UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXIII CICLO DEL DOTTORATO DI RICERCA IN

FISICA

EXPERIMENTAL CHARACTERIZATION OF SURFACES AND METAL-POLYMER HETEROSTACKS OF OPTICAL SYSTEMS FOR THE AUTOMOTIVE SECTOR

Settore scientifico-disciplinare: **FIS/03**

DOTTORANDO
TOMMASO FONTANOT

COORDINATORE
PROF. FRANCESCO LONGO

SUPERVISORE DI TESI
PROF. ERIK VESSELLI

CO-SUPERVISORE DI TESI
DOTT. SARA PARONI

ANNO ACCADEMICO 2019/2020

Abstract

Sfruttando le metodologie di indagine tipiche della fisica delle superfici e della scienza dei materiali, è stato possibile caratterizzare in dettaglio alcuni dei materiali utilizzati nel settore automotive. In particolare, l'attenzione è stata focalizzata sullo studio delle proprietà ottiche e strutturali di vari materiali, superfici ed interfacce di componenti polimeriche e metalliche da utilizzare per la produzione di fanaleria.

Durante il percorso di dottorato è stata realizzata e validata una macchina prototipale per la scansione automatica di trasparenti con difettosità, basata su un sistema di *computer vision*. Gli algoritmi del sistema d'indagine, sviluppati in modo tale da evitare il controllo umano e tutte le problematiche ad esso correlate, sono stati tarati con precisione al fine di aumentare le prestazioni e l'affidabilità del prototipo. Differenti modalità di detezione sono state sfruttate al fine di riconoscere e classificare una grande varietà di difetti superficiali e dei materiali stessi.

Oltre ai trasparenti, l'attenzione è stata focalizzata anche sui riflettori. L'obiettivo di un riflettore è quello di riflettere la luce generata da una sorgente e fornire al fanale l'aspetto estetico desiderato. Un generico riflettore è composto da un substrato plastico sopra il quale vengono depositati un layer metallico e uno protettivo: in questa tesi di dottorato ne sono state studiate dettagliatamente sia le proprietà strutturali (spessore dei film sottili, rugosità superficiali, …) che quelle ottiche (riflettanza speculare e totale, dipendenza spettrale del fascio riflesso, …). Substrati plastici planari (realizzati in policarbonato -PC- e acrilonitrile butadiene stirene -ABS-) sono stati metallizzati tramite un processo, del tutto identico a quello utilizzato per la produzione dei riflettori, consistente in una deposizione fisica da vapore (PVD) di alluminio (Al) in vuoto (10^{-4} mbar) con la successiva deposizione di un protettivo a base di esametildisilossano (HMDSO) per prevenire danni chimici e fisici al film metallico. Le proprietà strutturali di questi multistrati metallo-polimero sono state caratterizzate tramite microscopio a forza atomica (AFM), determinando la rugosità superficiale e lo spessore dei diversi layer. Le proprietà ottiche di campioni preparati utilizzando differenti trattamenti di metallizzazione sono state confrontate tramite funzione di distribuzione bidirezionale di riflettanza (BRDF). Sfruttando poi il sistema di riferimento di Harvey Shack e rappresentando i dati per mezzo del modello ABg, è stato possibile estrarre dei parametri fondamentali per meglio rappresentare i diversi trattamenti tramite i software di simulazione dedicati. I dati ottici, supportati da misure di profilometria, hanno evidenziato un contributo importante dovuto alla deformazione superficiale del substrato generata dal processo di stampaggio.

Sono state inoltre trovate le tecniche più adatte per investigare i difetti che si formano sulla superficie metallizzata e all'interfaccia metallo-polimero dei riflettori. Il microscopio elettronico e la tomografia a raggi X con luce di sincrotrone sono stati utilizzati con un approccio combinato, ottenendo dei risultati interessanti che chiarificano la natura e l'origine dei difetti. Sono stati osservati, infatti, problemi di adesione, problemi relativi allo stampaggio, contaminazioni cristalline e difetti strutturali.

Infine, è stata progettata e realizzata una macchina prototipale unica e innovativa. Essa ha un doppio scopo: identificare e classificare autonomamente i difetti presenti sulle superfici dei riflettori tramite riconoscimento visivo e verificare che il layer di alluminio abbia una riflettanza adeguata. Sfruttando il modello di apprendimento supervisionato delle macchine di supporto vettoriale (SVM), in seguito ad un preciso periodo di training, il prototipo è in grado

di riconoscere autonomamente i difetti e classificarli. Il ruolo del modello SVM è fondamentale perché permette di incrementare la qualità del riconoscimento all'aumentare dei componenti scansionati.

