces and used it as a match score. The two faces with a lower distance are the more likely to be the same person. Mohammadzade et Hatzinakos (2013) used the iterative closest normal point method for finding the corresponding points between a gallery face and input probe faces. Firstly, they sampled a set of points for each face to find the clo-

sest normal points. These corresponding points are denoted as the closest normal points (CNPs). Then, a discriminant analysis method is applied to the normal vectors at the CNPs of each face for recognition. From this method, authors proved that the normal vectors contain more discriminatory information than the coordinates of the points of a face. Irfanoglu et al. (2004) have applied a dense correspondence between faces using ICP-based approach. To do so, they have aligned faces using dense point to point matching method by means of a mesh containing points that are present in all faces. The distance between two different point clouds is computed using point set distance as an approximation of the volume between facial surfaces.

Although ICP is a powerful estimation tool of the similarity between two faces, it has a serious drawback. ICP-based methods treat the 3D shape of the face as a rigid object so they are not able to handle changes in expression (Abate et al. 2007).

2.2.4 Discussion

Although local features have proved a good accuracy for face description, they have several limitations. For instance, 3D faces often exhibit large inter-class and intra-class variability that cannot be captured with a single feature type. This triggers the need for combining different modalities or feature types. However, different shape features often have different dimensions, scales and variation range, which makes their aggregation difficult without normalizing or using blending weights.

The main challenge is to build a 3D face recognition system robust against the several variations such as expression, pose, illumination, occlusion and other disruptions. This allows maximizing inter-class variations and minimizing intra-class variations.

2.3 3D facial expression recognition- State of the art

Similar to face recognition problem, facial expression recognition with the presence of different intra-class variations (i.e. pose, illumination, image quality, etc) as well as inter-class variations, has become a very challenging issue. To overcome this problem, different approaches have been proposed in the literature, most of these approaches focus on recognizing six basic expressions include anger (AN), fear (FE), disgust (DI), sadness (SA), happiness (HA) and surprise (SU) (Ekman et Friesen 1971). The expressions are textually defined by Pandzic et Forchheimer (2002) as shown in Table 2.2. Analyzing the expression of a human face requires a number of steps, the main two steps are :

- **Facial feature extraction and selection :** discriminative features are extracted and used to describe the facial expression. These features should be robust against the different variations such as illumination and pose. This step can be followed by a feature selection phase in order to choose relevant features to construct the model. The selected features will then feed a classifier in the next step.
- **Facial expression recognition :** the most used techniques are machine learning classifiers in order to accurately distinguish between the expressions.

Figure 2.15 presents the general FER system.

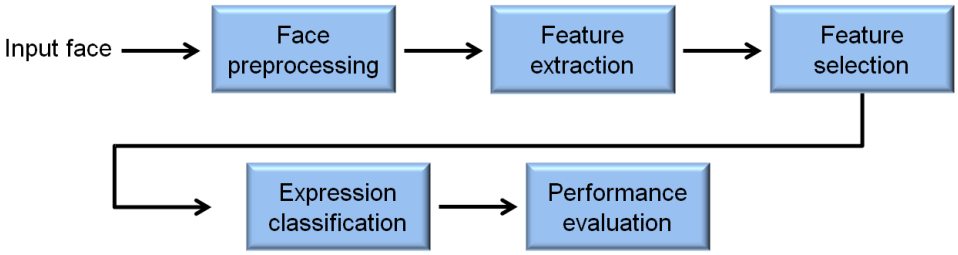


Figure 2.15 – *General FER system.*

2.3.1 Expression classification challenge

The recognition of facial expressions has attracted a great amount of researchers in the past decade. Detailed surveys of previous work can be found in (Fang et al. 2011; 2012, Zeng et al. 2009, Sandbach et al. 2012, Shan et al. 2009, Tian et al. 2003). Most of these previous works were developed for 2D data (Fasel et Luettin 2003, Pantic et Rothkrantz 2000, Zeng et al. 2009, Ilbeygi et Shah-Hosseini 2012, Mahersia et Hamrouni 2015, Chakrabarty et al. 2013). Although the remarkable performance achieved, most of these works are still sensitive to many variations, particularly illumination and pose. Recent progress in 3D acquisition techniques has provided a new alternative to overcome these issues (Yin et al. 2006). 3D data bring additional information which are more robust to illumination (Al-Osaimi et al. 2012, Patil et al. 2015) and pose changes (Ocegueda et al. 2013). State-of-the-art 3D FER methods are often based on a single descriptor which may fail to handle the large inter-class and intra-class variability of the human facial expressions.

Expression	**Textual Description**
Neutral	All face muscles are relaxed. Eyelids are tangent to the iris. The mouth is closed and lips are in contact.
Anger	The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are pressed against each other or opened to expose the teeth.
Sadness	The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is relaxed.
Surprise	The eyebrows are raised. The upper eyelids are wide open, he lower relaxed. The jaw is opened.
Hapiness	The eyebrows are relaxed. The mouth is open and the mouth corners pulled back toward the ears.
Disgust	The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often asymmetrically.
Fear	The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are tense and alert.

TABLE 2.2 – *Basic Facial Expressions (Pandzic et Forchheimer 2002).*

2.3.2 3D FER methods

Various methods have been proposed to analyze 3D FER to distinguish between the six prototypical expressions and AUs. We can classify them according to the used features into four categories as follows :

Distance-based features

Distance-based-features is one of the most known features used for 3D FER. The idea is to compute firstly the distance between certain facial landmarks from a neutral face. Next, after changing in the facial expression (i.e. deformation), the new distances between the aforementioned landmarks can be considered as features. The well known 3D dataset (i.e. BU-3DFE) provides 83 facial points (landmarks) located manually. These landmarks are widely used to compute this kind of distance features (Soyel et Demirel 2007; 2008b, Li et al. 2010, Tang et Huang 2008, Soyel et Demirel 2009; 2010, Tekguc et al. 2009, Sha et al. 2011, Srivastava et Roy 2009). For instance, Soyel et Demirel (2008b; 2010) used distance vectors computed between landmarks on the 3D face to describe facial features as presented in Figure 2.16(a). Probabilistic neural network is applied for expression classification. Sha et al. (2011) have extracted features by calculating the distances among all pairs of available facial landmarks as presented in Figure 2.16(b). Next, they classified each landmark into eight categories. The face has been divided into triangles using a subset of the given landmarks, and histograms have been formed for each triangle of the surface curvature types. Tang et Huang (2008) proposed an automatic feature selection method based on maximizing the average entropy. Next, they computed Euclidean distances between 83 facial features to a complete pool of candidate features composed of normalized points in the 3D space. Expression classification is then performed using a regularized multi-class AdaBoost classification algorithm. Soyel et Demirel (2007) uses six characteristic distances that are extracted from the distribution of eleven facial feature points from the given points in the BU-3DFE. This serves as in-

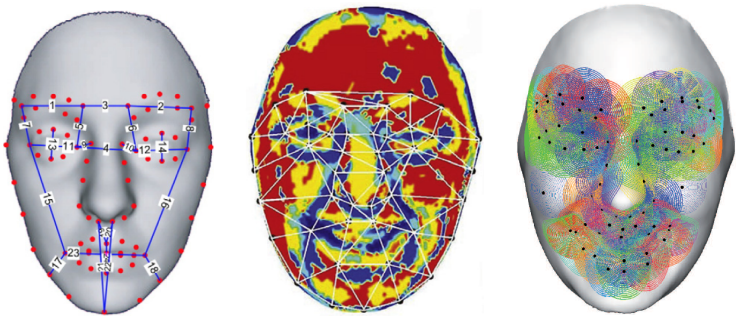
put to neural network classifier used for recognizing the different facial expressions. In Srivastava et Roy (2009), the authors computed the magnitude and the direction of the displacement of the given points in the BU-3DFE dataset instead of absolute distances. A feature matrix was then formed by concatenating the different matrices in each of the three spatial directions in order to form one 2D-matrix.

Patch-based features

The second category of 3D FER methods extract features on patches. it has also been widely used in expression recognition systems. They are used to capture information about the shape of the face over a small local region around either every point in the mesh, or around landmarks or feature points. Wang et al. (2006) computed a set of parameters for a smooth polynomial patch fitted to the local surface at each point in the mesh, which were subsequently used as inputs to rules that allowed the labeling of the surface at each point with primitives defining the type of curvature feature.

Maalej et al. (2011) proposed to represent each facial scan by a number of patches centered on considered points to describe the change in facial expression as presented in Figure 2.16(c). A Riemannian framework was then applied to compute the geodesic path between corresponding patches. Authors based on the assumption that people smile, or convey any other expression, the same way, or more appropriately certain regions taking part in a specific expression undergo practically the same dynamical deformation process. The association of those regions of two different expressions will deform differently. The geodesic distances between patches were labeled with respect to the six prototypical expressions and used as samples to train and test Multiboost algorithm classifier.

Lemaire et al. (2011) extracted patches around landmarks in the face through fitting of the Statistical Facial Feature Model, which is expressed as linear combinations of components of three different variations : shape,



(a) (b) (c)

FIGURE 2.16 – *Features based on the 83 facial points in BU-3DFE dataset. (a) : Distance between particular given facial points used in (Tang et Huang 2008, Soyel et Demirel 2008b; 2010).(b) : Distance and curvature features used in (Sha et al. 2011). (c) : 3D closed curves extracted around the landmarks used in (Maalej et al. 2011).*

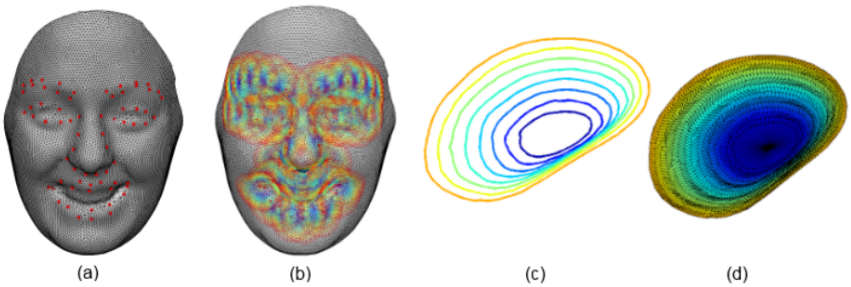


FIGURE 2.17 – *(a) 3D annotated facial shape model (68 landmarks) ; (b) closed curves extracted around the landmarks ; (c) example of 8 level curves ; (d) the mesh patch (Derkach et Sukno 2017).*

intensity and range value. These patches have then been compared to reference models representing the six prototypical expressions using ICP.

In (Derkach et Sukno 2017), authors proposed spectral methods as local shape descriptors. To do so, they proposed the use of Graph Laplacian features which result from the projection of local surface patches into a common basis obtained from the Graph Laplacian eigenspace as presented in Figure2.17.

Morphable models

The morphable models vary their shape in accordance with an unknown facial shape. They are also known as deformable models. Another methods used morphable models to extract facial expression features have been proposed. Ramanathan et al. (2006) proposed a Morphable Expression Model to model different expressions for a subject using his 3D face

surface. To do so, authors identified the corresponding points between expressive faces by reducing the energy function between points. These morphing parameters are used for emotion recognition and classification.

Mpiperis et al. (2008) proposed an elastically deformable model for establishing point correspondences among faces. This correspondence exploits both surface-to-model and model-to-surface distances during the model deformation as presented in Figure 2.18.

Rudovic et al. (2013) proposed a method for head-pose invariant facial expression recognition using a Coupled Scaled Gaussian Process Regression (CSGPR) model for head pose normalization. Next, they learned independently the mappings between the facial points in each pair of (discrete) non-frontal poses, and the frontal pose. Finally, they performed their coupling in order to capture dependencies between them.

A combination between 2D and 3D features is applied in (Huynh et al. 2016). To do so, authors proposed a convolutional neural network for 2D+3D feature-based FER. The proposed network consists of two CNNs, frontal view texture and 3D shape model. The network consists of three convolutional layers including max pooling as well as normalization layers, and two fully connected layers.

Furthermore, there are also a few FER systems that can process 3D dynamic sequences (i.e. 3D videos or 4D data). Shao et al. (2015) proposed an algorithm to videos retrieved by widespread and standard low-resolution RGB-D sensors, such as Kinect. After preprocessing, both RGB and depth image sequences, sparse features are learned from spatio-temporal local cuboids. Conditional Random Fields classifier is then employed for training and classification.

2D based features

Other methods map the 3D data into a 2D representation either to be able to directly apply 2D traditional techniques, or to reduce the high dimensionality of 3D faces. Authors in (Vretos et al. 2011) used depth images

FIGURE 2.18 – *Fitting base-mesh to a surface. a : base-mesh, b : original surface, c : base-mesh fitted to the surface (Mpiperis et al. 2008).*

of a 3D facial point cloud with Zernike moments. SVM are used in order to classify the six prototypical expressions. In order to tackle the problem of high modality using 3D faces, authors in (Azazi et al. 2015) transformed the 3D faces into the 2D plane using conformal mapping. Second, differential evolution based optimization algorithm is proposed to define the minimum and most relevant facial features for expression classification. The optimal features are selected from a pool of Speed Up Robust Features (SURF) descriptors of all the prospective facial points. Finally SVM is applied for the classification. Rosato et al. (2008) applied registration of vertex correspondence to convert the 3D meshes to 2D planar meshes. This mapping simplify the problem scope and allows for faster, more lightweight computations than the iterative-based approaches.

Lemaire et al. (2013) proposed Differential Mean Curvature Maps (DMCMs) to capture both global and local facial surface deformations caused by facial expressions. These DMCMs are directly extracted from a set of 2D maps by calculating the mean curvatures. Histograms of Oriented Gradients are applied to regions of DMCMs and to extract facial features. Multiclass-SVM classification algorithm is then performed to classify the six prototypical expressions. A few works have shown that salient keypoints and local descriptors can be effectively used to describe 3D facial expression. Berretti et al. (2010b) computed SIFT descriptors on a set of facial landmarks of depth images, and then selected the subset of most re-

levant features to characterize different facial expressions. SVM is applied for the classification.

Table 2.3 summarizes the existing approaches in 3D FER.

2.3.3 Discussion

3D face expressions often exhibit large inter-class and intra-class variabilities which require a robust representation to accurately distinguish between them. The state-of-the-art methods presented above use mostly local features in order to feed a classifier to recognize the six prototypical expressions. Methods used distance-based features generally use the 83 manually located points (landmarks) when dealing with BU-3DFE dataset, otherwise, manual annotation is required to precisely position landmarks. Although the good performance achieved by these methods, they are not full automatic since they require additional information about facial landmarks. Moreover, when dealing with automatic landmarking, distance-based features heavily rely on the accuracy of the landmark detection which may not be sufficiently discriminative. Methods using morphable-based features on the other hand are sensitive to inter-class and intra-class variability which decrease the recognition performance. 2D mapping allows applying 2D traditional methods for 3D FER problem. However, this mapping may lose some geometric characteristics of 3D facial expression and makes the FER task more subtle. Consequently, the challenge is to describe 3D facial expressions using robust features which must capture as accurately as possible facial surface deformations to enable the facial expression analysis.

Method	Morphable model (MM), Patch-based (P), Distance-based (D), 2D-based (2D), 2D+3D features	Landmarks	Expression	Dataset	Classifier	Average recognition rate (%)
Soyel et Demirel (2010)	D : Fisher criterion based feature selection	83 manual	7	BU-3DFE	Neural Network	93.23
Mpiperis et al. (2008)	MM : Bilinear models	Global registration	6	BU-3DFE	ML	90.51
Berretti et al. (2010b)	2D : SIFT	27 manual	6	BU-3DFE	SVM	77.53
Huynh et al. (2016)	MM : CNN	automatic	6	BU-3DFE	CNN	92.73
Azazi et al. (2015)	2D : Speed Up Robust Features descriptors	20 using SURF	6	BU-3DFE	SVM	81.81
Chun et al. (2013)	2D+3D : Curvature LBP	manual	6	Bosphorus	SVM+NN	76.98
Wang et al. (2013)	2D+3D : Curvature based descriptors	automatic	6	Bosphorus	SVM	76.56
Vretos et al. (2011)	2D : Zernike moments	automatic	6	Bosphorus	SVM	60.53
Tang et Huang (2008)	D : Normalized Euclidean distances	83 manual	6	BU-3DFE	Multi-class Adaboost	95.1
Lemaire et al. (2013)	2D : Differential Mean Curvature Maps + HOG	automatic	6	BU-3DFE	SVM	76.6
Maalej et al. (2011)	P : Geodesic distances between patches	automatic	6	BU-3DFE	SVM	89.81

TABLE 2.3 – *3D FER state-of-the-art methods.*

2.4 Conclusion

In this chapter, we first presented the 3D face acquisition techniques, the different 3D facial surface representations, and some available 3D face datasets. Second, we presented the different challenges of 3D face recognition which destruct the recognition performance so called *variations* (e.g. expression, pose, partial occlusion, etc). To alleviate these issues, various methods have been proposed, we reviewed the state-of-the-art methods proposed to handle these variations. Next, we introduced the limitation of the existing approaches for 3D face recognition.

We also presented the challenge of 3D facial expression recognition and we review the most interesting state-of-the-art methods proposed to tackle this problem. In this survey, we made the choice to categorize existing approaches according to the used features into four categories : distance based, patch based, morphable models and 2D based features. Finally, we discussed the limitations of the the existing methods to deal with the large inter-class and intra-class variabilities of the facial expressions.

In the next chapters, we present our proposed method to handle these issues and to efficiently combine heterogeneous features to construct a robust 3D face/facial expression recognition system.

3 3D Face Recognition Using Covariance Based Descriptors

Sommaire

In this chapter, we propose a new 3D face recognition method based on covariance descriptors. Unlike feature-based vectors, covariance-based

descriptors enable the fusion and the encoding of different types of features and modalities into a compact representation. The covariance descriptors are symmetric positive definite matrices which can be viewed as an inner product on the tangent space of (Sym_d^+) the manifold of Symmetric Positive Definite (SPD) matrices. In this chapter, we study geodesic distances on the Sym_d^+ manifold and use them as metrics for 3D face matching and recognition. We evaluate the performance of the proposed method on three well-known datasets including the FRGCv2, the GAVAB and the BU-3DFE datasets and demonstrate its superiority compared to other state-of-the-art methods in both identification and verification scenarios.

The reminder of this chapter is organized as follows. First, we present in Section 3.1 the covariance descriptors for 3D face recognition as well as the space of SPD matrices. In Section 3.2, the distance metrics used to compare covariance matrices as dissimilarity measure between 3D faces are reviewed. In Section 3.3, we present the two matching strategies applied in our proposed method to compare faces. Experimental results and comparative evaluation obtained on three well-known datasets are reported and discussed in Section 3.4. Hierarchical covariance description is presented in Section 3.5. Finally, conclusions are given in Section 3.6.

3.1 Covariance descriptors for 3D face recognition

Recently the image analysis community has shown a growing interest in characterizing image patches with the covariance matrix of local descriptors rather than the descriptors themselves. Covariance methods have been successfully used for object detection and tracking (Tuzel et al. 2008), texture (Tuzel et al. 2006) and image classification (Wang et al. 2012). Motivated by their success in image analysis, we propose a 3D face recognition method based on covariance descriptors as an extension of covariance based descriptors presented in (Tabia et al. 2014) for 3D shape retrieval. This chapter explores the usage of covariance matrices of features as discriminant representation for 3D face recognition problems. Our idea is to represent a 3D face with a set of m landmarks selected from its surface. Each landmark has a region of influence, which we characterize by the covariance of its geometric features instead of directly using the features themselves. These features, each of which captures some properties of the local geometry, can be of different type, dimension or scale. Covariance matrices provide a mean for their aggregation into a compact representation, which is then used for computing distances between 3D faces.

Covariance features extraction : Given a probe face (face to be recognized) F_1 and a gallery face (face in the database) F_0, we uniformly sample m feature points $\{p_1, \ldots, p_m\}$ from the gallery F_0. The m feature points of F_0 are the center of m patches of radius r, and form a paving of the face. We then align F_1 and F_0 by a coarse and fine registration using the Iterative Closest Point (ICP) (Besl et McKay 1992, Chen et Medioni 1991, Zhang 1994). After that, we select, from F_1, $N \leq m$ feature points $\{q_1, \ldots, q_N\}$, which are closest enough to the m points of F_0. In order to do so, we define a distance threshold $\delta = 0.1r$, and for each point p_i, we select its closest point q_i in the probe F_1. The point q_i is considered as a probe feature point only if the Euclidean distance $\|p_i - q_i\| < \delta$. The selected feature points q_j are the centers of the N patches in the probe face, and are used to com-

pute the similarity between F_0 and F_1. Figure 3.1 presents gallery feature points, and their respective extracted probe features points (in green). The probe is a left profile face after alignment which contains an occluded part. Red feature points on the occluded part are then ignored in the covariance description.