Durante questo percorso della durata di tre anni, è stato possibile ottenere maggiori informazioni sulla natura e sull'origine dei difetti presenti sui riflettori. Lo studio riguardante la caratterizzazione delle proprietà ottico-strutturali del multistrato eterogeneo metallo-polimero ha condotto ad una più precisa e maggiormente realistica rappresentazione delle superfici dei riflettori negli ambienti di simulazione. Le due macchine prototipali permettono, infine, una più economica, precisa ed affidabile alternativa per l'ispezione dei componenti, garantendo un'informazione in tempo reale essenziale al fine di monitorare la produzione e scovare potenziali problemi.

Abstract

The main target of the project consisted in applying the thorough experimental methodology typical of surface physics and materials science to the characterization of materials to be exploited in the automotive lighting sector. Several aspects were considered, thus covering a variety of surface, materials, optical, and physical properties of polymer and metal-polymer reflectors and lenses.

The development of a concept machine with the aim of automatically scanning defective outer lenses based on a computer vision system has also been validated. The algorithms of the inspection system, developed in order to avoid the human-eye control, have been fine-tuned to increase the performance and the reliability of the prototype machine. Different detection methods have been investigated and exploited, in order to recognize and classify a wide variety of surface and materials defects.

Following the experience on the lenses' inspection, the attention has been focused on reflecting elements. A reflector is one of the many components of rear automotive lamps. Its purpose is both to reflect light generated by a source and to provide a selected aesthetic aspect. It is composed by metallic and protective thin films deposited over plastic substrates: in particular, the structural (thickness of the metal layer, surface roughness, composition…) and optical (specular and total reflectance, spectral dependency of the scattered beam …) properties have been accurately studied. Flat plastic substrates (made of polycarbonate -PC- and acrylonitrile butadiene styrene -ABS-) have been metallized in vacuo (10^{-4} mbar) through Physical Vapor Deposition (PVD) of aluminum with the subsequent deposition of a protector made of hexamethyldisiloxane (HMDSO) to prevent both chemical and mechanical damaging. The preparation recipe of the samples was the same of the actual automotive reflector surfaces. The structural properties of metal-polymer heterostacks have been characterized by means of Atomic Force Microscopy (AFM), determining the surface roughness and the thickness of the diverse layers. Different metallization treatments were compared in a thorough characterization of the optical properties by measuring the spectral Bidirectional Reflectance Distribution Function (BRDF). Switching to the Harvey Shack reference system and fitting the BRDF data with the ABg model, it has been possible to extract important quantitative parameters that can be inserted in a simulation software to predict and model the reflector's properties and behavior. The optical data, corroborated with profilometry measurements, evidenced an important contribution from the surface deformations induced by the mold production process.

The most proper techniques to investigate defects forming at the metallized surfaces and interfaces of reflectors have been identified. By means of a combined approach exploiting Scanning Electron Microscopy (SEM) and synchrotron radiation based X-ray computed tomography (SR-CT) interesting results were obtained, yielding insight into the nature of the defects both at the surface and at the metal-polymer buried interfaces. Adhesion issues, press printing process-related problems, crystalline contaminations, and 3D structural defects have been revealed.

Finally, an innovative and unique prototype machine has been designed and commissioned. The machine has a double aim: on one side, it automatically recognizes the defects arising on the reflectors' surfaces by visual scanning, while on the other side it checks the reflectance of the metallized layer exploiting optical processes. In fact, by means of specifically designed algorithms, a standard camera is able to determine if the aluminum thin

film has the correct reflectance properties, i.e. if the reflector can be considered compliant or not. Furthermore, exploiting Support Vector Machines (SVM) supervised learning models the system, after a deep and precise training, can recognize autonomously the defects arising on the reflectors and classify them. The role of the SVM models is crucial, because the more scans are performed, the better the recognition of defects. Indeed, while scanning the machine learns how to better interpret the defects and therefore automatically increases its accuracy.

During this three-years-long project, detailed information regarding the nature and origin of materials defects has been obtained. Moreover, the thorough studies involving the structural and optical properties of metal-polymer heterostacks led to a more precise and realistic representation of the reflectors' surfaces in simulation environments. Finally, the two prototype machines allow for a faster, cheaper, more objective and reliable inspection of the components, providing real time information essential to monitor the production chain and detect potential problems.

Table of contents

1. Introduction 1
1.1 Marelli Automotive Lighting Italy – University of Trieste: purpose of the collaboration 1
1.2 Rear Lamps: general information 6
1.2.1 Light sources 7
1.2.2 Reflectors 7
1.2.3 Inner Lenses 8
1.2.4 Light guides 9
1.2.5 Outer lenses 10
1.2.6 Materials 10
1.3 The physics and the optical design of rear lamps: photometry, colorimetry and luminance requirements 12
2. Production of lamp optics: technical aspects and issues 17
2.1 Injection molding 17
2.1.1 Parameters 18
2.1.2 Defects 20
2.2 Coatings deposition 23
2.2.1 Metallization setup 23
2.2.2 Metallization process 25
2.3 Photolithography 27
3. Introduction to the characterization methods 33
3.1 Scanning Electron Microscopy 33
3.1.1 Electron-sample interactions 33
3.1.2 Backscattered Electrons 34
3.1.3 Secondary Electrons 36
3.1.4 X-rays 37
3.1.5 Instrumentation 39
3.1.6 Experimental Setup 40
3.2 Atomic Force Microscopy 41
3.2.1 Cantilever and tip 41
3.2.2 Laser and photodetector 43
3.2.3 Piezoelectric scanner 45
3.2.4 Feedback loop 45
3.2.5 Operation modes 46
3.2.6 Experimental setup 51
3.3 Synchrotron-radiation X-Ray Computed Tomography 52
3.3.1 Coherence 52
3.3.2 X-ray imaging 52
3.3.3 Propagation-based phase-contrast imaging 55
3.3.4 Computed Tomography 56
3.3.5 Physical background 58
3.3.6 Artifacts 62
3.3.7 Experimental setup 63
3.4 Bidirectional Reflectance Distribution Function 65
3.4.1 Radiance, irradiance, reflectance 65
3.4.2 Bidirectional Reflectance Distribution Function definition and properties 67
3.4.3 Experimental setup 70
4. Computer-based inspection and characterization approaches 73
4.1 Introduction to machine vision 73
4.1.1 Intensity transformation and spatial filtering 73
4.1.2 Sharpening spatial filters 77

4.1.3 Morphological image processing 80
4.1.4 Image segmentation 84
4.2 Introduction to Support Vector Machine 90
4.2.1 How does SVM work? 90
4.2.2 The Linear SVM 91
4.2.3 The nonlinear SVM 98
4.2.4 Regression by support vector machine 100
4.2.5 Case study 106
5. Automated characterization of outer lenses 107
5.1 Introduction 107
5.2 Layout and simulations 109
5.3 Working principle 114
5.3.1 Transmission region 114
5.3.2 Refraction region 121
5.3.3 Pattern region 124
5.4 Preliminary results 126
5.5 Discussion 128
6. Characterization of reflectors 129
6.1 Surface defects 129
6.1.1 Protrusions (PRT) 130
6.1.2 Depressions (DPR) 134
6.1.3 White spots (WS) 134
6.1.4 Black spots (BS) 137
6.1.5 Scratches (SCR) 137
6.1.6 Discussion 138
6.2 Surface morphology and optical properties 140
6.2.1 AFM analysis 141
6.2.2 BRDF analysis 145
6.2.3 Data evaluation and modelling 150
6.2.4 Discussion 155
6.3 Automatic quality check for reflectors 156
6.3.1 Why optical properties? 156
6.3.2 Setup of the machine 160
6.3.3 Working principle 164
6.3.4 Preliminary results 175
6.3.5 Discussion 176
7. Conclusions and perspectives 177
8. Scientific production 179
Publications 179
Conferences 179
9. Bibliography 181
Acknowledgements 189

1. Introduction

1.1 Marelli Automotive Lighting Italy – University of Trieste: purpose of the collaboration

The PhD project presented in this thesis has been developed within the framework of an extensive, dedicated scientific collaboration between a company (Marelli Automotive Lighting Italy) and a university (Università degli Studi di Trieste). Marelli is one of the world's leading global independent suppliers to the automotive sector. It was born in 2019 as a unification of the two well-known international players "Magneti Marelli" and "Calsonic Kansei". Marelli Automotive Lighting is one of Marelli's divisions and is fully committed to the design, production, and distribution of automotive lamps for the most important original equipment manufacturers (OEMs) all over the world. This thesis work has been performed in collaboration with Marelli Automotive Lighting Italy, plant of Tolmezzo (UD), which is the competence center for the design and manufacturing of rear lamps (examples represented in *Fig. 1.1.1*). The Tolmezzo plant also gives support to other rear lamps productive plants all over the world.

Fig. 1.1.1: examples of rear lamps designed and manufactured in the Marelli Automotive Lighting Italy plant of Tolmezzo (Udine).