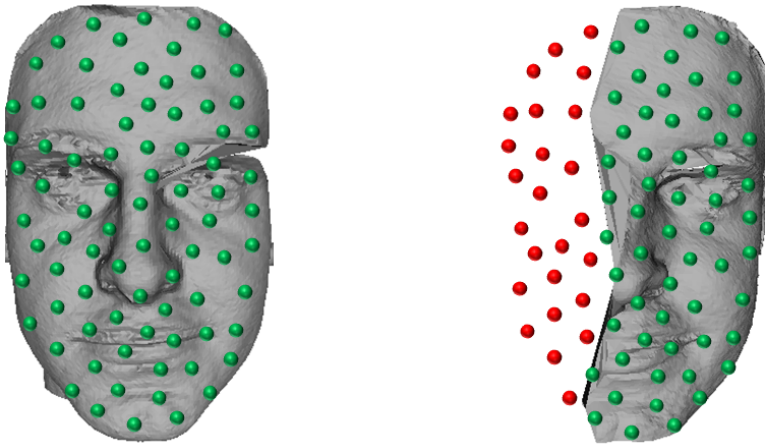


FIGURE 3.1 – *Feature points extracted from probe face under pose variation.*

Around each feature point, we extract a set of patches $\mathcal{P} = \{P_i, i = 1 \dots N\}$ from a 3D face. Each patch P_i defines a region around a feature point $p_i = (x_i, y_i, z_i)^t$. For each point p_j in P_i, we compute a feature vector f_j, of dimension d, which encodes the local geometric and spatial properties of the point. In our implementation, we considered the following feature vector :

$$f_j = \left[x_j, y_j, z_j, k_1, k_2, D_j\right] \tag{3.1}$$

Where :

— x_j, y_j and z_j are the three-dimensional coordinates of the point p_j.

— k_1 and k_2 are respectively the min and max principal curvatures.

— D_j is the distance of p_j from the origin defined by $\sqrt{x_j^2 + y_j^2 + z_j^2}$.

We characterize each face patch by the covariance matrix X_i :

$$X_i = \frac{1}{n}\sum_{j=1}^{n}(f_j - \mu)(f_j - \mu)^T \tag{3.2}$$

Where μ is the mean of the feature vectors $\{f_j\}_{j=1...n}$ computed in the patch P_i, and n is the number of points in P_i. The diagonal entries of X_i represent the variance of each feature and the non-diagonal entries represent their respective co-variations. Using covariance matrices as a region descriptors has several advantages, such as the ability of efficiently combining multiple features into a single descriptor and the invariance with respect to the ordering of points and number of feature vectors used for their computation. The size of covariance matrices does not depend on the size of the region from which they were extracted, but of the size of feature vectors, therefore, they can be computed from variable sized regions. Furthermore, covariance matrices are low dimensional compared to joint feature histograms.

An important aspect to consider is that building covariance-based descriptors requires local features that are correlated to each other otherwise covariance matrices become diagonal and will not provide additional benefits compared to using the individual features. Therefore, the parameters selected in the feature vector f_j need to be carefully selected, and could vary from one database to another. In Section 3.4, we have performed extensive performance simulations on two databases in order to select the best collection of parameters among the six which are defined in Equation (3.1).

Covariance matrices, however, lie on the manifold of Symmetric Positive Definite (SPD) tensors (Sym_d^+). Therefore, matching with covariance matrices requires the computation of geodesic distances on the manifold using proper metrics. Several geodesic distances on the Sym_d^+ manifold have been studied.

Once we have chosen the appropriate metric, the next step is to establish covariance matches between 3D faces and compute a global similarity measure. Two different strategies are proposed. The first strategy is to compute optimal match using a Hungarian solution for matching unordered set of covariance matrices. The total cost of matching is used as a

measure of dissimilarity between the pair of 3D faces. The second strategy is to compute a mean distance by integrating the chosen metric over the pairs of homologous regions, after spatial registration of the 3D faces. Figure 3.3 presents an overview of the proposed method.

The space of SPD matrices : Let $\mathcal{M} = Sym_d^+$ be the space of all $d \times d$ symmetric positive definite matrices and thus non-singular covariance matrices. Sym_d^+ is a non-linear Riemannian manifold, i.e. a differentiable manifold in which each tangent space T_X at X has an inner product $\langle \cdot, \cdot \rangle_{X \in \mathcal{M}}$ that smoothly varies from point to point. The inner product induces a norm for the tangent vectors $y \in T_X$ such that $\|y\|^2 = \langle y, y \rangle_X$. The shortest curve connecting two points X and Y on the manifold is called a geodesic. The length $d(X, Y)$ of the geodesic between X and Y is a proper metric that measures the dissimilarity between the covariance matrices X and Y. Let $y \in T_X$ and $X \in \mathcal{M}$. There exists a unique geodesic starting at X and shooting in the direction of the tangent vector y. The exponential map $\exp_X : T_X \mapsto \mathcal{M}$ maps elements y on the tangent space T_X to points Y on the manifold $\mathcal{M}$. The length of the geodesic connecting X to Y is given by $d(X, \exp_X(y)) = \|y\|_X$. Figure 3.2 depicts an example of two-dimensional manifold embedded in R^3.

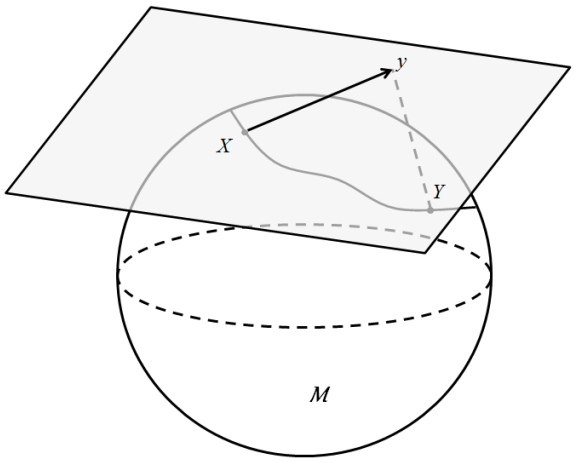


FIGURE 3.2 – *Two-dimensional manifold $\mathcal{M}$ embedded in R^3.*

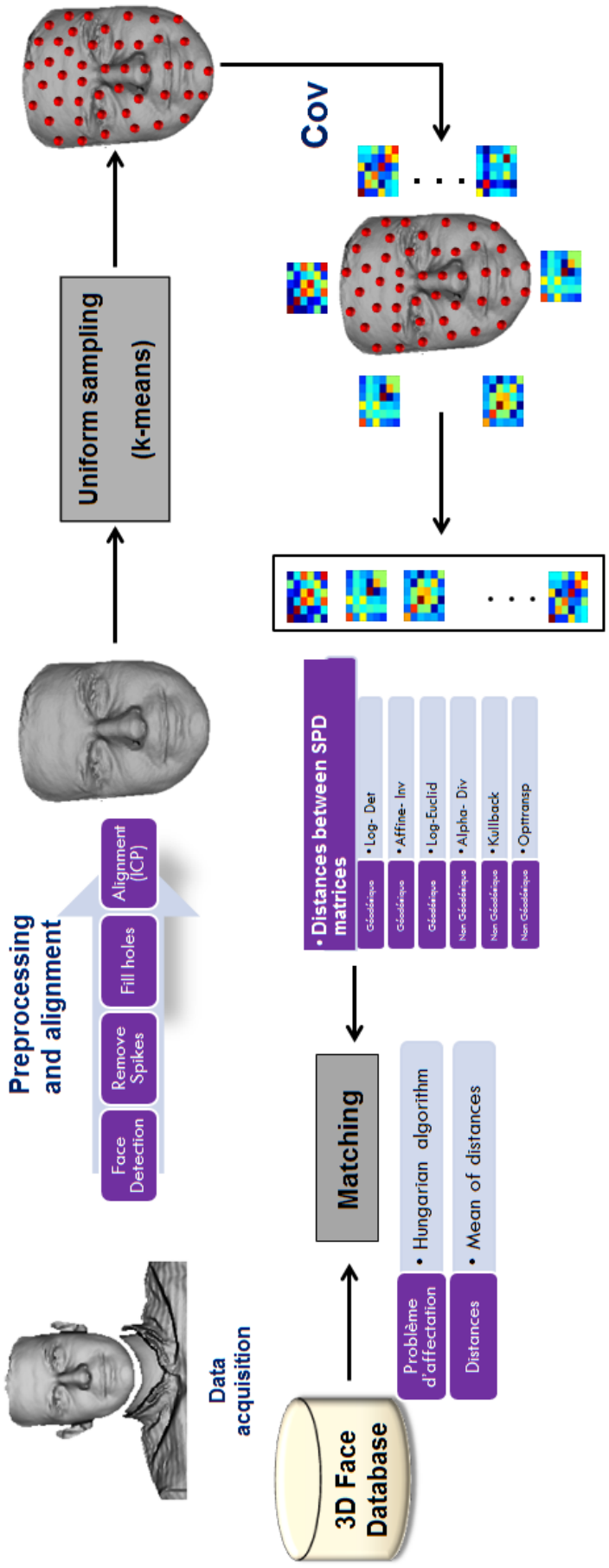


FIGURE 3.3 – *Overview of the proposed 3D face recognition method.*

3.2 Distances between covariance matrices

The space of covariance matrices $\mathcal{M} = Sym_d^+$ is a special type of homogeneous space which carries a natural Riemannian structure. More precisely, following the classification given in Moakher (2005), $\mathcal{M}$ is the Riemannian global symmetric space associated with the Lie algebra. Therefore, we can define a geodesic in the Riemannian space $\mathcal{M}$, or equivalently the space of Hermitian forms, as the shortest curve on $\mathcal{M}$, under the well chosen Riemannain metric or inner product, between two elements of the space $\mathcal{M}$. Since 3D face recognition task only requires a notion of distance between points on the manifold $\mathcal{M}$, we investigate in this chapter the computation using geodesic and non-geodesic distances as a comparative study.

3.2.1 Geodesic distances

In this section, we present the mathematical proprieties of three well-knwon geodesic distances which we used to compare covariance matrices.

The affine-invariant distance

The Riemannian metric of the tangent space T_X at a point X is given as $\langle y, z \rangle_X = \text{trace}\left(X^{-\frac{1}{2}} y X^{-1} z X^{-\frac{1}{2}}\right)$. The exponential map associated to the Riemannian metric $\exp_X(y) = X^{\frac{1}{2}} \exp\left(X^{-\frac{1}{2}} y X^{-\frac{1}{2}}\right) X^{\frac{1}{2}}$ is a global diffeomorphism (a one-to-one, onto, and continuously differentiable mapping in both directions). Thus, its inverse is uniquely defined at every point on the manifold : $\log_X(Y) = X^{\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) X^{\frac{1}{2}}$. The symbols exp and log are the ordinary matrix exponential and logarithm operators, while $\exp_X$ and $\log_X$ are manifold-specific operators, which depend on the point $X \in Sym_d^+$. The tangent space of Sym_d^+ is the space of $d \times d$ symmetric matrices and both the manifold and the tangent spaces are of dimension $m = d(d+1)/2$.

For symmetric matrices, the ordinary matrix exponential and loga-

rithm operators can be computed in the following way. Let $X = UDU^T$ be the eigenvalue decomposition of the symmetric matrix X. The exponential series is defined as : $\exp(X) = \sum_{k=0}^{\infty} \frac{X^k}{k!} = U \exp(D) U^T$, where $\exp(D)$ is the diagonal matrix of the eigenvalue exponentials. Similarly, the logarithm is given by $\log(X) = \sum_{k=1}^{\infty} \frac{-1^{k-1}}{k} (X - I)^k = U \log(D) U^T$. The exponential operator is always defined, whereas the logarithms only exist for symmetric matrices with strictly positive eigenvalues. The geodesic distance between two points on Sym_d^+ is then given by :

$$\begin{aligned} d^2_{ainv}(X, Y) &= \langle \log_X(Y), \log_X(Y) \rangle_X \\ &= \text{trace}\left(\log^2\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)\right) \end{aligned} \tag{3.3}$$

Log Determinant distance

From equation 3.3, it is apparent that computing the geodesic distance can be unattractive as it requires eigenvalue computations or sometimes even matrix logarithms, which for larger matrices causes significant slowdowns. For an application that must repeatedly compute distances between numerous pairs of matrices this computational burden can be excessive Cherian et al. (2011). Driven by such computational concerns, Cherian et al. (2011),Chebb et Moakher (2012),Sra (2012) introduced a symmetrized log-determinant based matrix divergence.

The greatest advantage of this metric against the affine-invariant metric is its computational speed, it requires only computation of determinants, which can be done rapidly via 3 Cholesky factorizations for ($X + Y$, X and Y), each at a cost of $\frac{1}{3}d^3$ flops (Golub et Van Loan 2012). Computing the affine-invariant on the other hand requires generalized eigenvalues, which can be done for positive-definite matrices in approximately $4d^3$ flops.

Let $X, Y \in Sym_d^+$ of $d \times d$ symmetric positive definite matrices which have positive eigenvalues. The log-determinant distance between X and Y is defined by :

$$d_{ld}(X,Y) = \sqrt{\log\left(\det\left(\frac{X+Y}{2}\right)\right) - \frac{1}{2}\log(\det(X.Y))} \tag{3.4}$$

It is easy to see that d_{ld} is symmetric, non-negative, and definite. Moreover, it is invariant under congruence transformations, ($d_{ld}(AXA^T, AYA^T) = d_{ld}(X,Y)$ for invertible A), and under inversion ($d_{ld}(X,Y) = d_{ld}(X^{-1}, Y^{-1})$.

Log Euclidean distance

The log-Euclidean framework Arsigny et al. (2006) proposed by Arsigny et. al. defines a class of Riemannian metrics called log-Euclidean metrics. The geodesic distances associated with log-Euclidean metrics are called log-Euclidean distances. Let $\odot$ be an operation on SPD matrices defined as $X \odot Y = \exp(\log(X) + \log(Y))$. Any inner product $\langle , \rangle$ defined on $T_I S_n^{++} = \{\log(X) | X \in S_n^{++}\} = S_n$ extended to the Lie group (S_n^{++} , $\odot$) by left or right multiplication is a bi-invariant Riemannian metric. The corresponding geodesic distance between $X \in S_n^{++}$ and $Y \in S_n^{++}$ is given by : $d(X,Y) = \|mlog(X) - mlog(Y)\|_I = \|\log(X) - \log(Y)\|_I$ where $\| \ \|_I$ is the norm induced by $\langle , \rangle$. Note that here $mlog_I$ is the inverse-exponential map at the identity matrix which is equal to the usual matrix logarithm in this case.

To compare SPD matrices there are many other distances that can be used.

3.2.2 Other distances

In this section we present other distances which have been proposed to compare covariance matrices.

Alpha divergence distance

Geometry and various divergence functions mostly related to Alpha-divergence through a unified approach based on convex functions, One of

important features of the considered family of divergences is that they can give some guidance for the selection and even development of new divergence measures if necessary. Moreover, these families of divergences are generally defined on unnormalized finite measures (not necessary normalized probabilities). This allows us to analyze patterns of different size to be weighted differently, e.g. images with different sizes or documents of different length. Such measures play also an important role in the areas of neural computation, pattern recognition, learning, estimation, inference, and optimization Cichocki et Amari (2010). The α-divergence between SPD matrices is defined by :

$$d_\alpha(P,Q) = \frac{4}{(1-\alpha^2)} \log \frac{\det(\frac{1-\alpha}{2}P + \frac{1+\alpha}{2}Q)}{\det(P)^{\frac{(1-\alpha)}{2}} \det(Q)^{\frac{(1+\alpha)}{2}}} \tag{3.5}$$

where P and Q are two unnormalized distributions and $\alpha \in (-\infty, +\infty)$ α-divergence is zero if $p = q$ and positive otherwise, As α approaches 0, α-divergence specializes to Kullback-Leibler (KL)-divergence from q to p.

Kullback distance

The Kullback-Leibler distance Kullback (1997) is perhaps the most frequently used information-theoretic distance measure from a viewpoint of theory, it is a special case of α-divergence where α approaches zero :

$$\lim_{\alpha \to 0} [q \| p] = KL[q \| p] \tag{3.6}$$

Let P and Q be probability measures on a set X with densities p and q with respect to a dominating measure λ . The relative entropy of P with respect to Q is defined as :

$$D_{KL}(P \| Q) = \lambda \left(p \log \frac{p}{q} \right) \tag{3.7}$$

Relative entropy is also known as Kullback-Leibler divergence, information gain, information divergence, and the Kullback-Leibler Information

Criterion (KLIC). In our experiments kullback distance between SPD matrices becomes :

$$KL = \sqrt{\frac{1}{2} trace(P/Q + P/Q - 2eye(size(P)))} \tag{3.8}$$

Optimal transportation distance

Optimal transportation distances Villani (2008) are a fundamental family of parameterized distances for histograms. Despite their appealing theoretical properties, excellent performance in retrieval tasks and intuitive formulation, Optimal transportation distances and their application to computer vision hold a special place among other distances in the probability simplex. Given a $d \times d$ cost matrix M, the cost of mapping r to c using a transportation matrix (or joint probability) P can be quantified as : $\langle P, M \rangle$ The following problem :

$$d_M(r, c) =^{def} \min_{P \in (r,c)} \langle P, M \rangle \tag{3.9}$$

is called an optimal transportation problem between r and c given cost M. The optimum of this problem, $d_M(r, c)$, is a distance Villani (2008) whenever the matrix M is itself a metric matrix. In Our case the optimal transportation distance between SPD matrices A and B is defined by :

$$d_{ot}(A, B) = \sqrt{trace(A) + trace(B) - 2trace((A^{\frac{1}{2}} B A^{\frac{1}{2}})^{\frac{1}{2}})} \tag{3.10}$$

Table 3.1 presents a comparison of computational complexities between affine-invariant, log-Euclidean, kullback metrics on covariance matrices and d_{ld} metric. This comparison further proves the computational speed of d_{ld} metric comparing with its counterparts defined above.

Metric	$D^2(X,Y)$	**Flops**	**Gradient (∇_X)**
Affine-invariant	$\text{trace}\left(\log^2\left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)\right)$	$4d^3$	$2X^{-1}log(XY^{-1})$
Log-Euclidean	$\|\log(X)-\log(Y)\|_I$	$\frac{8}{3}d^3$	$2X^{-1}(logX-logY)$
KL	$\sqrt{\frac{1}{2}trace(P/Q+P/Q-2eye(size(P)))}$	$\frac{8}{3}d^3$	$Y^{-1}-X^{-1}YX^{-1}$
d_{ld}	$\sqrt{\log\left(\det\left(\frac{X+Y}{2}\right)\right)-\frac{1}{2}\log(\det(X.Y))}$	d^3	$(X+Y)^{-1}-\frac{1}{2}X^{-1}$

TABLE 3.1 – *Comparison of computational complexities of d_{ld} metric and other metrics between SPD matrices. Cherian et al. (2013)*

3.3 3D FACE MATCHING USING SPD MATRICES

Similar to local features, covariance matrices computed on 3D surfaces can be used as local descriptors for matching two faces. Let us consider a patch center p_i, $i = 1,...,m$ represented by a covariance matrix X_i in a gallery 3D face F_0 and a patch center q_j, $j = 1...,N$ represented by the covariance matrix Y_j in a probe 3D face F_1. Let $c_{ij} = c\left(p_i, q_j\right)$ denotes the cost of matching these two points. This cost is defined as the distance between the two covariance matrices X_i and Y_j.

Given the set of costs c_{ij} between all pairs of points p_i on the gallery face F_0 and q_j on the probe face F_1, we define the total cost of matching the two 3D faces using two different ways :

Optimal match The total cost of matching is defined by :

$$Cost_1 = \sum_{i=1}^{m} c\left(p_i, q_{\varphi(i)}\right), \tag{3.11}$$

Minimizing $Cost_1$, subject to the constraint that the matching is one-to-one, gives the best permutation $\varphi(i)$. This is an assignment problem, which can be solved using the Hungarian algorithm (Kuhn 1955, Munkres 1957). The input to the assignment problem is a cost matrix with entries c_{ij}. The result is a permutation $\varphi(i)$ such that Equation (3.11) is minimized. Finally, once the permutation φ is computed, we use the total cost of matching, defined by Equation (3.11), as a measure of dissimilarity between the pair of 3D models.

Mean of distances An alternative matching cost, simpler than optimal matching consists in computing the mean of distances between each pair of homologous regions. Figure 3.4 presents homologous covariance matrices extracted form a probe and a gallery face. So Equation (3.11) becomes :

$$Cost_2 = \frac{1}{N}\sum_{j=1}^{N} c\left(p_j, q_j\right), \tag{3.12}$$

where N is the number of homologous patches.

The two matching strategies are presented in Figure 3.5.