The aim of this doctoral project was to provide solutions to typical industrial problems that are material-related, thus exploiting material science and surface physics approaches, based on thorough experimental methodologies. Particularly, during the three-year-itinerary, I have focused my attention in four main topics:

- the design and validation of a prototype machine to perform a real time automated quality control of outer lenses;
- the characterization of the optical and morphological properties of the highly reflective surfaces of reflectors;
- the study to determine the origin and nature of defects at the metal polymer heterostacks associated with the reflectors' manufacturing procedure;
- the design and validation of a unique and innovative prototype machine devoted to the quality and functional check of the reflectors.

An outer lens is the most external component of a lamp. Its function, as will be discussed in detail in *Section 1.2*, is to protect all the fragile components contained inside a rear lamp and to provide the vehicle the aesthetic aspect desired by the customer: for these reasons, its manufacturing quality must be guaranteed. The prototype machine to detect defects arising on the outer lenses has a fundamental role to prevent human errors during the quality check process. In fact, these defects are recognizable by human eye, but some issues can be tough obstacles in obtaining the right selection of the uncompliant pieces. The surface of an outer lens needs an extreme precision manufacturing procedure and the absence of defects is essential for the quality of the final product. The aim of the work involves the design and commissioning of a setup to extract and analyze information about the flaws present in an outer lens, exploiting different image processing techniques depending on the nature of the defects and, therefore, to support the production teams by providing a reliable tool to check the quality of the products. As benchmark samples to test this prototype we adopted the outer lenses of the Mercedes E-class car.

Reflectors are the other category of lamps' components investigated during the PhD project. Their primarily aim is to supply the light and optical properties required by the regulations and, secondarily, to contribute to the aesthetic aspect of the lamp, as it will be described in detail later in this chapter. A thorough optical and structural characterization of reflector surface materials has been performed to unravel the influence of substrate chemistry and thickness, aluminum and protective layer thickness, and surface roughness on the effective optical reflectance properties. Bidirectional Reflectance Distribution Function measurements, corroborated by surface morphological information obtained by Atomic Force Microscopy, correlate reflectance characteristics with the root mean square surface roughness, putting in evidence the role of the substrate and of the thin films' morphology.

Moreover, to have an insight on the origin and nature of the defects arising during the manufacturing process, a combined microscopy-, spectroscopy-, and tomography-based approach has been exploited to investigate morphological and compositional properties of both surfaces and buried interfaces. Synchrotron radiation X-ray Computed Tomography and Scanning Electron Microscopy, coupled with Energy Dispersive X-ray Spectroscopy indicate that both chemical contaminations, introduced in the fabrication process, and growth conditions strongly affect the presence of defects.

The design and validation of a prototype machine to check the aesthetic aspect and the functional optical properties of the reflectors follows directly from all the other three topics described above. The knowledge acquired during the validation of the prototype to scan the outer lenses, together to the information gained through the measurements performed to determine the optical and structural properties of the thin films deposited over the polymer-based substrates, played an important role in designing the machine. The prototype can efficiently detect defects arising at reflectors' surfaces ascribable to both the manufacturing and metallization process. Moreover, this novel machine gives the capability to perform a real time check of the optical properties of the scanned reflectors, leading the possibility to provide a quality feedback and to tune the metallization process parameters in case of insufficient performances.

The research topics previously introduced play a fundamental role in the innovative, trailblazing, and state-of-the-art route denominated Industry 4.0.[1,2] This defines an industrial automation trend consisting in a large-scale machine-to-machine communication (M2M) and Internet of Things (IoT) deployments to provide a real-time response to changings imposed by

shifting demands, unexpected faults or stock levels, improving the communication and increasing the automation of the production machines.[3–5] Machines are now able to perform self-monitoring and autonomous maintenance, giving the possibility to execute a real time data analysis and production parameters' tuning to satisfy the market requests without the need of human operators.[6–8] Smart factories are connected entities, able to interact with one another and adjust their performance, being every aspect of every process logged, assessable, and analyzable at any place and time.[9] These features are crucial to increase the efficiency and give the possibility to satisfy the more complex and detailed customers' needs.[10–12] From the production point of view, machine vision and machine learning are fundamental technologies to guide the transition from Industry 3.0 to Industry 4.0.[13] Machine vision consists of a set of technologies and methods to extract precise information from an image on an automated basis, while machine learning represents a set of algorithms which improve the automatically thought experience of the computer.[14–16] The technological improvement is underlined, for example, by the different quality inspection methods raised in the last years. In the past, in fact, monitoring the quality of products was a human task. The operator had to examine thoroughly each piece and decide if it was compliant or not. The agent, even if well trained and performing his/her work with excellent results, as a human being suffers from fatigue and illnesses, which can influence the judgment. Moreover, non-constant ambient lighting and environmental conditions could have led to a quality check failure, yielding both false positive and false negative results. In addition, depending on the typology and dimensions of the component, the analysis could have taken a considerable amount of time, thus slowing all the production chain. The automatic testing of components and processes is born to avoid the abovementioned fatalities. Machine-based inspection permits to isolate the analysis from possible human interferences. Computer vision and machine learning systems can provide a fast and reliable analysis, and nowadays they are exploited in multiple and various branches: in fact, medical,[17,18] agricultural,[19–21] manufacturing,[22] safety,[23] social,[24] and industrial fields are just some example of implementation of these powerful techniques.[25–27] The automotive lighting industry is not an exception in this constant upgrading movement: new technologies are exploited, and novel inspection methods are required to monitor different steps of the production chain. Particularly in this field, in order to fulfill the market requests asking for progressively new and more performing components, the need of an automated analysis is strongly rising, to maintain high productive standards and lowering the inspection time.[28,29]

The new tools provided by the Industry 4.0 technological development offer a faster and more precise way to design novel products. In order to decrease the time and raw material wastes, computer-based simulation has become the most exploited way to conceive new solutions to satisfy the market request.[30–32] In fact, concerning the automotive lighting sector, before prototyping a new lamp, several studies consisting on different kinds of simulations are performed: optical, electronic, and validation simulations are just some examples of the high amount of examinations that must be considered before validating the first prototype of a rear lamp. In addition, process-related simulations are also performed, thus determining the most proper process parameters in the molding, film deposition, and welding processes. This PhD project has been carried out working in close contact with the optical engineers, which perform optical simulations to design new prototype lamps. For the optical engineers, the difficulties lay in correctly modelling the components to successfully simulate the path of the light beam from the source to the final target. In the past, the desire of predicting the propagation of electromagnetic (e.m.) waves traveling in space has motivated the research of theoretical techniques to model such phenomena.[33] In the specific case, in this thesis only the e.m. waves

with energies mainly in the infrared and visible ranges of the e.m. spectrum will be treated. These techniques must consider the kinematics of e.m. waves along with their associated propagation and reflection properties.[34] Correct light simulations, together with realistic surface texture representations, add a lot more of information to a picture. In fact, most applications of computer graphics strongly depend on the attainable degree of realism.[35,36] The correct simulation of various lighting effects is the most important aspect of true photographic realism, especially for the design of a new lamp, which has to comply with some important optical requirements imposed by severe regulations. A simulation technique that satisfies these requirements is the ray-tracing. In its primordial form, a ray is traced for each pixel of the image from the viewpoint into the 3D scene to calculate its first intersection with an object. If the object is reflecting or refracting, an appropriate ray is determined by the physical laws of reflection or refraction. These new rays are treated analogously to the previous ones, following their propagation in space and modelling their interaction with reflective or refractive surfaces and interfaces. The ray-tracing simulation approximates closely the human vision perception process and therefore can produce highly realistic outputs. Moreover, it can easily handle complicated lighting effects, shading, and shadows. In the ages, ray-tracing has been upgraded in a wide variety of different methods as stochastic (a Monte-Carlo or Metropolis-based)[37–40] and deterministic ray-tracing.[41] Also, it is possible to define a backward ray-tracing (usually called back-tracing), when rays are emitted from the detector, and a forward ray-tracing, when rays are emitted from a light source and then reach the detector following the correct path. Although in nature the eyes do not emit rays to sense the environment, it does not mean that only forward ray-tracing is physically correct. In fact, all kinds of ray-tracing are only mathematical means to solve the self-consistent “global illumination problem", and are in a sense similar to integral equations with stochastic trajectories. In order to perform a correct ray-tracing simulation, all the optical and structural properties of the different materials employed must be well defined, in order to better describe the diverse light scattering properties. As a matter of fact, if the totality of the attributes is not correctly classified, the simulation could lead to a wrong representation of the desired object. Before starting with the lamp’s chain production, each component is designed exploiting a CAD environment and the lighting process is simulated by means of ray-tracing. It is therefore fundamental to thoroughly classify all the optical and structural properties of each component, in order to perform a correct and reliable simulation.

In this chapter (*Chapter 1*) the general information regarding the rear lamps are highlighted, presenting their main components. Moreover, the optical quantities and parameter involved in the design and validation of a rear lamp will be presented.

In *Chapter 2* the injection molding and the metallization processes will be described. In addition, the photolithography (fundamental to perform thickness measurements) will be introduced.

In *Chapter 3* the experimental technique and the methods exploited to study the samples will be revealed, while in *Chapter 4* the basic tools to understand computer vision and machine learning will be supplied.