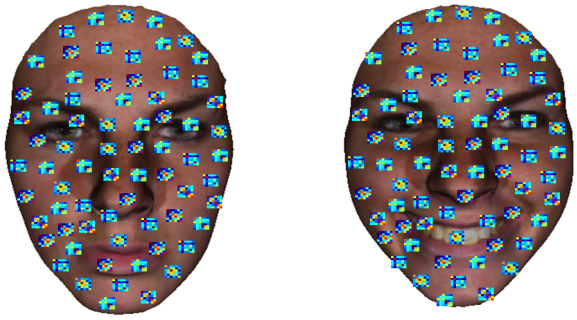


FIGURE 3.4 – *Covariance matrices extracted from probe and gallery faces.*

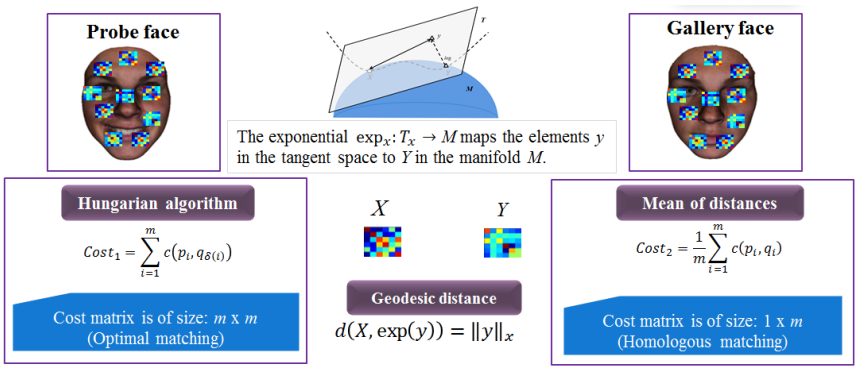


FIGURE 3.5 – *The two matching strategies between covariance matrices.*

3.4 EXPERIMENTAL RESULTS

We present results from different experiments in which we evaluate the performance of the proposed covariance descriptors. The performance is measured according to the percentage of the correctly recognized faces. We have also studied the impact of the chosen distance and the matching procedure on the recognition performance.

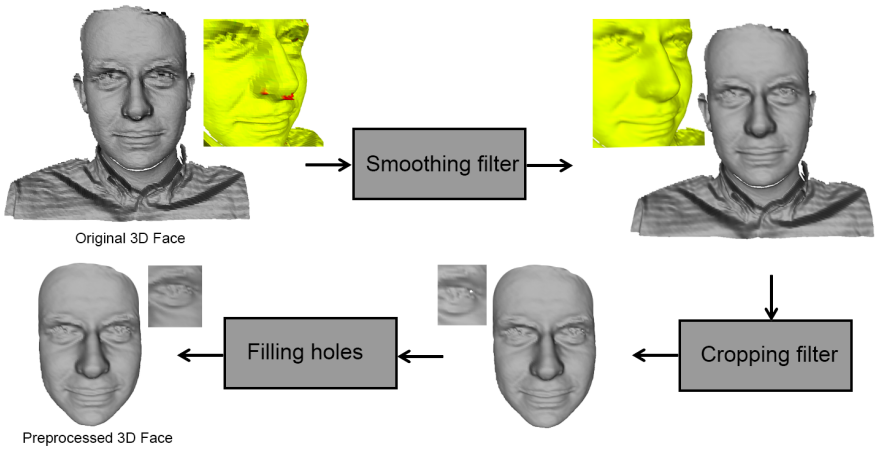


FIGURE 3.6 – *Automatic 3D face preprocessing.*

3.4.1 Preprocessing and alignment

After the acquisition step, the input face surface is preprocessed. The preprocessing helps improving the quality of the input face which may contain some imperfections (e.g. holes, spikes) as well as some undesired parts (e.g. clothes, neck, ears, hair, etc.) and so on. It consists of applying successively a set of filters (Figure 3.6). First, a smoothing filter is applied, which reduces spikes in the mesh surface, followed by a cropping filter which cuts and returns parts of the mesh inside an Euclidean sphere. Finally a filling holes filter is applied, which identifies and fills holes in input meshes. Note that spikes mainly occur in three regions : the eyes, the nose tip and the teeth. To remove these spikes, we apply a median filter on 3D face vertices. The filter starts by sorting the z coordinate within a neighborhood, finding then the median, and finally replacing the original z coordinate with the value of the median.

After preprocessing, we align each probe face to the gallery face using Iterative closest point algorithm (ICP). The aim of ICP based alignment approach is to determine relation and translation parameters iteratively in order to transform one point cloud in the gallery face such that it lies as close as possible to other point cloud on the probe face.

In order to extract features points, we apply a uniform clustering using k-means algorithm as follows : Let $\{X_i\}$, $i = 1, ..., n$ be the set

of n dimensional facial vertices to be clustered into a set of k clusters, $C = \{c_k, k = 1, ..., K\}$. k-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Let μ_k be the mean of cluster c_k. The squared error between μ_k and the points in cluster c_k is defined as :

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

The goal of k-means is to minimize the sum of the squared error over all K clusters :

$$J(C) \sum_{k=1}^{K} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2$$

The main steps of k-means algorithm are summarized as follows (Dubes et Jain 1988) :

1. Select an initial partition with K clusters ; repeat steps 2 and 3 until cluster membership stabilizes.
2. Generate a new partition by assigning each pattern to its closest cluster center.
3. Compute new cluster centers.

The obtained K centers are then used as facial feature points in our proposed method.

3.4.2 Experiments on the FRGCv2 dataset

We have first preprocessed the 3D surfaces and selected $m = 40$ feature points on each 3D face in the gallery as described in Section 3.4.1. We have then extracted one patch P_i around each point p_i. Each patch has a radius $r = 15\%$ of the radius of the shape's bounding sphere. For each patch, we compute a 5×5 covariance matrices computed from the feature vector $[x, y, z, k_1, k_2]$ (details about the impact of the features, the size of the patch radius r and the number of patches m are given in Section 3.4.4).

Identification scenario In this section, we evaluate the performance of the proposed method in the identification scenario on FRGCv2 dataset. Table 3.2 presents a comparison of recognition performance on the FRGCv2 database using the different proposed distances (Section 3.2) with respect to the two proposed matching methods (Section 3.3). In this experiment, we evaluate "Neutral versus All" identification experiment, where the first 3D face scan with neutral expression from each subject is used as gallery and the remaining face scans are treated as probes. When using the Hungarian algorithm, the highest recognition rate is achieved by the log-Euclidean distance 97.9%, followed by the log-determinant distance which achieves slightly lower rate 96.0%. When using the mean of distances algorithm, the log-determinant distance achieves the highest recognition rate 99.2%. The affine-invariant distance performs 99.1%.

From this experiment, one can notice that when using the geodesic distances (i.e. log-determinant, affine-invariant and log-Euclidean distances), both Hungarian and mean of distances matching techniques behave better than using non-geodesic distances. This demonstrates that the geodesic distances are more discriminative for covariance matrices than the other distances. This is the behavior that one would expect since the non-geodesic distances void one of the benefits of considering the Riemannian structure of the (Sym_d^+) manifold. On the other hand, Table 3.2 also shows that using the mean of distances matching technique is more suitable for 3D face recognition. This result shows that the spatial relations between covariance matrices are also an important component in the matching process.

Table 3.3 presents the recognition performance using "Neutral versus non-Neutral" protocol. In this experiment, the best recognition performance is achieved by log-determinant distance when dealing with the two matching algorithms, followed by affine-invariant and log-Euclidean distances. From this comparison, we can also notice that geodesic distances give higher recognition performance comparing to non-geodesic

Distance	Hungarian algorithm	Mean of distances
Log-determinant	96.0%	**99.2%**
Affine-invariant	90.0%	99.1%
Log-Euclidean	**97.9%**	98.7%
Alpha divergence	91.0%	98.9%
KL divergence	70.4%	92.9%
Optimal transportation	64.1%	78.5%

TABLE 3.2 – *Recognition rates on FRGCv2 dataset using the different distance metrics presented in Section 3.2. Reported results are obtained using Neutral vs All protocol.*

Distance	Hungarian algorithm	Mean of distances
Log-determinant	**97.4%**	**97.6%**
Affine-invariant	97.2%	97.2%
Log-Euclidean	96.7%	96.9%
Alpha divergence	88.0%	90.1%
KL divergence	68.4%	72.9%
Optimal transportation	62.1%	63.6%

TABLE 3.3 – *Recognition rates on FRGCv2 dataset using the different distance metrics presented in Section 3.2. Reported results are obtained using Neutral vs Non-Neutral protocol.*

distances. This proves our claim about the robustness of geodesic distances as a dissimilarity metric for 3D face recognition.

Table 3.4 presents a comparison of our method to several state-of-the-art methods using the two protocols, i.e. Neutral versus All, Neutral versus non-Neutral. From this table, we can see that our method outperforms the other state-of-the-art methods. This performance can be explained by the fact that covariance matrices provide an elegant way for combining multiple heterogeneous features without normalization or joint probability estimation. This combination significantly boosts the performance of our approach.

Method	Neutral vs. All	Neutral vs. non-Neutral
Queirolo et al. (2010)	98.4%	-
Spreeuwers (2011)	99.0%	-
Wang et al. (2010)	98.3%	-
Drira et al. (2013)	97.0%	-
Mian et al. (2008)	-	92.1%
Huang et al. (2012)	97.6%	95.1%
Faltemier et al. (2008)	98.1%	95.0%
Alyuz et al. (2010)	97.5%	96.4%
Ratyal et al. (2015)	98.9%	-
Al-Osaimi et al. (2008)	93.7%	-
Our method	**99.2%**	**97.4%**

TABLE 3.4 – *Comparison with state-of-the-art methods on the FRGCv2 dataset.*

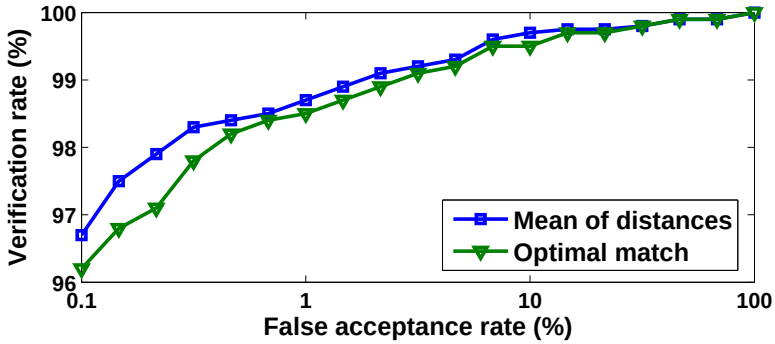
Verification scenario We further evaluate the proposed method in the verification (authentication) scenario on FRGCv2 dataset. To do so, we plot the Receiver Operating Characteristic (ROC) curves for the "All versus All" experiment as shown in Figure3.7. The horizontal axis of the ROC curve is the False Accept Rate (FAR), while the vertical axis is the Verification Rate (VR) also called true acceptance rate (TAR). They are defined over the square similarity matrix with a dimensionality of 4007×4007.

When dealing with log-determinant distance, our method provides 96.7% VR at 0.1% FAR using the mean of distances matching method and 96.2% using the optimal match one as shown in Figure3.7(a). With the Affine invariant distance, our method gives slightly lower VR comparing with the its performance using log-determinant distance as shown in Figure3.7(b).

Table3.5 presents a comparison of verification rates at FAR=0.1% on the FRGCv2 dataset with state-of-the-art results. From this comparison we can see that our method is slightly lower but still close to the best of ones in the literature in the verification scenario.

Note that FRGCv2 contains mostly frontal scans with high quality, so the missing data issues are not treated in this dataset, therefore, many existing methods achieved good performance. In order to evaluate the efficiency of our method against other variations such as pose changes, we have evaluated it on the GAVAB dataset as presented in the next Section.

(a) Log-determinant distance



(b) Affine-Invariant distance

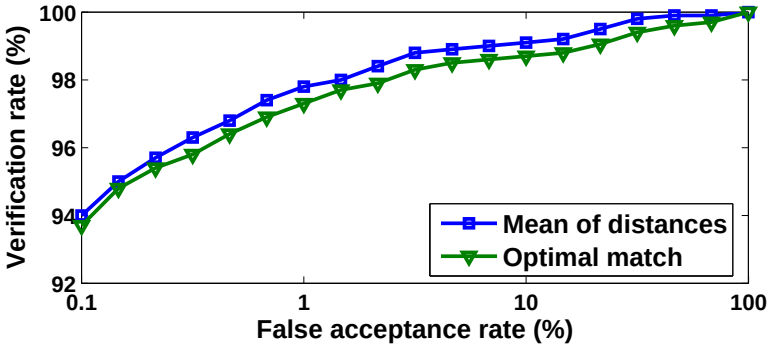


FIGURE 3.7 – *ROC curves for all versus All verification experiment.*

3.4.3 Experiments on GAVAB dataset

For the experiment on GAVAB dataset, we use $m = 50$ feature points and 6×6 covariance matrices computed from the feature vector $[x, y, z, k_1, k_2, D]$.

Table 3.6 presents the recognition performance of the proposed method using the different distances presented in Section 3.2. In this experiment, the first frontal facial scan of each subject was used as gallery while the others were treated as probes. The reported results are obtained using the optimal match when dealing on expressive faces. The best recognition rate is obtained by geodesic distances (i.e. log-determinant, affine invariant and log-Euclidean distances respectively).

Table 3.7 presents the recognition performance using the different distances presented in Section 3.2. The mean of distances matching method shows that the log-determinant and the affine-invariant distances give the

Method	All vs. All
Queirolo et al. (2010)	96.5%
Spreeuwers (2011)	94.6%
Wang et al. (2010)	**98.1%**
Drira et al. (2013)	93.9%
Huang et al. (2012)	94.2%
Faltemier et al. (2008)	93.2%
Our method	96.7%

TABLE 3.5 – *Comparison of verification rates at FAR=0.1% on the FRGCv2 dataset with state-of-the-art results.*

Distance	Neutral	Neutral+Expressive	Expressive	Looking down	Looking up
Log-determinant	100%	97.54%	97.26%	96.72%	95.08%
Affine-invariant	100%	97.95%	97.81%	96.72%	95.90%
Alpha Divergence	95.08%	92.21%	91.80%	90.98%	90.16%
KL divergence	91.80%	90.16%	89.61%	88.52%	85.24%
Optimal transportation	91.39%	87.97%	89.61%	86.06%	83.60%
Log-Euclidean	100%	96.72%	95.08%	91.80%	93.44%

TABLE 3.6 – *Recognition rates on the GAVAB dataset using the optimal match method.*

highest recognition rates, followed by the alpha divergence, KL divergence, optimal transportation, and the log-Euclidean respectively. From this comparison, we can see that the log-determinant distance gives the highest recognition rate with expressive scans, whereas the affine-invariant distance performs quite well with pose scans. This behavior also demonstrates that the geodesic distances are more efficient for covariance matrices than the other distances. Note that it is also possible to combine the results from each distance, i.e. using vote or training method to further improve the recognition rate. Comparing with Table 3.6, we can clearly see that the second matching method (i.e. Mean of distances) is more suitable comparing to the optimal match method. This performance is obtained due to the homologous matching that is applied only between complete regions, whereas occluded regions are excluded from the matching process (see Figure3.1).

Table 3.8 compares the results of our method to results from state-of-the-art methods following the same protocol. We calculated rank-one face recognition rates which show the matching accuracies for different categories of probe faces : including the results with and without expression and

Distance	Neutral	Neutral+Expressive	Expressive	Looking down	Looking up	Right profile	Left profile
Log-determinant	**100%**	**100%**	**100%**	97.54%	94.26%	75.80%	81.96%
Affine-invariant	100%	99.59%	99.45%	**99.18%**	**98.36%**	78.69%	**83.60%**
Alpha Divergence	100%	99.18%	98.90%	95.08%	93.44%	**81.96%**	80.32%
KL divergence	98.36%	97.95%	96.72%	95.08%	91.80%	-	-
Optimal transportation	97.54%	97.13%	95.62%	93.44%	91.80%	-	-
Log-Euclidean	97.54%	96.72%	95.08%	92.34%	90.16%	-	-

TABLE 3.7 – *Recognition rates on the GAVAB dataset using mean of distances matching method.*

pose variations. The highest recognition rate achieved by each method is highlighted.

From this comparison, we can see that our method outperforms the majority of the other state-of-the-art approaches in terms of the recognition rate. From Table 3.8, we can see that for frontal neutral probes, our method provides high recognition rate (100%) similarly as in (Drira et al. 2013, Huang et al. 2012, Tabia et al. 2014), note that this rate is obtained by the log-determinant, the affine-invariant and the alpha divergence distances. For expressive faces, our method with the log-determinant distance provides the highest recognition rate with non-neutral expressions faces (100%) and its performance surpasses all the other methods. The results on (Neutral+Expressive) faces also demonstrate that the proposed method efficiently outperforms the other methods, since we achieve an accuracy of (100%). With looking down faces, our method provides a good recognition rate (99.18%) which is better than the results given by Huang et al. (2012) and Mahoor et Abdel-Mottaleb (2009) and slightly lower than the result of Drira et al. (2013). Our method also gives the highest recognition rate (98.36%) on looking up faces similarly as in (Drira et al. 2013), and 97.81% with overall faces. Note that, the performance decreases on left or right sides scanned faces which include many occluded regions, but still outperforms state-of-the-art methods on right side scanned faces. The experimental results on the GAVAB dataset clearly demonstrate that the proposed method can deal with large pose changes and even partial occlusions.

3.4.4 Effects of the features, the patch size and the number of patches

In this section, we study the performance of the proposed method with respect to the main parameters of the recognition system. First, we studied the impact of the local features that are selected to form the feature vector f (see Equation (3.1)). In Table 3.9, for various choices of feature vectors, we present the performance results on the GAVAB and the FRGCv2 da-

Protocol	**Drira et al. (2013)**	**Li et al. (2009)**	**Tabia et al. (2014)**	**Mahoor et Abdel-Mottaleb (2009)**	**Huang et al. (2012)**	**Tang et al. (2017)**	**Our method**
Neutral	**100%**	96.67%	**100%**	95%	**100%**	**100%**	**100%**
Expressive	94.54%	93.33%	93.30%	72%	93.99%	79.2%	**100%**
Neutral+expressive	95.9%	94.68%	94.91%	78%	95.49%	-	**100%**
Rotated looking down	**100%**	-	-	85.3%	96.72%	98.4%	99.18%
Rotated looking up	**98.36%**	-	-	88.6%	96.72%	**100%**	98.36%
Overall	96.99%	-	-	-	-	-	**97.81%**
Right profile	70.49%	-	-	-	78.69%	-	**81.96%**
Left profile	86.89%	-	-	-	**93.44%**	-	83.60%

TABLE 3.8 – *Comparison with state-of-the-art methods on the GAVAB dataset.*

Features	GAVAB	FRGCv2
$f = [x, y, z]$	95.08%	92.7%
$f = [k_1, k_2]$	53.16%	79.0%
$f = [x, y, z, k_1]$	96.72%	95.6%
$f = [x, y, z, k_2]$	95.08%	93.4%
$f = [x, y, z, k_1, k_2]$	94.84%	**99.2%**
$f = [x, y, z, D]$	97.18%	92.8%
$f = [k_1, k_2, D]$	65.10%	78.6%
$f = [x, y, z, k_1, k_2, D]$	**97.81%**	98.5%

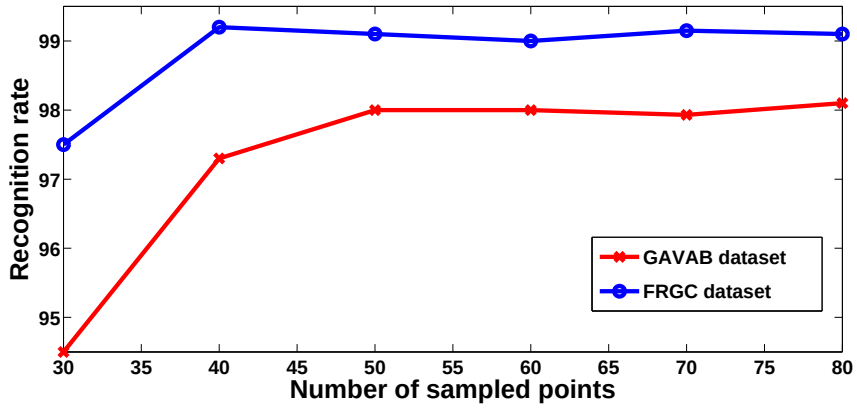
TABLE 3.9 – *Effects of the various geometric features on the performance of our face recognition method. Reported results are on both FRGCv2 and GAVAB datasets over all faces.*

tasets over all faces, using the best performing geodesic distance and the best matching technique, i.e. log-determinant distance and the mean of distances matching algorithm. We can clearly see that the performance of the covariance method highly depends on the chosen features. Although the combination of the six features performs the best in the GAVAB dataset, this experiment shows that the performance of our recognition system does not necessarily improves with the number of selected features. For instance, as shown in Table 3.9, co-varying $[x, y, z]$ features gives slightly better performance than co-varying the $[x, y, z, k_1, k_2]$ features. This behavior can be explained by the fact that some feature types are almost orthogonal (i.e. their correlation is low). Thus, their covariance matrix is almost diagonal and therefore not sufficiently discriminative.

We also analyze how the recognition performance of the proposed method varies with respect to the number of sample points. In this experiment, we set the patch radius $r = 15\%$ of the cropped face's bounding sphere and we vary the number of sample points between 30 to 80. We use the best performing distance and matching technique. Results are summarized in Figure 3.8(a). It shows that the performance over all faces becomes stable when the number of sample points is larger than 40 for the FRGCv2 and 50 for the GAVAB dataset. This is predictable since small number of points will result in a coarse representation of the 3D face.