Chapters 5 and *6* are the heart of the thesis and inside them the experimental results obtained during the PhD project are contained. *Chapter 5* will be focused on the commissioning of a prototype machine for the automatic scan of outer lenses, while in *Chapter 6* all the measurements performed to characterize the optical and structural properties of the reflectors together to the studies on the defects arising on them will be presented. Finally, *Chapter 6* will

illustrate the design and validation of a unique and innovative prototype machine to scan reflectors to detect defects and check their reflective properties.

The conclusions will be drawn in *Chapter 7*.

1.2 Rear Lamps: general information

Rear lamps are safety devices, in fact they are made to report to other drivers the vehicle's location on the road and its future moves (stop, turn, etc.). Beside the functional aim, in the last years rear lamps have been awarded with another purpose: provide the automobile the aesthetical lighting signature. In fact, as it is perceivable from *Fig. 1.2.1*, the rear lamp contributes as style identifier: it can be noticed how a sport car requires an aggressive shape, while, for example, an economy car needs a more sober style.

Fig. 1.2.1: example of different rear lamps' styles.

An exploded view, showing the main pieces inside a rear lamp, is represented in *Fig. 1.2.2*. Clearly, high technology is needed to design, manufacture and assemble each component involved in the lamp.

Fig. 1.2.2: exploded view of a rear lamp, showing all its internal components.

From *Fig. 1.2.3*, a clear understanding regarding a rear lamp working principle can be guesses. In *Fig. 1.2.3a* the section of a rear lamp is represented, while a ray-tracing view of the

same section is depicted in *Fig. 1.2.3b*. In brief, the LEDs contained in the PCB shine light on the reflector. The reflected light then crosses the inner lens and the outer lens, being therefore emitted by the lamp.

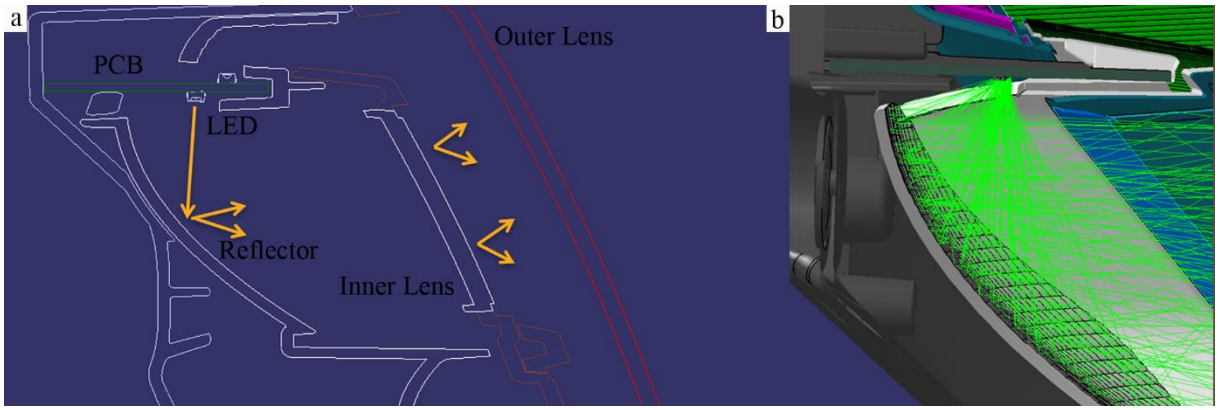

Figure 1.2.3: schematic crosscut section of rear lamp (a), and ray-tracing view of the same crosscut section (b).

1.2.1 Light sources

The most common light sources for rear lamps are filament bulbs, LEDs and, to a smaller fraction, OLEDs. Filament bulbs are the old, conventional light sources, which are still exploited for economy cars. They cost less with respect to the LEDs and OLEDs, but their service life is shorter and the power consumption higher. Moreover, their volume size makes it difficult to design complex lamp geometries and their optical properties are limited. On the contrary, LEDs last longer, their vibration resistance is better than bulbs' one and their packaging is shallower. LEDs can be placed in many different positions on the printed circuit board (PCB) allowing to tune the geometry of the light source itself and creating lamps with diverse shapes. Unfortunately, LEDs require greater attention because the emitted luminous flux is extremely sensitive to heat due to the semiconductor junction temperature dependence. This means that more strict validation simulations (thermal and stability tests) must be performed to be sure the light source is working in the right condition. The OLEDs are the newest light sources which can guarantee a better lighting homogeneity and solve the temperature-related problems. However, they are more sensitive to vibrations and they are more expensive than the LEDs.

1.2.2 Reflectors

Different kinds of reflectors can be exploited to design a lamp. The typical reflector is the so-called "parabolic metallized reflector", consisting of a plastic (PC or ABS) substrate metallized with aluminum (*Fig. 1.2.4*). "Parabolic" means that the reflecting surface has a paraboloid shape and the light source is positioned in its focus to better shine light. Sometimes it is not possible, due to volume constrains imposed by the design, to exploit parabolic reflectors, but in these cases a "freeform surface reflector" can be used. This type of reflector is a compromise to collect most of the light emitted by the source without the obstacle of the paraboloid volume. Even those reflectors are metallized in order to increase the reflectivity of the plastics. Usually, to better control and focus light in specific directions, optics with precise

geometric shape (pillow, cylindric, prims) are engraved/molded on the reflectors. The metallization process will be explained in detail in *Chapter 2*. Last types of reflectors are the "white reflectors", characterized by white non-metallized plastic. In this case, the white surface helps spreading light following a Lambertian distribution. They are mainly used when the goal is to produce a high luminance homogeneity level from a large viewing angle field, despite a lower efficiency with respect to the metallized ones.

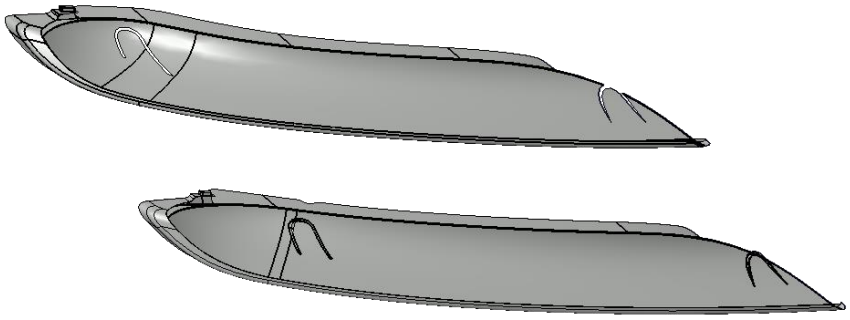

Fig. 1.2.4: example of reflector.

1.2.3 Inner Lenses

There can be multiple typologies of inner lenses: Fresnel lenses (*Fig. 1.2.5a*), lenses with standard optics, and lenses with micro optics (*Fig. 1.2.5b*). They are all produced using semi-transparent PC or PMMA, leaving the possibility for the light to pass through the medium. Fresnel lens is a semi-transparent planar element with prisms to redirect the light emitted by the source to the normal direction of the lens. Principles of refraction and total reflection are used to design these lenses. Lenses with standard and micro-optics are semi-transparent elements with an array of local calculated deviation from the original basic surfaces, which is used to spread the incident light within a desired angle, based on the refraction principle. When the size of the array is greater than 1 mm the optics are called "standard", while for lower size they are named "micro" optics. Particular typologies of inner lenses are the diffusive elements, which are intended for spreading the light by means of controlled roughness pattern on the component surface or volume scattering, obtaining satisfactory luminance homogeneity, but reducing the intensity efficiency.

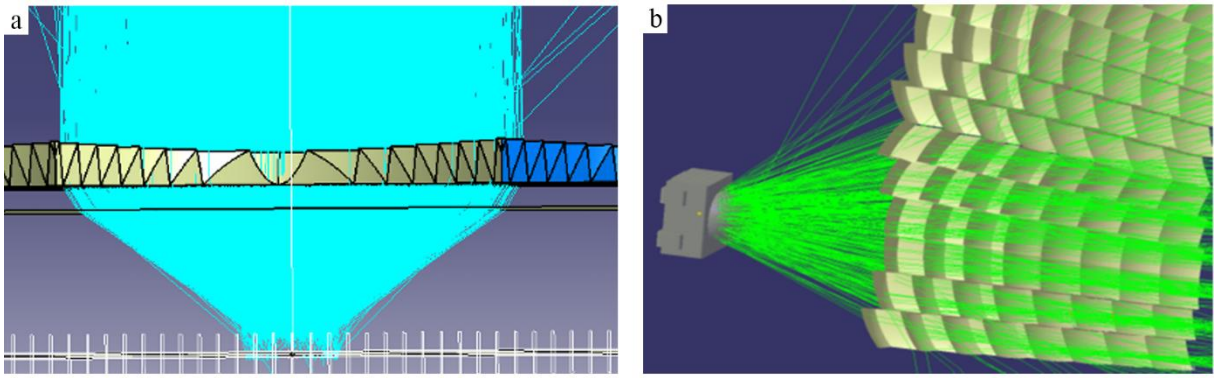

Fig. 1.2.5: ray-tracing simulations to demonstrate the working principle of Fresnel lens (a) and micro optics (b).

1.2.4 Light guides

Light guides (example shown in *Fig. 1.2.6*) are tubular semi-transparent elements, made by PC or PMMA, whose operating principle is based on Total Internal Reflection (TIR).