We also analyzed how the recognition performance of the proposed method varies with respect to the patch radius r. For this end, we set the

(a) Varying the number of patches



(b) Varying patch radius

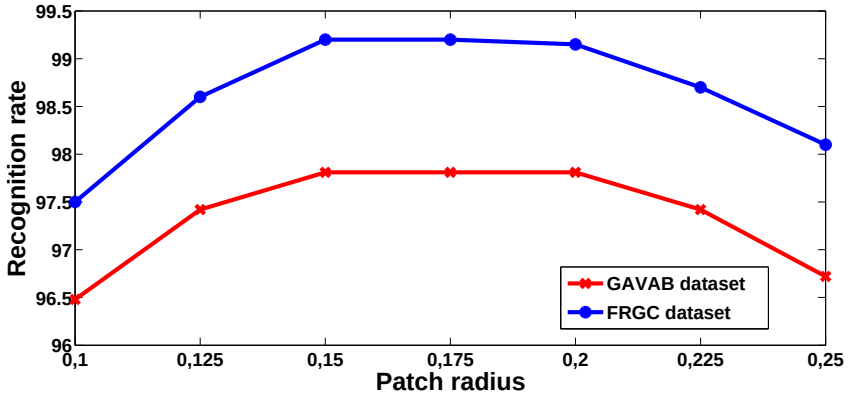


FIGURE 3.8 – *Effect of the patch radius and the number of patches on the recognition performance of the proposed covariance based method. The reported results are on both FRGCv2 and GAVAB datasets over all faces.*

number of sample points $m = 40$ for the FRGCv2 and $m = 50$ for the GAVAB dataset and vary the patch radius between 10% to 25% of the total radius of the cropped face's bounding sphere. Please note that in this setting the patches may overlap. Figure 3.8(b) shows that the performance remains stable when r varies between 15% and 20%. The performance starts to drop when choosing values outside this interval. Note that, similar to all local descriptor, this behavior was predictable since very small patches do not capture sufficient geometric properties of the shapes. Large

patches on the other hand capture only coarse features, which may not be sufficiently discriminative.

3.5 Hierarchical covariance description

In this section we present an extension of the covariance matching method proposed in our previous work presented above, in which we demonstrated the usefulness of covariance matrices as local descriptors for 3D face recognition. Here, we further focus on the issue of face recognition under facial expression variation. To do so, we propose to represent a 3D face using a set of feature points, around each of which we consider three description levels starting from a small region to a bigger overlapped region as presented in Figure 3.9. We use a covariance based descriptor to represent each region. The performance of the proposed method has been evaluated on the BU-3DFE, the GAVAB and the FRGCv2 datasets.

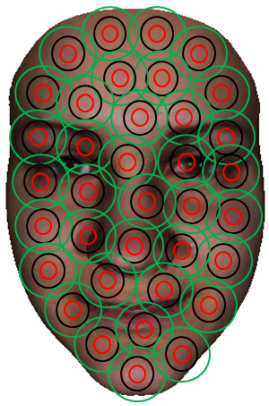


Figure 3.9 – *Hierarchical covariance extraction. Green circles refer to grand patches, black for average patches and red for small patches.*

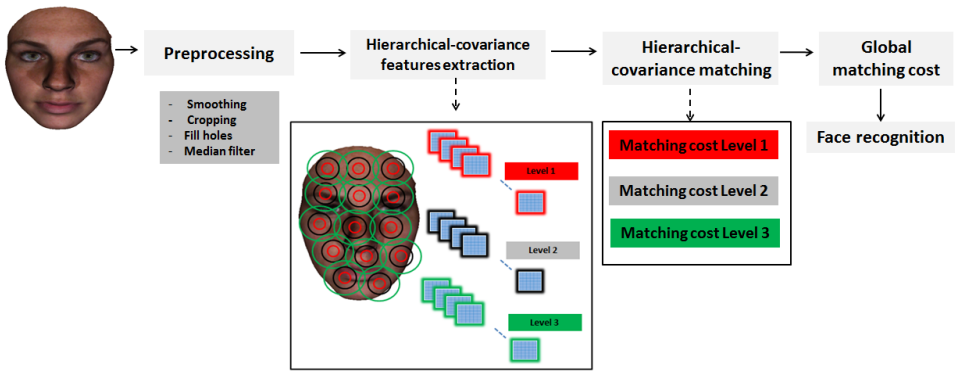


Figure 3.10 – *Overview of the proposed hierarchical covariance method.*

The advantage of covariance descriptors is that the size of covariance matrices does not depend on the size of the region from which they were extracted, but of the size of feature vectors, therefore, they can be com-

puted from variable sized regions (three different patch sizes) as shown in Figure 3.10. Covariance matrices however are not elements of an Euclidean space; they are elements of a Lie group, which has a Riemannian structure. Therefore, matching with covariance matrices requires the computation of geodesic distances on the manifold using a proper metric. In this contribution, we have applied log-determinant distance.

Formally, we uniformly sample $m = 30$ feature points and 6×6 covariance matrices computed from the feature vector : f_j=$[x, y, z, k1, k2, D]$. Next, around each feature point, we extract three covariance descriptors with respect to three patch radius ($r_1 = 10\% \times R$, $r_2 = 20\% \times R$, $r_3 = 30\% \times R$), where R is the radius of the cropped face's bounding sphere. To compare probe and gallery faces, we compute the mean of distances which measures their dissimilarity as follows :

Given the set of costs c_{ij} between all pairs of points p_i on the gallery face F_0 and q_j on the probe face F_1, we define the total cost of matching the two 3D faces by computing the mean of distances between each pair of homologous regions over the three levels as follows :

$$Cost = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} c\left(p_j, q_j\right) \right) \tag{3.13}$$

Where m is the number of sampled feature points, n is the number of levels.

Finally, the class of each probe face is the identity of the gallery face which minimizes the matching cost.

Results on GAVAB, BU-3DFE and FRGCv2 datasets To evaluate our method performance we present a Cumulative Match Characteristic curve (CMC) which plots the recognition rate versus the rank number. The rank-1 recognition rate is the percentage of all probes for which the best match in the gallery belongs to the same person, which is a popular evaluation criterion for face identification. The percentage of the best and the second-best correct matches is the rank-2 recognition rate and so on for higher

ranks. To obtain the recognition rate, we compute the ratio of the correctly classified query images to the total number of query images.

Figure 3.11 reports the CMC curves of the proposed method on BU-3DFE dataset using two protocols (i.e. Neutral versus Expressive and Low-intensity versus High-intensity). We can clearly see that the recognition performance with respect to the rank number increases faster using the first protocol. This behavior can be explained by the fact that gallery faces are neutral which is a helpful for face identification. In the other hand, using the second protocol (i.e. Low intensity versus High intensity), gallery faces are expressive and this is more binding for the matching task.

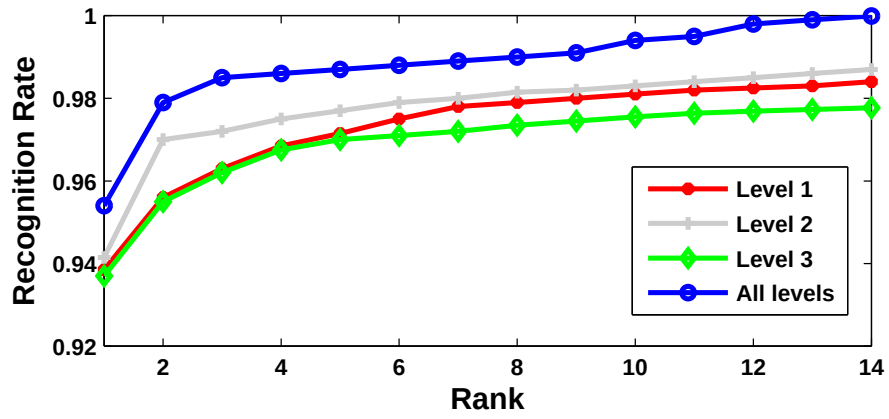
Table 3.10 shows the rank-1 recognition performance using different protocols on the three datasets. From these results, it appears that the matching of the hierarchical covariance levels combination gives higher recognition performance compared to the use of each level individually. This behavior can be explained by the fact that small patches do not capture sufficient geometric properties of the shapes. Large patches on the other hand capture only coarse features, which may not be sufficiently discriminative. The combination of the three patch levels captures both fine and coarse features and therefore provides a more accurate representation.

Dataset	Protocol	Level 1	Level 2	Level 3	All levels
GAVAB	Neutral-Vs-Expressive	98.00%	99.45%	98.90%	**100%**
FRGCv2	Neutral-Vs-Non-Neutral	97.1%	97.4%	96.9%	**97.6%**
BU-3DFE	Neutral-Vs-Expressive	93.85%	94.15%	93.70%	**95.40%**
BU-3DFE	Low int-Vs-High int	97.25%	97.60%	96.90%	**98.25%**

TABLE 3.10 – *Face recognition rates using hierarchical covariance method on the three datasets.*

In comparison to the state-of-the-art methods reported in Table3.4 and Table3.8, when dealing with the FRGCv2 dataset (Neutral versus non-Neutral protocol) and the GAVAB dataset (Neutral versus Expressive protocol), we can see that our hierarchical covariance method outperforms the state-of-the-art methods with 97.6% and 100% respectively.

(a) Neutral-Vs-Expressive protocol



(b) Low intensity-Vs-High intensity protocol

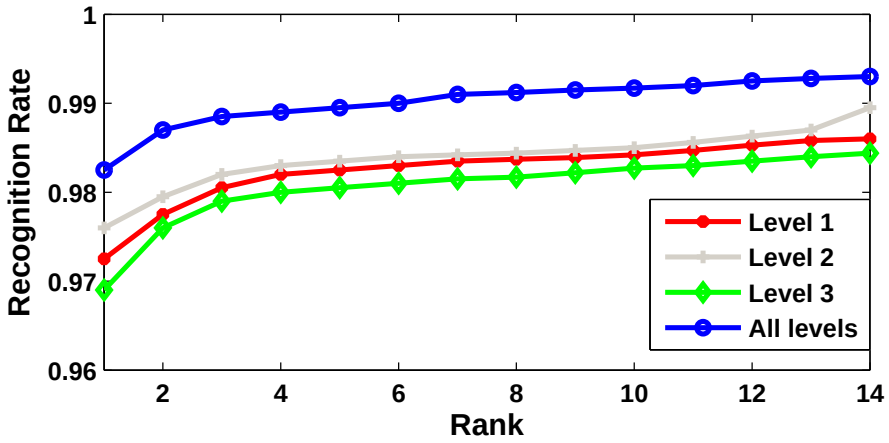


FIGURE 3.11 – *The CMC curves of our proposed method on BU-3DFE dataset. Reported results are obtained using the three hierarchical covariance levels individually and on their combination.*

When dealing with BU-3DFE dataset (see Table3.11), our proposed method achieves higher rank-1 recognition rate using "Low-intensity versus High-intensity" identification protocol compared to Lei et al. (2013)'s method (i.e. 98.25% vs 97.70%). When dealing with "Neutral versus All" protocol, our method outperforms Lei et al. (2016)'s method (i.e. 95.70% vs 93.25%). This performance is achieved due to the accurate facial description obtained by covariance descriptors and reinforced by the hierarchical representation.

Method	Neutral vs. All	Low intensity vs. High intensity
Lei et al. (2016)	93.25%	-
Lei et al. (2013)	-	97.70%
Our method	**95.70%**	**98.25%**

TABLE 3.11 – *Comparison with state-of-the-art methods on the BU-3DFE dataset.*

3.6 CONCLUSION

In this chapter, a new approach for comparing 3D faces using covariance matrices of features instead of the features themselves is proposed. We studied various distances for dissimilarity measure between two covariance matrices and proposed two different ways for 3D face matching. Covariance matrices provide an elegant way for combining multiple heterogeneous features without normalization or joint probability estimation. Therefore, analyzing 3D faces with covariance matrices has several advantages compared to individual descriptors. First, covariance matrices enable the fusion of multiple heterogeneous features of arbitrary dimension without normalization, blending weights, or joint probability distribution estimation. Also, spatial relationships can be naturally encoded in the covariance matrices. Moreover, covariance matrices are compact, compared to histogram-based representations, and can be efficiently computed. Finally, although we have experimented in our work with only three types of features, our approach is generic and thus various types of features can be added to the framework. An important aspect to consider is that building covariance-based descriptors requires local features that are correlated to each other otherwise covariance matrices become diagonal and will not provide additional benefits compared to using the individual features instead of their covariance.

We also proposed a hierarchical covariance description for 3D face matching and recognition, under expression variations. We represented a 3D face using a set feature points, around each of which we considered three description levels. The levels start from a small region to a bigger overlapped region. We used a covariance based descriptor to represent each

region. The log-determinant geodesic distance is used for the face matching. Experimental results on BU-3DFE, GAVAB and FRGCv2 datasets showed that the use of the three hierarchical levels improves the recognition performance compared to the use of each level individually. This performance can be explained by the fact that each hierarchical level captures some specific characteristics which are complementary.

4

3D Facial Expression Recognition Using Kernel Methods on Riemannian Manifold

Sommaire

In this chapter, we handle the problem of 3D facial expression recognition regardless to the face identity. We focus on the six prototypical expressions (i.e. Happiness, Angry, Disgust, Sadness, Surprise and fear).

The majority of work conducted in this area has been done using 2D data. Most of these systems are still highly sensitive to the different variations such as illumination, occlusions and other changes in facial appearance like makeup and facial hair.

Furthermore, 2D FER systems are very sensitive to pose variation, therefore it is necessary to maintain a consistent facial pose (preferably a frontal one) in order to achieve a good recognition performance.

Due to the development of 3D image capturing technologies, the acquisition of 3D data is becoming a more feasible task. The 3D data bring a more effective solution in addressing the issues faced by its 2D counterpart. State-of-the-art 3D FER methods are often based on a single descriptor which may fail to handle the large inter-class and intra-class variability of the human facial expressions.

In this chapter, we explore, for the first time, the usage of covariance matrices of descriptors instead of the descriptors themselves in 3D FER. Since covariance matrices are elements of the non-linear manifold of SPD matrices, we particularly look at the application of manifold-based classification to the problem of 3D FER. We have performed comprehensive experiments on two well-known datasets, and demonstrate the superiority of our proposed method compared to the state-of-the-art methods.

The rest of this chapter is organized as follows. In Section 4.1, covariance descriptors for 3D FER are addressed. We explain in Section 4.2 how to classify covariance matrices on manifold using conventional classification algorithm. In Section 4.3, the classification of 3D facial expressions using kernel-SVM on Riemannian manifold is addressed. Experimental results on BU-3DFE and Bosphorus datasets are reported and evaluated in Section 4.4. Conclusions end the chapter.

4.1 Covariance descriptors for 3D FER

In this section, we first discuss the advantages of covariance descriptors for 3D FER task comparing to the use of local features. Second, we present our proposed covariance description which we use for the classification of the six prototypical expressions.

4.1.1 Covariance descriptors versus local features for 3D FER

Proposed methods often use 3D local features which capture the geometrical and topological properties of the face surface to distinguish between expressions or Action Units Fang et al. (2011; 2012), Zeng et al. (2009), Sandbach et al. (2012), Shan et al. (2009), Tian et al. (2003). One of the main strengths of local features is their flexibility in terms of type of analysis that can be performed with. Wang et al. (2006) proposed to extract geometric based features to describe facial expressions. These features have been estimated using the principle curvature information calculated on the 3D triangulated mesh model of a face. A linear discriminant analysis classifier has been used for features classification. Soyel et Demirel (2008a; 2010) used distance vectors computed between landmarks on the 3D face to describe facial features, and used probabilistic neural network for expression classification. Shao et al. (2015) proposed to learn sparse features from spatio-temporal local cuboids extracted from the face. They applied conditional random fields classifier to train and classify the expressions.

The use of local features in 3D facial expression recognition, however, has several limitations. For instance, 3D face expressions often exhibit large inter-class and intra-class variability that cannot be captured with a single feature type. This triggers the need for combining different modalities or feature types. However, different shape features often have different dimensions, scales and variation range, which makes their aggregation difficult without normalizing or using blending weights.

Covariance matrices successfully have been used as region descriptors (Krizaj et al. 2013, Tabia et al. 2014, Tabia et Laga 2015). The use of covariance matrices has several advantages. First, they provide a natural way for fusing multi-modal features, eventually of different dimensions, without normalization or joint distribution estimation. Second, covariance matrices extracted from different regions have the same size, which is significantly compact compared to the features themselves and to their statistics. This enables comparing any regions without being restricted to a constant window size or specific feature dimension. Covariance matrices, however, lie on the manifold of Symmetric Positive Definite (SPD) tensors Sym_d^+, a special type of Riemannian manifolds. Therefore, matching with covariance matrices requires the computation of geodesic distances on the manifold using proper metrics. In the previous chapter, we have shown how such geodesic distances can be computed in an efficient way.

4.1.2 The proposed covariance description for 3D FER

Once the 3D face mesh has been preprocessed (see Section 3.4.1 in the previous Chapter), we uniformly select m feature points over the whole 3D surface. The feature points are the center of m patches from a paving of the face. Each point has a region of influence, which we characterize by the covariance of its geometric features instead of directly using the features themselves. Each feature captures some properties of the local geometry. They can be of different type, dimension or scale.

Let $\mathcal{P} = \{P_i, i = 1 \dots m\}$ be the set of patches extracted from a 3D face. Each patch P_i defines a region around a feature point $p_i = (x_i, y_i, z_i)^t$. For each point p_j in P_i, we compute a feature vector f_j, of dimension d, which encodes the local geometric and spatial properties of the point. In our implementation, we considered the following feature vector :

$f_j = \left[x_j, y_j, z_j, k_1, k_2, D_j\right]$,

where x_j, y_j and z_j are the three-dimensional coordinates of the point p_j. k_1 and k_2 are respectively the min and max principal curvatures. D_j is

the distance of p_j from the origin defined by $\sqrt{x_j^2 + y_j^2 + z_j^2}$. We characterize each face patch by the covariance matrix X_i :

$X_i = \frac{1}{n}\sum_{j=1}^{n}(f_j - \mu)(f_j - \mu)^T$,

where μ is the mean of the feature vectors $\{f_j\}_{j=1...n}$ computed in the patch P_i, and n is the number of points in P_i. The diagonal entries of X_i represent the variance of each feature and the non-diagonal entries represent their respective co-variations.

Covariance matrices lie on the (Sym_d^+) which lacks Euclidean structures such as norm and inner product. This makes impossible the application of conventional clustering algorithms in their original forms. In the next section, we review the existing strategies to classify manifold valued data.

4.2 Classification on Riemannian manifold

Support vector machine (SVM) classifier (Cortes et Vapnik 1995) is a supervised machine learning method which is popular for addressing binary classification problems.

Given a set of labeled feature vectors $\{x_i, y_i\}_{i=1}^{N}$ where $x_i \in R^d$ and $y_i \in \{-1, +1\}$, a SVM aims to find a classifier that has the minimum generalization error on the test set. This is related to finding maximum margin hyperplan. For non-linear separable classes, a mapping $(R^d \to \mathcal{H})$ is usually applied to map the feature vectors $x_i \in R^d$ to a higher dimensional space where classes may be more close to linearly separable. This produces a kernel Hilbert space $\mathcal{H}$ with an inner product (kernel function) $\mathcal{K}(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$. For extension of a binary SVM to a multiclass SVM, one-against-all or one-against-one strategies can be applied (Hsu et Lin 2002).

Since manifolds (i.e. Sym_d^+) lack a vector space structure and other Euclidean structures such as norm and inner product, popular techniques developed for Euclidean spaces do not apply such as machine learning al-

gorithms including support vector machines (SVM), principal component analysis (PCA) and clustering (Tabia et Laga 2015, Jayasumana et al. 2015).

To overcome this problem, one can neglect the non-linear geometry of manifold-valued data and apply Euclidean methods directly. As a result, this approach often yields poor accuracy and undesirable effects (Pennec et al. 2006, Arsigny et al. 2006). Recently, this problem has been addressed in two ways :

Approximation using tangent space : which can be achieved by flattening the manifold to approximate it onto tangent space. Tuzel et al. (2008) proposed to train several weak classifiers on the tangent spaces and combining them through boosting for pedestrian detection. Authors in (Wang et al. 2012) modeled a set of 2D images by a single covariance matrix. Next, they applied Log-Euclidean distance to map each covariance matrix from the Riemannian manifold to a Euclidean space for image set classification.

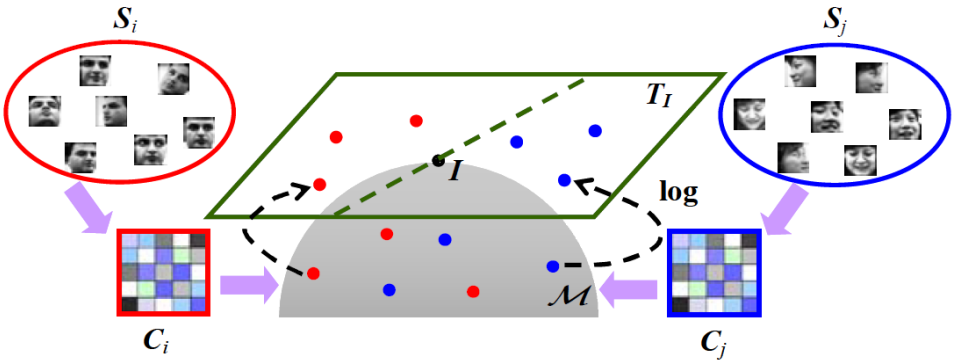


FIGURE 4.1 – *The problem of image set (i.e. S) classification is formulated as classifying its covariance matrix C on the Riemannian manifold $\mathcal{M}$. (Wang et al. 2012)*

Yun et al. (2013) applied a two-layer mapping for the manifold points, by first using the logarithmic mapping under the log-Euclidean metric, and then by using a Radial Basis Function. Similar strategy has been applied in (Yun et Gu 2016), in which authors encoded three layer levels of features using covariance descriptors to address the problem of classifying activities in video. SVM classifier under the one-against-all strategy by exploiting the Riemannian geometry is then applied for the classification.

Approximation using Reproducing kernel Hilbert space : exploiting a positive definite kernel function to embed the manifold into a RKHS. Au-

thors in (Jayasumana et al. 2015) used covariance matrices for pedestrian detection, visual object categorization, texture recognition and image segmentation. To be able to utilize algorithms developed for linear spaces on non-linear manifold, they applied Gaussian radial basis function (RBF)-based positive definite kernels on manifolds which embed the manifold with a corresponding metric in a high dimensional reproducing kernel Hilbert space. Since the Gaussian RBF defined with any given metric is not always positive definite, authors presented a unified framework for analyzing the positive definiteness of the Gaussian RBF on a generic metric space. They then used the proposed framework to identify positive definite kernels on two specific manifolds (i.e. Riemannian manifold of SPD matrices and the Grassmann manifold).

In (Hamm et Lee 2008), authors have treated each subspace as a point in the Grassmann space, and have performed feature extraction and classification in the same space. In the same way, Harandi et al. (2013) have proposed to model the actions by subspaces elements of a Grassmann manifold. Then, they embed this manifold into reproducing kernel of Hilbert spaces in order to tackle the problem of action classification on such manifolds. Different kernels to embed Grassmannian manifold into a Hilbert space and represent each entity (image set, video) using a single subspace have been introduced in Harandi et al. (2014b).

Vemulapalli et al. (2013) developed extrinsic classifiers for features that lie on Riemannian manifolds using the kernel learning approach. Based on the log-Euclidean framework, they have shown how geodesic distance functions can be learned for Sym_d^+ matrices by learning Mahalanobis distance functions in the logarithm domain. Figure4.2 depicts the difference between Vemulapalli et al. (2013)'s method and tangent space or poor choice of kernels. Deng et al. (2017) have applied local covariance descriptor and Riemannian kernel sparse coding. SPD matrices are mapped to the RKHS, and the log-Euclidean Gaussian kernel sparse coding method is applied to identify the faces.

In both cases, Euclidean based techniques can be applied to the embedded data, since Hilbert spaces obey Euclidean geometry. Recent studies, however, report superior results with RKHS embedding over flattening the manifold using its tangent spaces (Hamm et Lee 2008, Vemulapalli et al. 2013). Intuitively, this can be attributed to the fact that a tangent space is a first order approximation to the true geometry of the manifold, whereas, being higher-dimensional, an RKHS has the capacity of better capturing the non-linearity of the manifold (Harandi et al. 2014b).

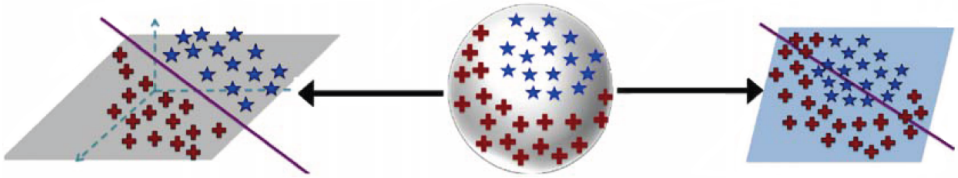


FIGURE 4.2 – *Tangent-space mapping or poorly-chosen kernel can often result in a bad classifier (right). Good classification using a learned mapping which uses the classifier cost in the optimization (left). (Vemulapalli et al. 2013)*

Based on the above discussion, we take advantages of recent works on kernel methods on manifold-valued data (Harandi et al. 2014a;b, Tabia et Laga 2015) and explore, for the first time, their usage in 3D FER. Since covariance matrices are considered in this work as local descriptors, we propose to apply the SVM algorithm to this local representation. For this end, we build a global kernel function so that one can compare two 3D facial expressions by using the covariance descriptors. The proposed 3D FER method has been evaluated on the two well known datasets, namely the BU-3DFE and the Bosphorus. Figure 4.3 presents an overview of the proposed method. In the next section, we give more details about the classification of 3D facial expressions on Riemannian manifold of SPD matrices.

4.3 3D FACIAL EXPRESSION RECOGNITION

Once covariance matrices have been extracted and the geodesic distance has been defined, the expression recognition task can be reduced to covariance classification. However, the non-linear structure of Sym_d^+

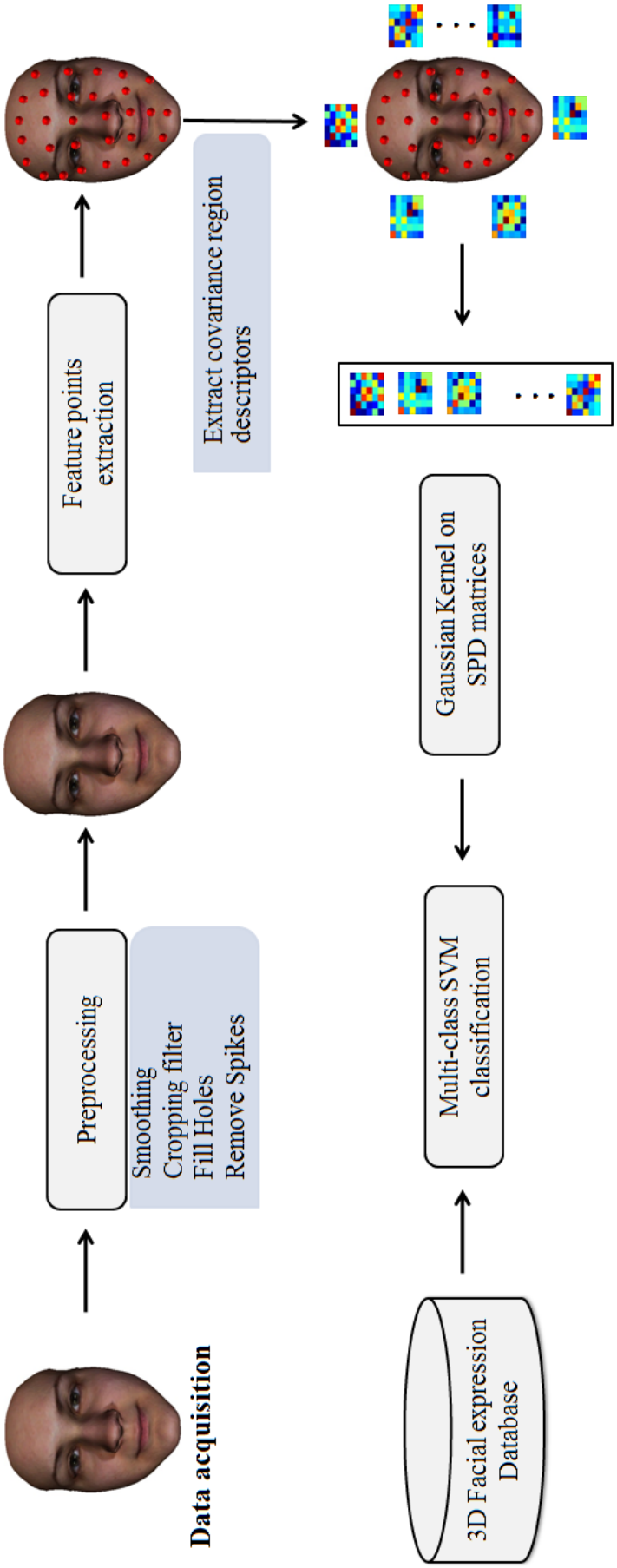


FIGURE 4.3 – *Overview of the proposed 3D facial expression recognition method.*

FIGURE 4.4 – *The four levels of the six expression variations for the same person from BU-3DFE dataset.*

makes the classification of covariance matrices using conventional algorithms such as SVM unsuitable. To overcome this issue, we apply a Gaussian radial basis function (RBF) which maps the covariance matrices to an infinite dimensional Hilbert space. This intuitively, yields a very rich representation. In R^d, the Gaussian kernel can be expressed as :

$$K_G(x_i, x_j) = \exp(\|x_i - x_j\|^2 / 2\sigma^2), \tag{4.1}$$

which makes use of the Euclidean distance between two data points x_i and x_j. To define a kernel on a Riemannian manifold, we would like to replace the Euclidean distance by a more accurate geodesic distance on the manifold.

The advantage of computing positive definite kernels on a Riemannian manifold of the SPD matrices is that it directly allows us to make use of algorithms developed for R^d while still accounting for the geometry of the manifold.

In the following, we use $\mathcal{K}(:,:)$, $\mathcal{H}$ and $\Phi(X)$ to denote the kernel function, the reproducing kernel Hilbert space, generated by $\mathcal{K}$, and the feature vector in $\mathcal{H}$ to which X is mapped, respectively. In this chapter, the function Φ is not explicitly expressed and solely the Gaussian kernel is used by changing the Euclidean distance by the distance on the Riemannian manifold defined by Equation (3.3) :

$$\mathcal{K}(X_i, X_j) = \exp(-d_g(\mathrm{X}_i, \mathrm{X}_j)^2 / 2\sigma^2), \tag{4.2}$$

where Φ is a mapping from $\mathcal{M}$ to $\mathcal{H}$ such that $d_g(X_i, X_j) = \left\|\Phi(X_i) - \Phi(X_j)\right\|_{\mathcal{H}}$.

Given a set of labeled samples $\{(X_i, y_i)\}_{i=1}^N$ where $X_i \in \mathcal{M}$ and the labels $y_i \in \{-1, 1\}$, the basic idea of SVM is to construct a hyperplane or set of hyperplanes, which is/are used for feature classification. A good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class, called support vectors. The

class of a test point is determined by the position of the mapping $\Phi(X)$ in $\mathcal{H}$ relative to the separating hyperplane. SVM is known to possess good generalization properties and to perform well in high-dimensional feature spaces. The mapping Φ is generally non-linear and the decision function is based on the sign of :

$$h(X) = b + \sum_{i=1}^{N} \alpha_i y_i \langle \Phi(X_i), \Phi(X) \rangle . \tag{4.3}$$

The kernel $\mathcal{K}(:,:)$ is defined by $\mathcal{K}(X_i, X_j) = \langle \Phi(X_i), \Phi(X) \rangle$.

Since covariance matrices are considered in this work as local descriptors, we propose to apply the SVM algorithm to this local representation. For this end, we build a global kernel function so that one can compare two 3D facial expressions by using the covariance descriptors.

In order to do so, we propose to use the matching kernel method proposed in (Wallraven et al. 2003), which satisfies the Mercer condition and thus is suitable for our application.

Given two expressive faces represented by two sets of covariance matrices $S_1 = \{X_i\}_{i=1..n}$ and $S_2 = \{X_j\}_{j=1..n}$, we first compute a matrix of similarity scores between S_1 and S_2. Common choices for the similarity measure, called also the minor kernel, are the RBF-kernel given by Equation 4.1. The kernel value can then be computed as the average over the best matching scores of the elements in S_1 and S_2 as :

$$K(S_1, S_2) = \frac{1}{2} \left[\hat{K}(S_1, S_2) + \hat{K}(S_2, S_1) \right], \tag{4.4}$$

where $\hat{K}(S_1, S_2) = \frac{1}{|S_1|} \sum_{i=1}^{|S_1|} \max_{j=1..|S_2|} \mathcal{K}(X_i, X_j)$.

Our manifold kernel SVM classification method can easily be extended to the multi-class case with standard one-versus-one or one-versus-all procedures.

4.4 Experimental results

In order to demonstrate the performance of the proposed method, we have first preprocessed the 3D surfaces. Then, we have uniformly sampled $m = 30$ feature points. Around each feature point p_i, we extract one patch P_i of radius r = 15% of the radius of the cropped face's bounding sphere. For each patch, we extract 6×6 covariance matrices computed from the feature vector : $[x, y, z, k_1, k_2, D]$. We then demonstrate the use of our kernel-based method for the task of FER with the proposed kernel SVM on Sym_d^+ as described in Section 4.3.

4.4.1 Experimental results on BU-3DFE dataset

To evaluate our approach, we perform a 10 fold-cross validation, where BU-3DFE subjects are ten times randomly divided into two parts; a training set consisting of 90 subjects, and a test set consisting of the rest 10 subjects. We then use our manifold kernel SVM in a multi-class (one-against-all) setting. Results hereinafter are averaged across the ten-folds. Table 4.1 reports the resulting confusion matrix where the columns represent the predicted expressions and the rows represent the actual expression. The recognition accuracy of each expression is presented in **bold**, remaining values present the percentage of miss-classified items.

Expression	HA	FE	DI	AN	SA	SU
HA	**97.75**	0.75	0	0	0	1.5
FE	4.7	**91.67**	3.63	0	0	0
DI	2.56	2.77	**94.67**	0	0	0
AN	0	0	1.5	**88.00**	10.5	0
SA	0	0	2.41	7.66	**85.33**	4.6
SU	0	0	0.75	0.92	0	**98.33**

Table 4.1 – *Confusing matrix of facial expression recognition (%) on BU-3DFE dataset : Happiness (HA), Fear (FE), Disgust (DI), Anger (AN), Sadness (SA), Surprise (SU).*

From Table 4.1, we can see that our method gives higher performance on happiness and surprise expressions. This is due to the distinctive large

deformations they make on face surfaces. The difference between sadness and anger expressions is more subtle and thus explains their confusion.

Table 4.2 presents the comparison results of our method performance with respect to the state-of-the-art ones. The results are reported on the BU-3DFE dataset following the same evaluation protocol over the six facial expressions.

Overall the dataset, Soyel et Demirel (2010) achieve 93.23% using probabilistic neural network for expression classification. Berretti et al. (2010b) applied SIFT descriptor and SVM for classification and achieved 77.53%. Huynh et al. (2016), on the other hand, used Convolutional neural network and achieved 92.73%. Azazi et al. (2015) used a pool of Speed Up Robust Features descriptors and yielded an average recognition accuracy of 85.81%.

Method	**HA**	**FE**	**DI**	**AN**	**SA**	**SU**	**Overall**
Soyel et Demirel (2010)	94.1	90.0	93.9	**91.7**	**90.8**	98.9	**93.23**
Mpiperis et al. (2008)	99.2	**97.9**	**100**	83.6	62.4	**100**	90.51
Berretti et al. (2010b)	86.9	63.6	73.6	81.7	64.6	94.8	77.53
Huynh et al. (2016)	**100**	86.7	95.2	91.3	87.5	95.7	92.73
Azazi et al. (2015)	93.50	73.67	90.83	78.67	83.67	94.50	85.81
Our method	97.75	91.67	94.67	88.00	85.33	98.33	92.62

TABLE 4.2 – *Comparison of classification rates (%) with state-of-the-art method on BU-3DFE dataset.*

Method	**Modality**	**Landmark**	**Classifier**
Soyel et Demirel (2010)	3D mesh	83 manual	NN
Mpiperis et al. (2008)	3D mesh	Global registration	ML
Berretti et al. (2010b)	2D	27 manual	SVM
Huynh et al. (2016)	2D+3D	-	CNN
Azazi et al. (2015)	2D	20 using SURF	SVM
Our method	3D mesh	30 automatic	SVM

TABLE 4.3 – *Protocol comparison with state-of-the-art method on BU-3DFE dataset.*

It should be noted that Soyel et Demirel (2010) testing setup is different from ours as shown in Table 4.3. Soyel and Demirel's method provides results using 83 manually annotated facial landmarks, while our approach automatically extracts the set of feature points. Notes that the proposed approach outperforms Soyel et Demirel (2010)'s one when dealing with these three different expressions : (HA, FE, DI). Our approach gives a

better performance compared to Mpiperis et al. (2008)'s approach when dealing with the following expressions : (AN, SA). With respect to Huynh et al. (2016)'s work, our approach performs better when dealing with FE and SU expressions. From the reported results, one can also notice that for all the methods, the happiness and the surprise expression are the easiest to be recognized, whereas the sadness and anger expressions are more challenging.

Moreover, the comparison with state-of-the-art methods demonstrates that our method gives challenging results (92.62% overall recognition rate). This performance is achieved due to the discrimination power of the covariance descriptors and the accurate classification of the manifold kernel SVM.

To give more insight about the efficiency of our proposed method and the confused expressions, we presented in Table 4.4 a comparison between the items of our confusion matrix with those of Huynh et al. (2016). From this table, we can see that AN and SA expressions are the more confusing with each other, which is similar to the finding of Huynh et al. (2016) and Azazi et al. (2015)'s methods. Our method on the other hand has successfully distinguished between FE and SU expressions compared with Huynh et al. (2016)'s method which easily confused to classify FE expression into SU.

According to the above comparisons with state-of-the-art methods, we can clearly see that our method performance is steadier over the six expressions as presented in Figure4.5. This strengthens our first claim about the robustness of covariance representation with respect to intra-class facial expression variabilities as well as its efficiency in capturing inter-class ones.

Expression	HA	FE	DI	AN	SA	SU
HA	**97.75 / 100**	0.75 / 0	0 / 0	0 / 0	0 / 0	1.5 / 0
FE	4.7 / 0	**91.67 / 86.7**	3.63 / 0	0 / 0	0 / 0	0 / 13.3
DI	2.56 / 0	2.77 / 4.8	**94.67 / 95.2**	0 / 0	0 / 0	0 / 0
AN	0 / 0	0 / 4.4	1.5 / 0	**88.00 / 91.3**	10.5 / 4.3	0 / 0
SA	0 / 0	0 / 0	2.41 / 0	7.66 / 12.5	**85.33 / 87.5**	4.6 / 0
SU	0 / 0	0 / 0	0.75 / 4.3	0.92 / 0	0 / 0	**98.33 / 95.7**

TABLE 4.4 – *Confusing matrix of our proposed method (left items) compared to Huynh et al. (2016)'s method (right items) on BU-3DFE dataset : Happiness (HA), Fear (FE), Disgust (DI), Anger (AN), Sadness (SA), Surprise (SU).*

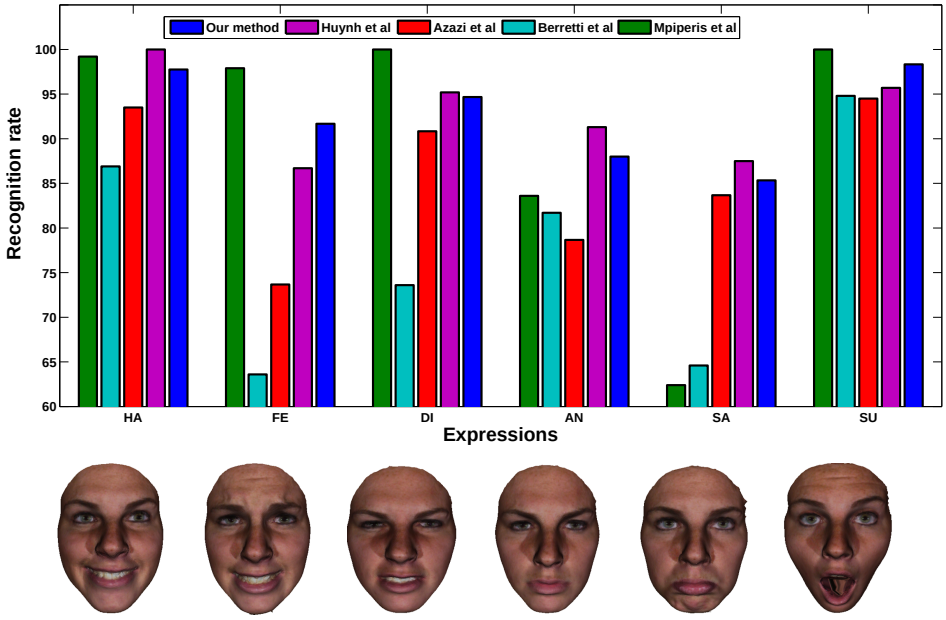


FIGURE 4.5 – *Recognition rate comparison on BU-3DFE dataset over the six prototypical expressions.*

4.4.2 Experimental results on Bosphorus dataset

As a second experiment, we further evaluated our proposed method on Bosphorus dataset. To this end, we followed the same experimental protocol of Azazi et al. (2015) : we applied 10 fold-cross validation technique over 420 faces from 60 subjects (7 expressions for each subject). Note that 60 subjects are selected randomly from 65 subjects. Results are averaged in Table 4.5.

From Table 4.5, we can see that happiness and surprise expressions have the best recognition rate, whereas, sadness and fear expressions are more challenging. This behavior is similar to that observed on BU-3DFE

Expression	HA	FE	DI	AN	SA	SU	NE
HA	**93.00**	2.50	4.50	0	0	0	0
FE	5.00	**81.00**	1.00	0	3.50	9.50	0
DI	3.25	3.75	**85.25**	0	1.75	0	6.00
AN	0	0	7.25	**86.25**	3.50	0	3.00
SA	0	0	9.25	0	**79.75**	1.50	9.50
SU	1.50	8.00	0	0	0	**90.50**	0
NE	0	0	0	10.75	1.75	0	**87.50**

TABLE 4.5 – *Confusing matrix of facial expression recognition (%) on Bosphorus dataset : Happiness (HA), Fear (FE), Disgust (DI), Anger (AN), Sadness (SA), Surprise (SU), Neutral (NE).*

Table 4.6 presents a comparison of our recognition performance with state-of-the-art methods. We can clearly see that our method gives the highest recognition performance (86.17%) followed by Azazi et al. (2015) with 84.10%. Chun et al. (2013) on the other hand employed the depth images and the 2D texture images and obtained 76.98%. Note that in Chun et al. (2013)'s method, landmarks were located manually (see Table 4.7). Wang et al. (2013) used curvature based descriptors with LBP, and achieved 76.56%. Vretos et al. (2011) achieved 60.53% using Zernike moments.

More specifically, our method gives the highest classification performance compared to the other state-of-the-art methods when dealing with AN, SA and NE expressions. Furthermore, compared to Azazi et al. (2015), our method gives better performance on AN, SA, SU and NE expressions. Our approach gives better performance compared to Wang et al. (2013)'s approach when dealing with HA, FE, DI, AN, and SA expressions. Note that Wang et al. (2013) and Vretos et al. (2011) didn't use Neutral expression (See table 4.7). Finally, compared to Vretos et al. (2011)'s method, our method gives better recognition performance over the six prototypical expressions.

To further study the efficiency of our method, we present in Table 4.8 a comparison between items of our confusion matrix with those of the best state-of-the-art performed method in Table 4.6 (i.e. Azazi et al. (2015)). This comparison shows that both methods confused in recognizing NE and AN expressions with each other, as well as FE and SU expressions. SA expression on the other hand is confused with DI and NE expressions. Although the two methods in comparison give quite similar confused expressions, our method has lower error compared to Azazi et al. (2015)'s method. This explains why our method delivers the highest overall performance.

4.4.3 System evaluation

In this section, we evaluate the performance of the proposed method with respect to the main parameters of the proposed approach when dea-

Method	HA	FE	DI	AN	SA	SU	NE	Overall
Azazi et al. (2015)	**97.50**	**86.25**	**90.00**	82.50	67.50	83.75	81.25	84.10
Chun et al. (2013)	-	-	-	-	-	-	-	76.98
Wang et al. (2013)	92.50	62.80	70.60	63.50	74.50	**95.60**	-	76.56
Vretos et al. (2011)	92.30	43.10	58.50	70.80	50.80	47.70	-	60.53
Our method	93.00	81.00	85.25	**86.25**	**79.75**	90.50	**87.50**	**86.17**

TABLE 4.6 – *Comparison of classification rates (%) with state-of-the-art method on Bosphorus dataset.*

Method	Modality	Landmark	Expressions	Classifier
Azazi et al. (2015)	2D	automatic	7	SVM with EPE
Chun et al. (2013)	2D+3D	manual	6	SVM+NN
Wang et al. (2013)	2D+3D	automatic	6	SVM
Vretos et al. (2011)	2D	automatic	6	SVM
Our method	3D mesh	automatic	7	SVM

TABLE 4.7 – *Protocol comparison with state-of-the-art method on Bosphorus dataset.*

ling with BU-3DFE and Bosphorus datasets. We study the effect of the number of sampled points as well as the effect of the patch size, and the position of the sampled points on the classification performance. In the first experiment, we set the patch radius r = 15% of the cropped face's bounding sphere and we vary the number of sample points m between 10 to 40. The reported results in Figure 4.6 show that the performance over all faces becomes stable when the number of sample points is larger than 30. This is predictable since small number of points will result in a coarse representation of the 3D face.

We also analyzed how the classification performance of the proposed method varies with respect to the patch radius r . For this end, we set the number of sample points m = 30, and vary the patch radius between 10%

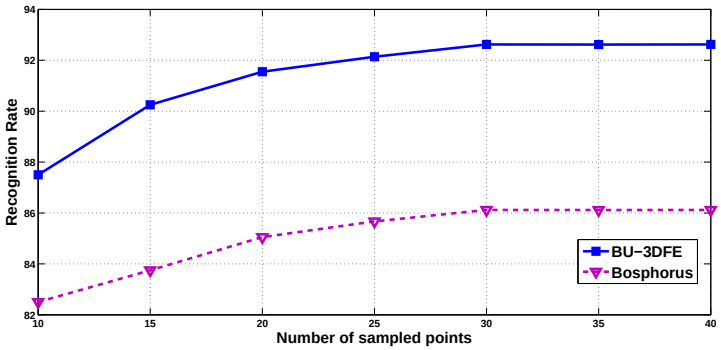


FIGURE 4.6 – *Effect of the number of patches on the classification performance of the proposed method on BU-3DFE and Bosphorus datasets.*

Expression	HA	FE	DI	AN	SA	SU	NE
HA	**93.00 / 97.50**	2.50 / 0	4.50 / 2.50	0 / 0	0 / 0	0 / 0	0 / 0
FE	5.00 / 1.25	**81.00 / 86.25**	1.00 / 1.25	0 / 1.25	3.50 / 0	9.50 / 10.00	0 /0
DI	3.25 / 3.75	3.75 / 1.25	**85.25 / 90.00**	0 / 2.50	1.75 / 0	0 / 0	6.00 / 2.50
AN	0 / 0	0 /1.25	7.25 / 5.00	**86.25 / 82.50**	3.50 / 2.50	0 / 1.25	3.00 / 7.50
SA	0 / 0	0 / 0	9.25 / 11.25	0 / 5.00	**79.75 / 67.50**	1.50 / 1.25	9.50 / 15.00
SU	1.50 / 0	8.00 / 13.75	0 / 2.50	0 / 0	0 / 0	**90.50 / 83.75**	0 / 0
NE	0 / 0	0 / 2.50	0 / 3.75	10.75 / 10.00	1.75 / 1.25	0 / 1.25	**87.50 / 81.25**

TABLE 4.8 – *Confusing matrix of our proposed method (left items) compared to Azazi et al. (2015)'s method (right items) on Bosphorus dataset : Happiness (HA), Fear (FE), Disgust (DI), Anger (AN), Sadness (SA), Surprise (SU), Neutral (NE).*

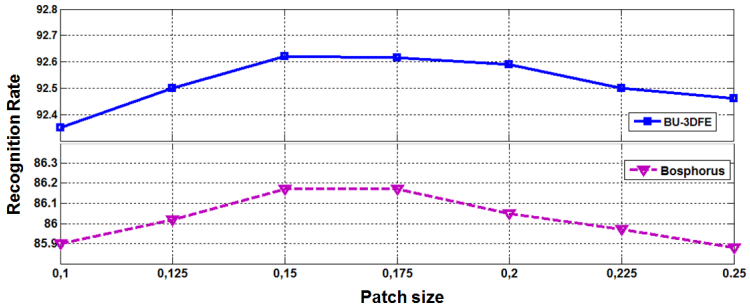


FIGURE 4.7 – *Effect of the patch radius on the classification performance of the proposed method. The reported results are on both BU-3DFE and Bosphorus datasets.*

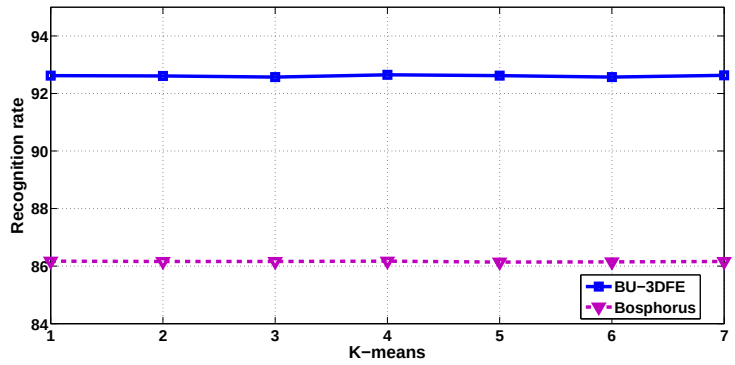


FIGURE 4.8 – *Effect of the position of the sampled points on the classification performance of the proposed method on BU-3DFE dataset.*

to 25% of the total radius of the cropped face's bounding sphere. Please note that in this setting, the patches may overlap. Figure 4.7 shows that the performance remains stable when r varies between 15% and 20%. The performance starts to drop when choosing values outside this interval. Note that, similar to all local descriptor, this behavior was predictable since very small patches do not capture sufficient geometric properties of the shapes. Large patches on the other hand capture only coarse features, which may not be sufficiently discriminative.

Furthermore, we also study how the classification performance varies with respect to the position of the m uniformly sampled points. For this end, we apply a uniform sampling using k-means algorithm, for seven times. We hence obtain seven different distributions of feature points. In this experiment, we set the number of sampled points m=30, and the patch radius $r = 15\%$. From Figure 4.8, we can see that the classification perfor-

mance is almost stable with respect to different positions of the sampled points.

4.5 Conclusion

In this chapter, we proposed a novel method for 3D facial expression recognition that uses the local covariance descriptor and kernel-based classifier on Riemannian manifold. Since covariance matrices fuse different types of features, their description power are very interesting for FER where expressions exhibit large intra-class variability. This description considers covariance matrices as an unordered set of descriptors and therefore we used a kernel on sets rather than directly using Gaussian kernel in the SVM classifier. For this end, we build a global kernel function so that one can compare two facial expressions. The Gaussian kernel is then used to map the covariance matrices into a high dimensional Hilbert space. This mapping allows to extend the standard kernel methods, originally introduced for the analysis on Euclidean spaces, to the non-linear Riemannian manifold of covariance matrices. Moreover, Gaussian kernel utilized in our proposed method helps understand the inherent structure of the 3D facial data to enhance the 3D FER accuracy. Reported results on BU-3DFE and Bosphorus datasets show that our method efficiently succeed to classify facial expressions independently to the face identity, and demonstrate its superiority compared to state-of- the-art methods.

5 Conclusion

5.1 Summary

In this thesis we presented two main contributions to 3D face analysis including 3D face recognition/verification and 3D facial expression recognition. In the first contribution, we proposed a 3D face recognition method based on covariance matrices of descriptors rather than the descriptors themselves. We demonstrated the efficiency of our proposed method to tackle the problem of expression variation as well as pose and partial occlusions. This robustness is achieved due to the ability of covariance descriptors to efficiently combine multiple features into a single descriptor and the invariance with respect to the ordering of points and number of feature vectors used for their computation. Moreover, the size of covariance matrices does not depend on the size of the region from which they were extracted, but of the size of feature vectors, therefore, they can be computed from variable sized regions. Furthermore, covariance matrices are low dimensional compared to joint feature histograms.

Since covariance matrices are elements of the non-linear manifold of SPD matrices, their comparison requires the computation of geodesic distances on the manifold using proper metrics. Therefore, we assessed six distances (geodesic and non-geodesic) to match covariance matrices, and demonstrated that geodesic distances are more suitable to match 3D faces using their covariance descriptors.

In the matching process, we have assessed two different strategies after spatial registration of the 3D faces. The first strategy is to compute

optimal match using a Hungarian solution for matching unordered set of covariance matrices. The total cost of matching is used as a measure of dissimilarity between the pair of 3D faces. The second strategy is to compute a mean distance by integrating the chosen metric over the pairs of homologous regions.

We have also proposed an extension of the our covariance-based 3D face recognition method using hierarchical description. To do so, we represented a 3D face using a set feature points, around each of which we consider three description levels starting from a small region to a bigger overlapped region, each region is represented by a covariance matrix. Experimental results showed that the use of the three hierarchical levels improves the recognition performance compared to the use of each level individually. This performance can be explained by the fact that each hierarchical level captures some specific characteristics which are complementary. In comparison to the previously published methods on the three challenging datasets (i.e. GAVAB, FRGCv2, BU-3DFE), our method achieves a superior performance in terms of recognition performance.

In the second contribution, we proposed a new method for 3D facial expression classification regardless to the face identity using covariance descriptors and kernel methods on Riemannian manifold. We proved that covariance descriptors are very interesting for FER where expressions exhibit large intra-class variability. This description considers covariance matrices as an unordered set of descriptors and therefore we used a kernel on sets rather than directly using Gaussian kernel in the SVM classifier. For this end, we build a global kernel function so that one can compare two facial expressions. The Gaussian kernel is then used to map the covariance matrices into a high dimensional Hilbert space. This mapping allows to extend the standard kernel methods, originally introduced for the analysis on Euclidean spaces, to the non-linear Riemannian manifold of covariance matrices. Moreover, Gaussian kernel utilized in our experiments helps understand the inherent structure of the 3D facial data to enhance the 3D FER

accuracy. Reported results on BU-3DFE and Bosphorus datasets show that our method efficiently succeed to classify facial expressions independently to the face identity. Average recognition rate attained on BU-3DFE 92.62%, and 86.17% on Bosphorus dataset. The reported results demonstrate the superiority of our proposed method compared to state-of- the-art ones.

Since 3D data provide naturally more information on the geometric proprieties of the facial shape, 3D FER methods are preferable than their 2D counterparts, especially because of their invariance against illumination changes. In the other hand, our covariance based method is generic. The covariance matrices can be computed from different type of features including 2D ones and its performance depends on the used features.

An important aspect of our 3D FER is that it is applicable to real situations due to the automatic feature points extraction. On the other hand, most state-of-the-art methods use a predefined set of manually selected points or landmarks, which makes these methods difficult to use in practice.

5.2 Perspectives

Further work can be conducted in order to enhance our proposed methods for 3D face recognition and 3D facial expression recognition, future directions are also presented as follows :

1. In this thesis, we proposed to extract covariance descriptors around a set of feature points (i.e. landmarks). These landmarks are extracted after a uniform sampling using k-means algorithm. However, according to the purpose of facial analysis, different parts of the face possess different levels of importance. For example, the eyes and nose are the most robust parts for the face recognition task. On the other hand, if the purpose of facial analysis is facial expression recognition, then the mouth region will naturally provide important information that has to be taken into account for classification. In either cases, landmarking salient points in our method allows to reduce the number of covariance descriptors per face and may provide more robust results.

2. In our 3D FER contribution, we applied the global kernel presented in (Wallraven et al. 2003) in order to compare the unordered set of covariance descriptors, our method is generic so that other kernel methods can also be applied which should satisfy Mercer condition.

3. Since Bosphorus and BU-3DFE datasets provide 2D face scans for each 3D facial shape, a multi-modal (2D+3D) covariance based method can be applied. This combination may improve the FER performance thanks to the elegant way for combining multiple heterogeneous features without normalization provided by covariance matrices.

4. This thesis deals with recognizing static 3D facial expressions using two well-known 3D datasets. Recently, 4D datasets of dynamic 3D facial samples have become available (i.e. BU-4DFE (Yin et al. 2008), Hi4D-ADSIP (Matuszewski et al. 2012) and D3DFACS (Petrovska-

Delacrétaz et al. 2008)). These datasets encode temporal cues that are indicative of more complex expressions and give the most accurate representation of facial articulations by including temporal information of dynamic facial movements. Furthermore, since 3D dynamic facial scans help to handle the low intensity of expressions, 4D FER is more similar to the real life which can be a future direction of our work. Our covariance based method can also deal with 4D FER since several studies have successfully extended the use of covariance descriptors to the temporal dimension in action and gesture recognition (e.g. (Bhattacharya et al. 2016, Sanin et al. 2013)).

5. In this thesis, we only handled the six prototypical expressions to understand the human emotions. In our future work, we can further investigate the ability of recognizing the action units of a face. These action units refer a measuring of specific facial muscle movements, and are automatically related to their contraction. Thus recognizing the action units can be useful for further understanding the behaviors and the appearance of the face.

Annexes

A

Sommaire

A.1 Hungarian algorithm

The Hungarian algorithm is a combinatorial optimization algorithm that solves the assignment problem in polynomial time. It was proposed by Harold Kuhn (Kuhn 1955, Munkres 1957). The assignment problem deals with assigning machines to tasks, workers to jobs, soccer players to positions, and so on. The goal is to determine the optimum assignment that, for example, minimizes the total cost or maximizes the team effectiveness. In our contribution, we apply it to find the best permutation between probe and gallery faces which minimizes the cost of matching using their covariance matrices so that we can classify the probe face to the nearest gallery face.

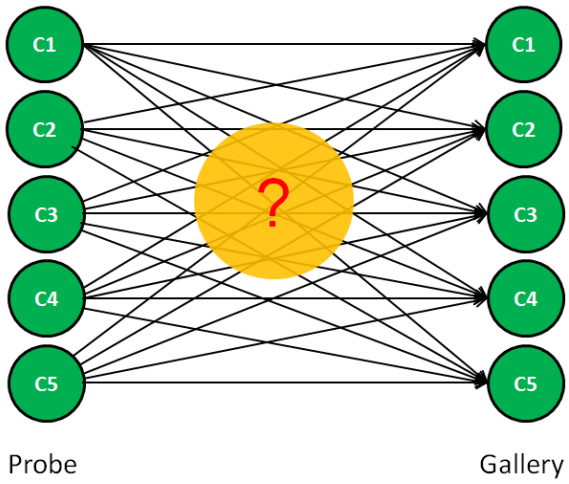


Figure A.1 – *The assignment problem to match probe and gallery faces using their covariance descriptors.*

A.2 Iterative closest point (ICP)

Besl et McKay (1992) proposed the ICP method, which computes a rigid transformation and aligns a data point set to a model point set. The alignment is performed by minimizing the value reported by an objective function that adds the sum of squared distances between pairs of nearest neighbors with one element in the model shape and the other in the aligned data shape. Figure A.2 shows the steps of the alignment process. In

this thesis, we applied ICP to align probe faces to gallery faces in order to make them of the same position.

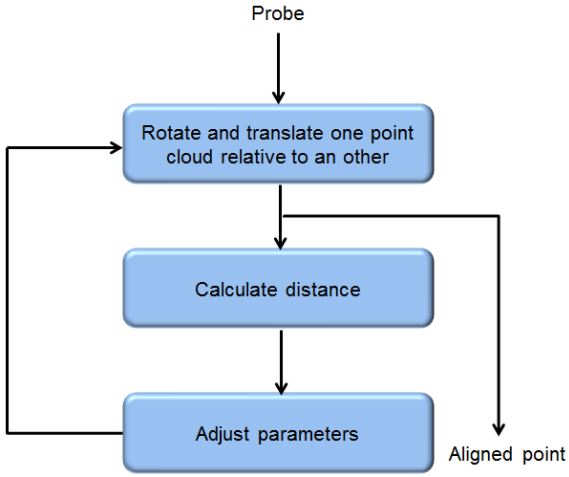


FIGURE A.2 – *Steps of aligning one point cloud to the other using Iterative Closest Point algorithm*

A.3 Principal curvatures

Principal curvatures are one of the most used features on 3D shape analysis. On a 2D plane or 3D surface, the curvature at a particular point measures how the curve bends by different amounts in different directions at that point. It is given by the inverse radius of the osculating circle at that point. To compute the 3D surface curvature, only two angles are selected : the ones giving the maximal and minimal curvature values known as first (k1) and second (k2) principal curvatures (Creusot et al. 2013). Table A.1 presents the surface classes according to their curvatures.

	$k_1 < 0$	$k_1 = 0$	$k_1 > 0$
$k_1 < 0$	Concave ellipsoid	Concave cylinder	Hyperboloid surface
$k_1 = 0$	Concave cylinder	Plane	Convex cylinder
$k_1 > 0$	Hyperboloid surface	Convex cylinder	Convex ellipsoid

TABLE A.1 – *Surface classes*

A.4 Support Vector Machine (SVM)

SVM is a supervised machine learning method which is popular for addressing binary classification problems using functions that can opti-

mally separate data. In the case of two linearly separable classes, there exists an infinite number of hyperplanes for separating the data. The aim of SVM is to find the optimal hyperplane that separates data with maximizing the distance between the two classes (i.e. maximizing the margin). The nearest points, which are used only for the determination of the hyperplane, are called support vectors.

We can distinguish between two models of SVM according to the separability of data, linear-SVM and non-linear-SVM. The linear-SVM are the simplest because they are linearly separable as presented in FigureA.3. In the non-linear-SVM, the data are transformed to be represented in a large space where they are linearly separable (See FigureA.4). To evaluate the classification, the most used technique is 10-cross validation which we applied in this thesis.

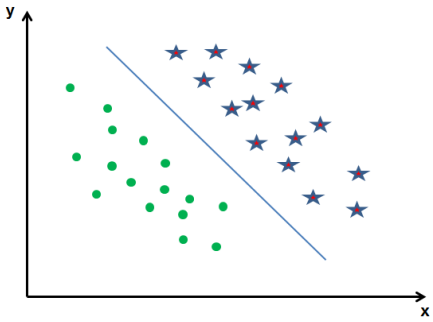


FIGURE A.3 – *Linearly separable data separated by a straight line.*

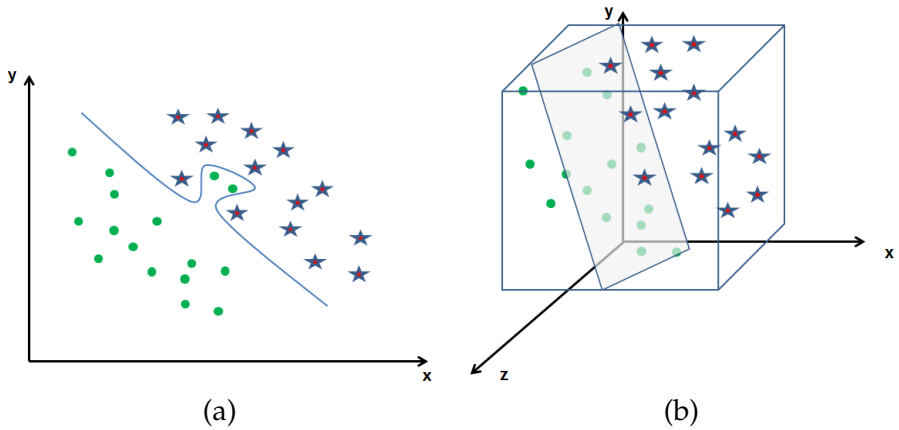


FIGURE A.4 – *(a) : Non-linearly separable data separated by a curved line, (b) : Plan separation after a transformation of the data into a 3D plane.*

A.5 Measuring biometric system performance

The performance of a biometric system in the verification or identification scenarios can be evaluated based on the match scores obtained by the used matching algorithm. In the case of face recognition system, for a set of test faces, let *id* be the number of identities and *sam* be the number of face samples per identity, the total number of samples is $N_T = id * sam$. By comparing each of N_T samples against the remaining $N_T - 1$ samples, a total of $\frac{1}{2}\ N_T\ (N_T - 1)$ similarity match scores can be computed. This comparison is referred to as *all versus all protocol*. In computing the match scores for an *all versus all* protocol, two classes of match scores are generated namely *genuine* and *impostor* match scores. Genuine match scores denote the scores generated when comparing two face samples belonging to the same individual. Impostor scores denote the scores generated when matching two face samples belonging to different individuals.

The verification performance is typically evaluated by assessing the false acceptance rate (FAR) and the false reject rate (FRR). As presented before in Figure1.4, the FAR denotes the percentage of impostor scores that exceed a numerical threshold and are incorrectly classified as matches. The FRR denotes the percentage of genuine scores that are below a threshold and are incorrectly classified as non-matches. Graphically, the FAR and FRR are often expressed by a Receiver Operating Characteristic (ROC) curve. In this thesis, we plot the true acceptance rate (TAR) versus FAR by varying the threshold to evaluate our verification system.

The true acceptance rate (also called recall, sensitivity) of a classifier is estimated as :

$$TAR \approx \frac{\text{Positives correctly classified}}{\text{Total positives}} \tag{A.1}$$

The false acceptance rate (also called false alarm rate) of the classifier is :

$$FAR \approx \frac{\text{Negatives incorrectly classified}}{\text{Total negatives}} \tag{A.2}$$

To evaluate the identification performance, a set of N_{probe} probe samples is compared against a set of N_{gal} gallery samples. This results N_{probe} sets of match scores, with each set containing N_{gal} match scores. The match scores in each set are sorted in descending order.

In closed-set identification, the ordered score sets from the N_{probe} are used to estimate the probability that the correct matching identity pertaining to a probe is observed within the top K ranks. These probabilities are typically expressed visually through the Cumulative Match Characteristic (CMC) curve. Unlike the ROC curve, which is generated by looking at genuine and impostor scores all-at-once, the data in the CMC curve is obtained based on the explicit ordering of genuine and impostor scores for each face probe.

A.6 Cross validation

In order to evaluate the classification performance using k-cross validation approach, the data is divided into (k=10 in the case of 10-cross validation) for testing and training partitions. Accordingly, k iterations of training and validation are performed so that in each iteration, one fold is used for validation and $k-1$ folds are used for training.

Bibliography

Abate, Andrea F, Nappi, Michele, Riccio, Daniel, et Sabatino, Gabriele. 2d and 3d face recognition : A survey. *Pattern Recognition Letters*, 28(14), p. 1885–1906, 2007. (Cited pages 7 et 34.)

Al-Osaimi, F, Bennamoun, Mohammed, et Mian, Ajmal. An expression deformation approach to non-rigid 3d face recognition. *International Journal of Computer Vision*, 81(3), p. 302–316, 2009. (Cited page 27.)

Al-Osaimi, Faisal R, Bennamoun, Mohammed, et Mian, A. Integration of local and global geometrical cues for 3d face recognition. *Pattern Recognition*, 41(3), p. 1030–1040, 2008. (Cited page 63.)

Al-Osaimi, Faisal R, Bennamoun, Mohammed, et Mian, Ajmal. Spatially optimized data-level fusion of texture and shape for face recognition. *IEEE Transactions on Image Processing*, 21(2), p. 859–872, 2012. (Cited pages 6, 7 et 36.)

Alyuz, Nese, Gokberk, Berk, et Akarun, Lale. Regional registration for expression resistant 3-d face recognition. *Information Forensics and Security, IEEE Transactions on*, 5(3), p. 425–440, 2010. (Cited page 63.)

Alyuz, Nese, Gokberk, Berk, et Akarun, Lale. 3-d face recognition under occlusion using masked projection. *Information Forensics and Security, IEEE Transactions on*, 8(5), p. 789–802, 2013. (Cited pages x et 30.)

Amenta, Nina, Choi, Sunghee, et Kolluri, Ravi Krishna. The power crust. Dans *Proceedings of the sixth ACM symposium on Solid modeling and applications*, pages 249–266. ACM, 2001. (Cited page 18.)

Arsigny, Vincent, Fillard, Pierre, Pennec, Xavier, et Ayache, Nicholas. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2), p. 411–421, 2006. (Cited pages 54 et 84.)

Azazi, Amal, Lutfi, Syaheerah Lebai, Venkat, Ibrahim, et Fernández-Martínez, Fernando. Towards a robust affect recognition : Automatic facial expression recognition in 3d faces. *Expert Systems with Applications*, 42(6), p. 3056–3066, 2015. (Cited pages 41, 43, 92, 93, 95, 96, 97 et 98.)

Beeler, Thabo, Bickel, Bernd, Beardsley, Paul, Sumner, Bob, et Gross, Markus. High-quality single-shot capture of facial geometry. Dans *ACM Transactions on Graphics (ToG)*, volume 29, page 40. ACM, 2010. (Cited page 15.)

Bellil, Wajdi, Brahim, Hajer, et Amar, Chokri Ben. Gappy wavelet neural network for 3d occluded faces : detection and recognition. *Multimedia Tools and Applications*, 75(1), p. 365–380, 2016. (Cited page 30.)

Benedikt, Lanthao, Cosker, Darren, Rosin, Paul L, et Marshall, David. Assessing the uniqueness and permanence of facial actions for use in biometric applications. *IEEE Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans*, 40(3), p. 449–460, 2010. (Cited page 15.)

Berretti, Stefano, Bimbo, Alberto Del, et Pala, Pietro. 3d face recognition using isogeodesic stripes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12), p. 2162–2177, 2010a. (Cited page 28.)

Berretti, Stefano, Del Bimbo, Alberto, et Pala, Pietro. Sparse matching of salient facial curves for recognition of 3-d faces with missing parts. *IEEE Transactions on information forensics and security*, 8(2), p. 374–389, 2013. (Cited page 30.)

Berretti, Stefano, Del Bimbo, Alberto, Pala, Pietro, Amor, Boulbaba Ben, et Daoudi, Mohamed. A set of selected sift features for 3d facial expres-

sion recognition. Dans *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 4125–4128. IEEE, 2010b. (Cited pages 41, 43 et 92.)

Besl, Paul J et McKay, Neil D. Method for registration of 3-d shapes. Dans *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992. (Cited pages 23, 47 et 108.)

Beymer, David James. Face recognition under varying pose. Dans *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 756–761. IEEE, 1994. (Cited page 26.)

Bhattacharya, Subhabrata, Souly, Nasim, et Shah, Mubarak. Covariance of motion and appearance featuresfor spatio temporal recognition tasks. *arXiv preprint arXiv :1606.05355*, 2016. (Cited page 105.)

Bowyer, Kevin W, Chang, Kyong, et Flynn, Patrick. A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. *Computer vision and image understanding*, 101(1), p. 1–15, 2006. (Cited pages 6 et 18.)

Brunelli, Roberto et Poggio, Tomaso. Face recognition : Features versus templates. *IEEE transactions on pattern analysis and machine intelligence*, 15(10), p. 1042–1052, 1993. (Cited page 26.)

Chakrabarty, Ankush, Jain, Harsh, et Chatterjee, Amitava. Volterra kernel based face recognition using artificial bee colonyoptimization. *Engineering Applications of Artificial Intelligence*, 26(3), p. 1107–1114, 2013. (Cited page 36.)

Chebb, Z et Moakher, M. Means of hermitian positivedefinite matrices based on the log-determinant divergence function. *Linear Algebra Appl*, 40, 2012. (Cited page 53.)

Chen, Yang et Medioni, Gérard. Object modeling by registration of multiple range images. Dans *Robotics and Automation, 1991. Proceedings.,*

1991 IEEE International Conference on, pages 2724–2729. IEEE, 1991. (Cited pages 23 et 47.)

Cherian, Anoop, Sra, Suvrit, Banerjee, Arindam, et Papanikolopoulos, Nikolaos. Efficient similarity search for covariance matrices via the jensen-bregman logdet divergence. Dans *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2399–2406. IEEE, 2011. (Cited page 53.)

Cherian, Anoop, Sra, Suvrit, Banerjee, Arindam, et Papanikolopoulos, Nikolaos. Jensen-bregman logdet divergence with application to efficient similarity search for covariance matrices. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(9), p. 2161–2174, 2013. (Cited page 57.)

Chew, Wei Jen, Seng, Kah Phooi, Ang, Li-Minn, et al. Nose tip detection on a three-dimensional face range image invariant to head pose. Dans *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 18–20. Citeseer, 2009. (Cited page 21.)

Chun, Soon-Yong, Lee, Chan-Su, et Lee, Sang-Heon. Facial expression recognition using extended local binary patterns of 3d curvature. Dans *Multimedia and Ubiquitous Engineering*, pages 1005–1012. Springer, 2013. (Cited pages 43, 96 et 97.)

Cichocki, Andrzej et Amari, Shun-ichi. Families of alpha-beta-and gamma-divergences : Flexible and robust measures of similarities. *Entropy*, 12(6), p. 1532–1568, 2010. (Cited page 55.)

Colombo, Alessandro, Cusano, Claudio, et Schettini, Raimondo. 3d face detection using curvature analysis. *Pattern recognition*, 39(3), p. 444–455, 2006. (Cited page 21.)

Colombo, Alessandro, Cusano, Claudio, et Schettini, Raimondo. Three-dimensional occlusion detection and restoration of partially occluded faces. *Journal of mathematical imaging and vision*, 40(1), p. 105–119, 2011. (Cited page 23.)

Cortes, Corinna et Vapnik, Vladimir. Support-vector networks. *Machine learning*, 20(3), p. 273–297, 1995. (Cited page 83.)

Creusot, Clement, Pears, Nick, et Austin, Jim. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International journal of computer vision*, 102(1-3), p. 146–179, 2013. (Cited page 109.)

Darwin, Charles, Ekman, Paul, et Prodger, Phillip. *The expression of the emotions in man and animals*. Oxford University Press, USA, 1998. (Cited page 7.)

Deng, Xing, Da, Feipeng, et Shao, Haijian. Efficient 3d face recognition using local covariance descriptor and riemannian kernel sparse coding. *Computers & Electrical Engineering*, 2017. (Cited page 85.)

Derkach, Dmytro et Sukno, Federico M. Local shape spectrum analysis for 3d facial expression recognition. *arXiv preprint arXiv :1705.06900*, 2017. (Cited pages x et 39.)

Drira, Hassen, Ben Amor, Boulbaba, Srivastava, Anuj, Daoudi, Mohamed, et Slama, Rim. 3d face recognition under expressions, occlusions, and pose variations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(9), p. 2270–2283, 2013. (Cited pages 28, 29, 63, 65, 67 et 68.)

Dubes, Richard C et Jain, Anil K. Algorithms for clustering data, 1988. (Cited page 60.)

Ekman, Paul et Friesen, Wallace V. Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2), p. 124, 1971. (Cited pages 7, 8 et 35.)

El-Hakim, Sabry F, Beraldin, Jean-Angelo, et Blais, Francois. Comparative evaluation of the performance of passive and active 3d vision systems. Dans *Digital Photogrammetry and Remote Sensing'95*, pages 14–25. International Society for Optics and Photonics, 1995. (Cited page 17.)

Elaiwat, S, Bennamoun, M, Boussaid, F, et El-Sallam, A. 3-d face recognition using curvelet local features. *IEEE Signal processing letters*, 21(2), p. 172–175, 2014. (Cited page 33.)

Faltemier, Timothy C, Bowyer, Kevin W, et Flynn, Patrick J. A region ensemble for 3-d face recognition. *Information Forensics and Security, IEEE Transactions on*, 3(1), p. 62–73, 2008. (Cited pages 63 et 65.)

Fang, Tianhong, Zhao, Xi, Ocegueda, Omar, Shah, Shishir K, et Kakadiaris, Ioannis A. 3d facial expression recognition : A perspective on promises and challenges. Dans *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 603–610. IEEE, 2011. (Cited pages 36 et 81.)

Fang, Tianhong, Zhao, Xi, Ocegueda, Omar, Shah, Shishir K, et Kakadiaris, Ioannis A. 3d/4d facial expression analysis : an advanced annotated face model approach. *Image and vision Computing*, 30(10), p. 738–749, 2012. (Cited pages 36 et 81.)

Fasel, Beat et Luettin, Juergen. Automatic facial expression analysis : a survey. *Pattern recognition*, 36(1), p. 259–275, 2003. (Cited pages ix, 7, 8 et 36.)

Fua, Pascal. Regularized bundle-adjustment to model heads from image sequences without calibration data. *International Journal of Computer Vision*, 38(2), p. 153–171, 2000. (Cited page 16.)

Gokberk, Berk, Dutagaci, Helin, Akarun, Lale, Sankur, Bülent, et al. Representation plurality and fusion for 3-d face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(1), p. 155–173, 2008. (Cited page 17.)

Golub, Gene H et Van Loan, Charles F. *Matrix computations*, volume 3. JHU Press, 2012. (Cited page 53.)

Guo, Yulan, Lei, Yinjie, Liu, Li, Wang, Yan, Bennamoun, Mohammed, et Sohel, Ferdous. Ei3d : Expression-invariant 3d face recognition based

on feature and shape matching. *Pattern Recognition Letters*, 2016. (Cited pages ix, x, 21, 28 et 33.)

Hamm, Jihun et Lee, Daniel D. Grassmann discriminant analysis : a unifying view on subspace-based learning. Dans *Proceedings of the 25th international conference on Machine learning*, pages 376–383. ACM, 2008. (Cited pages 85 et 86.)

Harandi, Mehrtash T, Salzmann, Mathieu, et Hartley, Richard. From manifold to manifold : geometry-aware dimensionality reduction for spd matrices. Dans *Computer Vision–ECCV 2014*, pages 17–32. Springer, 2014a. (Cited page 86.)

Harandi, Mehrtash T, Salzmann, Mathieu, Jayasumana, Sadeep, Hartley, Richard, et Li, Hongdong. Expanding the family of grassmannian kernels : An embedding perspective. Dans *European Conference on Computer Vision*, pages 408–423. Springer, 2014b. (Cited pages 85 et 86.)

Harandi, Mehrtash T, Sanderson, Conrad, Shirazi, Sareh, et Lovell, Brian C. Kernel analysis on grassmann manifolds for action recognition. *Pattern Recognition Letters*, 34(15), p. 1906–1915, 2013. (Cited page 85.)

Hsu, Chih-Wei et Lin, Chih-Jen. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2), p. 415–425, 2002. (Cited page 83.)

Huang, Di, Ardabilian, Mohsen, Wang, Yunhong, et Chen, Liming. 3-d face recognition using elbp-based facial description and local feature hybrid matching. *Information Forensics and Security, IEEE Transactions on*, 7(5), p. 1551–1565, 2012. (Cited pages 63, 65, 67 et 68.)

Huang, Peisen S, Zhang, Chengping, et Chiang, Fu-Pen. High-speed 3-d shape measurement based on digital fringe projection. *Optical Engineering*, 42(1), p. 163–168, 2003. (Cited page 15.)

Huynh, Xuan-Phung, Tran, Tien-Duc, et Kim, Yong-Guk. Convolutional neural network models for facial expression recognition using bu-3dfe

database. Dans *Information Science and Applications (ICISA) 2016*, pages 441–450. Springer, 2016. (Cited pages 40, 43, 92, 93 et 94.)

Ilbeygi, Mahdi et Shah-Hosseini, Hamed. A novel fuzzy facial expression recognition system based on facial feature extraction from color face images. *Engineering Applications of Artificial Intelligence*, 25(1), p. 130–146, 2012. (Cited page 36.)

Introna, Lucas et Nissenbaum, Helen. Facial recognition technology a survey of policy and implementation issues. 2010. (Cited page 2.)

Irfanoglu, M Okan, Gokberk, B, et Akarun, Lale. 3d shape-based face recognition using automatically registered facial surfaces. Dans *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 183–186. IEEE, 2004. (Cited page 34.)

James, William. What is an emotion ? *Mind*, 9(34), p. 188–205, 1884. (Cited page 8.)

Jarvis, Ray. Range sensing for computer vision. *Three-dimensional object recognition systems*, 1, p. 17–56, 1993. (Cited page 15.)

Jayasumana, Sadeep, Hartley, Richard, Salzmann, Mathieu, Li, Hongdong, et Harandi, Mehrtash. Kernel methods on riemannian manifolds with gaussian rbf kernels. *IEEE transactions on pattern analysis and machine intelligence*, 37(12), p. 2464–2477, 2015. (Cited pages 84 et 85.)

Kanade, T. *Picture processing by computer complex and recognition of human faces*. PhD thesis, PhD thesis, Kyoto University, 1973. (Cited page 14.)

Kemelmacher-Shlizerman, Ira et Basri, Ronen. 3d face reconstruction from a single image using a single reference face shape. *IEEE transactions on pattern analysis and machine intelligence*, 33(2), p. 394–405, 2011. (Cited page 15.)

Kittler, J, Hilton, A, Hamouz, M, et Illingworth, J. 3d assisted face recognition : A survey of 3d imaging, modelling and recognition approachest.

Dans *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, pages 114–114. IEEE, 2005. (Cited page 16.)

Krizaj, Janez, Struc, Vitomir, et Dobrisek, Simon. Combining 3d face representations using region covariance descriptors and statistical models. Dans *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–7. IEEE, 2013. (Cited page 82.)

Kuhn, Harold W. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2), p. 83–97, 1955. (Cited pages 57 et 108.)

Kullback, Solomon. *Information theory and statistics*. Courier Dover Publications, 1997. (Cited page 55.)

Lee, Deokwoo et Krim, Hamid. 3d face recognition in the fourier domain using deformed circular curves. *Multidimensional systems and signal processing*, 28(1), p. 105–127, 2017. (Cited page 28.)

Lee, Mun Wai et Ranganath, Surendra. Pose-invariant face recognition using a 3d deformable model. *Pattern recognition*, 36(8), p. 1835–1846, 2003. (Cited page 31.)

Lei, Yinjie, Bennamoun, Mohammed, et El-Sallam, Amar A. An efficient 3d face recognition approach based on the fusion of novel local low-level features. *Pattern Recognition*, 46(1), p. 24–37, 2013. (Cited pages 29, 75 et 76.)

Lei, Yinjie, Guo, Yulan, Hayat, Munawar, Bennamoun, Mohammed, et Zhou, Xinzhi. A two-phase weighted collaborative representation for 3d partial face recognition with single sample. *Pattern Recognition*, 52, p. 218–237, 2016. (Cited pages 75 et 76.)

Lemaire, Pierre, Ardabilian, Mohsen, Chen, Liming, et Daoudi, Mohamed. Fully automatic 3d facial expression recognition using differential mean

curvature maps and histograms of oriented gradients. Dans *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–7. IEEE, 2013. (Cited pages 41 et 43.)

Lemaire, Pierre, Ben Amor, Boulbaba, Ardabilian, Mohsen, Chen, Liming, et Daoudi, Mohamed. Fully automatic 3d facial expression recognition using a region-based approach. Dans *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, pages 53–58. ACM, 2011. (Cited page 38.)

Li, Xiaoli et Da, Feipeng. Efficient 3d face recognition handling facial expression and hair occlusion. *Image and Vision Computing*, 30(9), p. 668–679, 2012. (Cited page 29.)

Li, Xiaoli, Ruan, Qiuqi, et Ming, Yue. 3d facial expression recognition based on basic geometric features. Dans *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 1366–1369. IEEE, 2010. (Cited page 37.)

Li, Xiaoxing, Jia, Tao, et Zhang, Hao. Expression-insensitive 3d face recognition using sparse representation. Dans *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2575–2582. IEEE, 2009. (Cited page 68.)

Liu, Shuai-Shi, Tian, Yan-Tao, et Li, Dong. New research advances of facial expression recognition. Dans *Machine Learning and Cybernetics, 2009 International Conference on*, volume 2, pages 1150–1155. IEEE, 2009. (Cited page 8.)

Maalej, Ahmed, Amor, Boulbaba Ben, Daoudi, Mohamed, Srivastava, Anuj, et Berretti, Stefano. Shape analysis of local facial patches for 3d facial expression recognition. *Pattern Recognition*, 44(8), p. 1581–1589, 2011. (Cited pages x, 38, 39 et 43.)

Mahersia, Hela et Hamrouni, Kamel. Using multiple steerable filters and

bayesian regularization for facial expression recognition. *Engineering Applications of Artificial Intelligence*, 38, p. 190–202, 2015. (Cited page 36.)

Mahoor, Mohammad H et Abdel-Mottaleb, Mohamed. Face recognition based on 3d ridge images obtained from range data. *Pattern Recognition*, 42(3), p. 445–451, 2009. (Cited pages 32, 67 et 68.)

Maiti, Somsukla, Sangwan, Dhiraj, et Raheja, Jagdish Lal. Expression-invariant 3d face recognition using k-svd method. Dans *International Conference on Applied Algorithms*, pages 266–276. Springer, 2014. (Cited page 29.)

Matuszewski, Bogdan J, Quan, Wei, Shark, Lik-Kwan, Mcloughlin, Alison S, Lightbody, Catherine E, Emsley, Hedley CA, et Watkins, Caroline L. Hi4d-adsip 3-d dynamic facial articulation database. *Image and Vision Computing*, 30(10), p. 713–727, 2012. (Cited page 104.)

Medioni, Gérard, Choi, Jongmoo, Kuo, Cheng-Hao, Choudhury, Anustup, Zhang, Li, et Fidaleo, Douglas. Non-cooperative persons identification at a distance with 3d face modeling. Dans *Biometrics : Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on*, pages 1–6. IEEE, 2007. (Cited page 16.)

Mehrabian, Albert. *Nonverbal communication*. Transaction Publishers, 1972. (Cited page 7.)

Mian, Ajmal, Bennamoun, Mohammed, et Owens, Robyn. Automatic 3d face detection, normalization and recognition. Dans *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 735–742. IEEE, 2006. (Cited page 21.)

Mian, Ajmal S, Bennamoun, Mohammed, et Owens, Robyn. Keypoint detection and local feature matching for textured 3d face recognition. *International Journal of Computer Vision*, 79(1), p. 1–12, 2008. (Cited pages 33 et 63.)

Ming, Yue. Robust regional bounding spherical descriptor for 3d face recognition and emotion analysis. *Image and Vision Computing*, 35, p. 14–22, 2015. (Cited page 29.)

Moakher, Maher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3), p. 735–747, 2005. (Cited page 52.)

Moeini, Ali, Moeini, Hossein, et Faez, Karim. Expression-invariant face recognition via 3d face reconstruction using gabor filter bank from a 2d single image. Dans *2014 22nd International Conference on Pattern Recognition (ICPR)*, pages 4708–4713. IEEE, 2014. (Cited page 29.)

Mohammadzade, Hoda et Hatzinakos, Dimitrios. Iterative closest normal point for 3d face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(2), p. 381–397, 2013. (Cited pages 6 et 33.)

Moreno, AB et Sanchez, A. Gavabdb : a 3d face database. Dans *Proc. 2nd COST275 Workshop on Biometrics on the Internet, Vigo (Spain)*, pages 75–80, 2004. (Cited pages 19 et 22.)

Mpiperis, Iordanis, Malassiotis, Sotiris, et Strintzis, Michael G. 3-d face recognition with the geodesic polar representation. *IEEE Transactions on Information Forensics and Security*, 2(3-2), p. 537–547, 2007. (Cited page 28.)

Mpiperis, Iordanis, Malassiotis, Sotiris, et Strintzis, Michael G. Bilinear models for 3-d face and facial expression recognition. *Information Forensics and Security, IEEE Transactions on*, 3(3), p. 498–511, 2008. (Cited pages x, 40, 41, 43, 92 et 93.)

Munkres, James. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial & Applied Mathematics*, 5(1), p. 32–38, 1957. (Cited pages 57 et 108.)

Ocegueda, Omar, Fang, Tianhong, Shah, Shishir K, et Kakadiaris, Ioannis A. 3d face discriminant analysis using gauss-markov posterior marginals. *IEEE transactions on pattern analysis and machine intelligence*, 35(3), p. 728–739, 2013. (Cited pages 6, 7 et 36.)

Pandzic, Igor S et Forchheimer, Robert. Mpeg-4 facial animation. *The standard, implementation and applications. Chichester, England : John Wiley&Sons*, 2002. (Cited pages 35 et 36.)

Pantic, Maja et Rothkrantz, Leon JM. Automatic analysis of facial expressions : The state of the art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12), p. 1424–1445, 2000. (Cited page 36.)

Passalis, Georgios, Perakis, Panagiotis, Theoharis, Theoharis, et Kakadiaris, Ioannis A. Using facial symmetry to handle pose variations in real-world 3d face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10), p. 1938–1951, 2011. (Cited page 31.)

Patil, Hemprasad, Kothari, Ashwin, et Bhurchandi, Kishor. 3-d face recognition : features, databases, algorithms and challenges. *Artificial Intelligence Review*, 44(3), p. 393–441, 2015. (Cited page 36.)

Pears, Nick, Liu, Yonghuai, et Bunting, Peter. *3D imaging, analysis and applications*, volume 3. Springer, 2012. (Cited page 17.)

Pennec, Xavier, Fillard, Pierre, et Ayache, Nicholas. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66 (1), p. 41–66, 2006. (Cited page 84.)

Pentland, Alex, Moghaddam, Baback, et Starner, Thad. View-based and modular eigenspaces for face recognition. Dans *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 84–91. IEEE, 1994. (Cited page 26.)

Perakis, P, Passalis, G, Theoharis, T, Toderici, G, et Kakadiaris, IA. Partial matching of interpose 3d facial data for face recognition. Dans *Biome-*

trics : Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on, pages 1–8. IEEE, 2009. (Cited page 31.)

Petrovska-Delacrétaz, Dijana, Lelandais, Sylvie, Colineau, Joseph, Chen, Liming, Dorizzi, Bernadette, Ardabilian, M, Krichen, E, Mellakh, M-A, Chaari, A, Guerfi, S, et al. The iv 2 multimodal biometric database (including iris, 2d, 3d, stereoscopic, and talking face data), and the iv 2-2007 evaluation campaign. Dans *Biometrics : Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pages 1–7. IEEE, 2008. (Cited pages 6 et 104.)

Phillips, P Jonathon, Flynn, Patrick J, Scruggs, Todd, Bowyer, Kevin W, Chang, Jin, Hoffman, Kevin, Marques, Joe, Min, Jaesik, et Worek, William. Overview of the face recognition grand challenge. Dans *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, volume 1, pages 947–954. IEEE, 2005. (Cited pages 19 et 22.)

Phillips, P Jonathon, Grother, Patrick, Micheals, Ross, Blackburn, Duane M, Tabassi, Elham, et Bone, Mike. Face recognition vendor test 2002. Dans *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, page 44. IEEE, 2003. (Cited pages 6 et 14.)

Plutchik, Robert. *Emotions and life : Perspectives from psychology, biology, and evolution.* American Psychological Association, 2003. (Cited page 8.)

Queirolo, Chaua C, Silva, Luciano, Bellon, Olga RP, et Segundo, Mauricio Pamplona. 3d face recognition using simulated annealing and the surface interpenetration measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(2), p. 206–219, 2010. (Cited pages 63 et 65.)

Ramanathan, Subramanian, Kassim, Ashraf, Venkatesh, YV, et Wah, Wu Sin. Human facial expression recognition using a 3d morphable mo-

del. Dans *Image Processing, 2006 IEEE International Conference on*, pages 661–664. IEEE, 2006. (Cited page 39.)

Rara, Ham M, Elhabian, Shireen Y, Ali, Asem M, Miller, Mike, Starr, Thomas L, et Farag, Aly A. Distant face recognition based on sparse-stereo reconstruction. Dans *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 4141–4144. IEEE, 2009. (Cited page 16.)

Ratyal, Naeem Iqbal, Taj, Imtiaz Ahmad, Bajwa, Usama Ijaz, et Sajid, Muhammad. 3d face recognition based on pose and expression invariant alignment. *Computers & Electrical Engineering*, 46, p. 241–255, 2015. (Cited pages 31 et 63.)

Rosato, Matthew, Chen, Xiaochen, et Yin, Lijun. Automatic registration of vertex correspondences for 3d facial expression analysis. Dans *Biometrics : Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pages 1–7. IEEE, 2008. (Cited page 41.)

Rudovic, Ognjen, Pantic, Maja, et Patras, Ioannis. Coupled gaussian processes for pose-invariant facial expression recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(6), p. 1357–1369, 2013. (Cited page 40.)

Russ, Trina, Boehnen, Chris, et Peters, Tanya. 3d face recognition using 3d alignment for pca. Dans *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1391–1398. IEEE, 2006. (Cited page 27.)

Salazar, Augusto, Wuhrer, Stefanie, Shu, Chang, et Prieto, Flavio. Fully automatic expression-invariant face correspondence. *Machine Vision and Applications*, 25(4), p. 859–879, 2014. (Cited page 32.)

Sandbach, Georgia, Zafeiriou, Stefanos, Pantic, Maja, et Yin, Lijun. Static and dynamic 3d facial expression recognition : A comprehensive survey. *Image and Vision Computing*, 30(10), p. 683–697, 2012. (Cited pages ix, 20, 36 et 81.)

Sanin, Andres, Sanderson, Conrad, Harandi, Mehrtash T, et Lovell, Brian C. Spatio-temporal covariance descriptors for action and gesture recognition. Dans *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 103–110. IEEE, 2013. (Cited page 105.)

Savran, Arman, Alyüz, Neşe, Dibeklioğlu, Hamdi, Çeliktutan, Oya, Gökberk, Berk, Sankur, Bülent, et Akarun, Lale. Bosphorus database for 3d face analysis. Dans *European Workshop on Biometrics and Identity Management*, pages 47–56. Springer, 2008. (Cited pages 20 et 22.)

Segundo, Mauricio P, Queirolo, Chaua, Bellon, Olga RP, et Silva, Luciano. Automatic 3d facial segmentation and landmark detection. Dans *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*, pages 431–436. IEEE, 2007. (Cited page 21.)

Seitz, Steven M, Curless, Brian, Diebel, James, Scharstein, Daniel, et Szeliski, Richard. A comparison and evaluation of multi-view stereo reconstruction algorithms. Dans *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 519–528. IEEE, 2006. (Cited page 15.)

Sha, Teng, Song, Mingli, Bu, Jiajun, Chen, Chun, et Tao, Dacheng. Feature level analysis for 3d facial expression recognition. *Neurocomputing*, 74 (12), p. 2135–2141, 2011. (Cited pages x, 37 et 39.)

Shan, Caifeng, Gong, Shaogang, et McOwan, Peter W. Facial expression recognition based on local binary patterns : A comprehensive study. *Image and Vision Computing*, 27(6), p. 803–816, 2009. (Cited pages 36 et 81.)

Shao, Jie, Gori, Ilaria, Wan, Shaohua, et Aggarwal, JK. 3d dynamic facial expression recognition using low-resolution videos. *Pattern Recognition Letters*, 65, p. 157–162, 2015. (Cited pages 40 et 81.)

Smeets, Dirk, Claes, Peter, Vandermeulen, Dirk, et Clement, John Gerald.

Objective 3d face recognition : Evolution, approaches and challenges. *Forensic science international*, 201(1), p. 125–132, 2010. (Cited page 7.)

Smeets, Dirk, Keustermans, Johannes, Vandermeulen, Dirk, et Suetens, Paul. meshsift : Local surface features for 3d face recognition under expression variations and partial data. *Computer Vision and Image Understanding*, 117(2), p. 158–169, 2013. (Cited pages x, 30 et 31.)

Soyel, Hamit et Demirel, Hasan. Facial expression recognition using 3d facial feature distances. Dans *Image Analysis and Recognition*, pages 831–838. Springer, 2007. (Cited page 37.)

Soyel, Hamit et Demirel, Hasan. 3d facial expression recognition with geometrically localized facial features. Dans *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*, pages 1–4. IEEE, 2008a. (Cited page 81.)

Soyel, Hamit et Demirel, Hasan. Facial expression recognition using 3d facial feature distances. Dans *Affective Computing*. InTech, 2008b. (Cited pages x, 37 et 39.)

Soyel, Hamit et Demirel, Hasan. Optimal feature selection for 3d facial expression recognition with geometrically localized facial features. Dans *Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control, 2009. ICSCCW 2009. Fifth International Conference on*, pages 1–4. IEEE, 2009. (Cited page 37.)

Soyel, Hamit et Demirel, Hasan. Optimal feature selection for 3d facial expression recognition using coarse-to-fine classification. *Turkish Journal of Electrical Engineering & Computer Sciences*, 18(6), p. 1031–1040, 2010. (Cited pages x, 37, 39, 43, 81 et 92.)

Spreeuwers, L. Fast and accurate 3d face recognition using registration to an intrinsic coordinate system and fusion of multiple region. *Proc of Int Journal of Computer Vision*, 93, p. 389–414, 2011. (Cited pages 63 et 65.)

Sra, Suvrit. A new metric on the manifold of kernel matrices with application to matrix geometric means. Dans *Advances in Neural Information Processing Systems*, pages 144–152, 2012. (Cited page 53.)

Srivastava, Ruchir et Roy, Sujoy. 3d facial expression recognition using residues. Dans *TENCON 2009-2009 IEEE Region 10 Conference*, pages 1–5. IEEE, 2009. (Cited pages 37 et 38.)

Sun, Yuehui. Expression invariant 3d face recognition based on gmds. Dans *2015 10th International Conference on Information, Communications and Signal Processing (ICICS)*, pages 1–5. IEEE, 2015. (Cited page 28.)

Tabia, Hedi et Laga, Hamid. Covariance-based descriptors for efficient 3d shape matching, retrieval, and classification. *Multimedia, IEEE Transactions on*, 17(9), p. 1591–1603, 2015. (Cited pages 82, 84 et 86.)

Tabia, Hedi, Laga, Hamid, Picard, David, et Gosselin, Philippe-Henri. Covariance descriptors for 3d shape matching and retrieval. Dans *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 4185–4192. IEEE, 2014. (Cited pages 47, 67, 68 et 82.)

Tang, Hao et Huang, Thomas S. 3d facial expression recognition based on automatically selected features. Dans *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2008. (Cited pages x, 37, 39 et 43.)

Tang, Hengliang, Yin, Baocai, Sun, Yanfeng, et Hu, Yongli. 3d face recognition using local binary patterns. *Signal Processing*, 93(8), p. 2190–2198, 2013. (Cited page 6.)

Tang, Yinhang, Li, Huibin, Sun, Xiang, Morvan, Jean-Marie, et Chen, Liming. Principal curvature measures estimation and application to 3d face recognition. *Journal of Mathematical Imaging and Vision*, pages 1–23, 2017. (Cited page 68.)

Tekguc, Umut, Soyel, Hamit, et Demirel, Hasan. Feature selection for person-independent 3d facial expression recognition using nsga-ii. Dans

Computer and Information Sciences, 2009. ISCIS 2009. 24th International Symposium on, pages 35–38. IEEE, 2009. (Cited page 37.)

ter Haar, Frank B et Veltkamp, Remco C. Expression modeling for expression-invariant face recognition. *Computers & Graphics*, 34(3), p. 231–241, 2010. (Cited page 27.)

Tian, Y, Kanade, T, et Cohn, J. Handbook of face recognition, chapter facial expression analysis, 2003. (Cited pages 36 et 81.)

Tomkins, Silvan. *Affect imagery consciousness : Volume I : The positive affects*, volume 1. Springer publishing company, 1962. (Cited page 8.)

Tuzel, Oncel, Porikli, Fatih, et Meer, Peter. Region covariance : A fast descriptor for detection and classification. Dans *Computer Vision–ECCV 2006*, pages 589–600. Springer, 2006. (Cited page 47.)

Tuzel, Oncel, Porikli, Fatih, et Meer, Peter. Pedestrian detection via classification on riemannian manifolds. *IEEE transactions on pattern analysis and machine intelligence*, 30(10), p. 1713–1727, 2008. (Cited pages 47 et 84.)

Vemulapalli, Raviteja, Pillai, Jaishanker K, et Chellappa, Rama. Kernel learning for extrinsic classification of manifold features. Dans *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1782–1789. IEEE, 2013. (Cited pages xi, 85 et 86.)

Villani, Cédric. *Optimal transport : old and new*, volume 338. Springer, 2008. (Cited page 56.)

Vretos, Nicholas, Nikolaidis, Nikos, et Pitas, Ioannis. 3d facial expression recognition using zernike moments on depth images. Dans *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 773–776. IEEE, 2011. (Cited pages 40, 43, 96 et 97.)

Wallraven, Christian, Caputo, Barbara, et Graf, Arnulf. Recognition with local features : the kernel recipe. Dans *Computer Vision, 2003. Proceedings.*

Ninth IEEE International Conference on, pages 257–264. IEEE, 2003. (Cited pages 90 et 104.)

Wang, Jun, Yin, Lijun, Wei, Xiaozhou, et Sun, Yi. 3d facial expression recognition based on primitive surface feature distribution. Dans *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1399–1406. IEEE, 2006. (Cited pages 38 et 81.)

Wang, Ruiping, Guo, Huimin, Davis, Larry S, et Dai, Qionghai. Covariance discriminative learning : A natural and efficient approach to image set classification. Dans *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2496–2503. IEEE, 2012. (Cited pages xi, 47 et 84.)

Wang, Shu-Fan et Lai, Shang-Hong. Reconstructing 3d face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold. *IEEE transactions on pattern analysis and machine intelligence*, 33(10), p. 2115–2121, 2011. (Cited page 15.)

Wang, Yiding, Meng, Meng, et Zhen, Qingkai. Learning encoded facial curvature information for 3d facial emotion recognition. Dans *Image and Graphics (ICIG), 2013 Seventh International Conference on*, pages 529–532. IEEE, 2013. (Cited pages 43, 96 et 97.)

Wang, Yueming, Liu, Jianzhuang, et Tang, Xiaoou. Robust 3d face recognition by local shape difference boosting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(10), p. 1858–1870, 2010. (Cited pages 63 et 65.)

Woodham, Robert J. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1), p. 191139–191139, 1980. (Cited page 15.)

Xu, Chenghua, Tan, Tieniu, Wang, Yunhong, et Quan, Long. Combining local features for robust nose location in 3d facial data. *Pattern Recognition Letters*, 27(13), p. 1487–1494, 2006. (Cited page 21.)

Xu, Chenghua, Wang, Yunhong, Tan, Tieniu, et Quan, Long. A new attempt to face recognition using 3d eigenfaces. Dans *Proceedings of the Asian Conference on Computer Vision*, volume 2, pages 884–889. Citeseer, 2004. (Cited page 18.)

Yagou, Hirokazu, Ohtake, Yutaka, et Belyaev, Alexander. Mesh smoothing via mean and median filtering applied to face normals. Dans *Geometric Modeling and Processing, 2002. Proceedings*, pages 124–131. IEEE, 2002. (Cited page 23.)

Yin, Lijun, Chen, Xiaochen, Sun, Yi, Worm, Tony, et Reale, Michael. A high-resolution 3d dynamic facial expression database. Dans *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, pages 1–6. IEEE, 2008. (Cited pages 15 et 104.)

Yin, Lijun, Wei, Xiaozhou, Sun, Yi, Wang, Jun, et Rosato, Matthew J. A 3d facial expression database for facial behavior research. Dans *Automatic face and gesture recognition, 2006. FGR 2006. 7th international conference on*, pages 211–216. IEEE, 2006. (Cited pages 19, 22 et 36.)

Yu, Xun, Gao, Yongsheng, et Zhou, Jun. Sparse 3d directional vertices vs continuous 3d curves : Efficient 3d surface matching and its application for single model face recognition. *Pattern Recognition*, 2016. (Cited page 32.)

Yun, Yixiao et Gu, Irene Yu-Hua. Exploiting riemannian manifolds for daily activity classification in video towards health care. Dans *e-Health Networking, Applications and Services (Healthcom), 2016 IEEE 18th International Conference on*, pages 1–6. IEEE, 2016. (Cited page 84.)

Yun, Yixiao, Gu, Irene Yu-Hua, et Aghajan, Hamid. Riemannian manifold-based support vector machine for human activity classification in

images. Dans *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 3466–3469. IEEE, 2013. (Cited page 84.)

Zeng, Zhihong, Pantic, Maja, Roisman, Glenn I, et Huang, Thomas S. A survey of affect recognition methods : Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1), p. 39–58, 2009. (Cited pages 36 et 81.)

Zhang, Lin, Ding, Zhixuan, Li, Hongyu, Shen, Ying, et Lu, Jianwei. 3d face recognition based on multiple keypoint descriptors and sparse representation. *PloS one*, 9(6), p. e100120, 2014. (Cited page 32.)

Zhang, Zhengyou. Iterative point matching for registration of free-form curves and surfaces. *International journal of computer vision*, 13(2), p. 119–152, 1994. (Cited pages 23 et 47.)

Zhang, Zhengyou, Liu, Zicheng, Adler, Dennis, Cohen, Michael F, Hanson, Erik, et Shan, Ying. Robust and rapid generation of animated faces from video images : A model-based modeling approach. *International Journal of Computer Vision*, 58(2), p. 93–119, 2004. (Cited page 16.)

Zhao, Xi, Dellandrea, Emmanuel, Chen, Liming, et Kakadiaris, Ioannis A. Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(5), p. 1417–1428, 2011. (Cited page 6.)

Ce document a été préparé à l'aide de l'éditeur de texte GNU Emacs et du logiciel de composition typographique LaTeX 2ε.