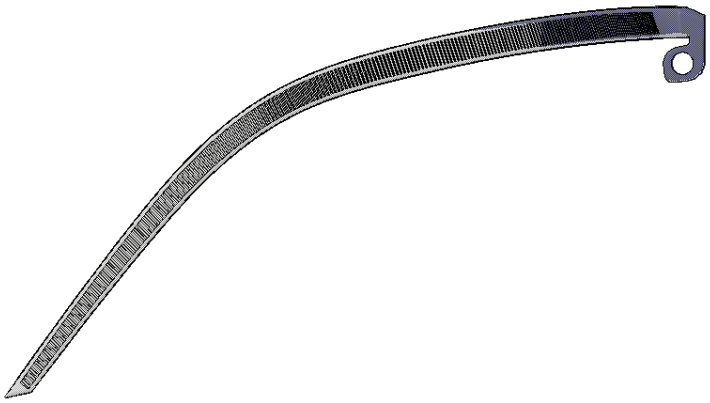

Fig. 1.2.6: example of light guide.

Their function is to distribute light coming from a punctual source in such a way that it can be emitted by an extended wide area (even greater than 500 cm^2) exploiting TIR phenomenon arising between the plastic of the component and the air. In order to emit light in the desired direction, a special set of prisms is inserted in the light-guide to generate a deviation from the TIR condition, changing the ray-path and spreading light from the surface to the defined directions. In *Fig. 1.2.7* a ray-tracing simulation has been performed to demonstrate graphically the working principle of a light guide.

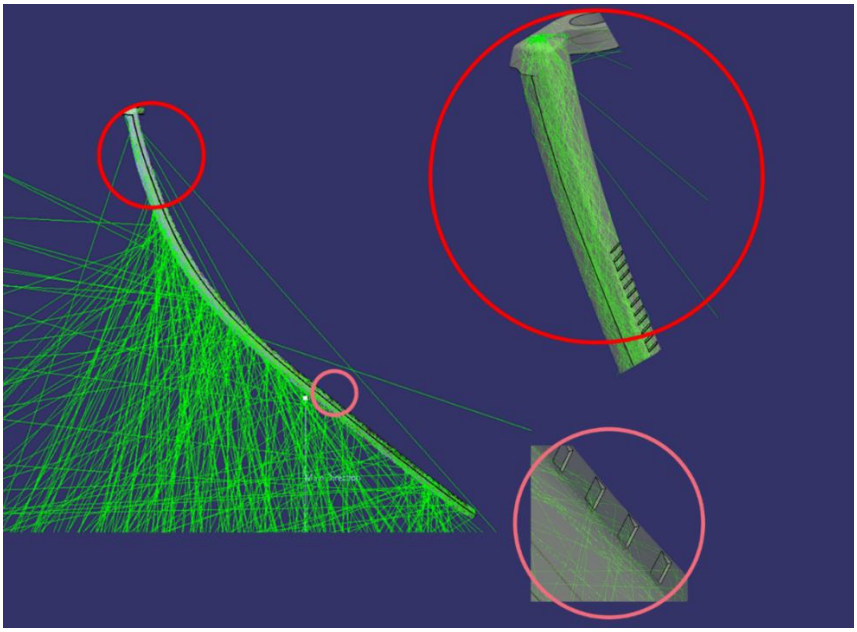

Figure 1.2.7: ray-tracing simulation to demonstrate the functionality of a light guide: in the red circle, the TIR is exploited. In the pink circle, by means of tailored prisms the propagation direction of light is modified.

In the first part of the light guide, highlighted by the red circle, the TIR is exploited in order to propagate the light only inside the plastic component. In fact, all the rays (in green) remain trapped in the light guide. To bring out the light in the chosen directions, prisms are used (pink circle).

1.2.5 Outer lenses

The outer lens is fundamental to protect the internal components of the lamp from damages, water leakage and every ambient impurity which could damage the interior of the lamp itself (example of outer lens represented in *Fig. 1.2.8*). In addition, the outer lens plays a key role in reaching the correct colorimetry imposed by the regulations and it defines the aesthetic aspect of the entire lamp, contributing to the appearance of the vehicle. It is welded together with the back housing and it must also resist to all the mechanical stress of the moving vehicle and to the heat produced by the light source during operation condition. Outer lenses are made of PMMA, which is subjected to severe chemical proofs to resist oil, fuel, wax and, moreover, to comply with a broad temperature ranges spanning from -40°C to 95°C, simulating the different environmental conditions that can be found in the entire world. The outer lens can be assembled to the housing (containing all the components described above) with several technique as vibration welding, hot plate welding and laser welding depending on the geometry of the lens itself.

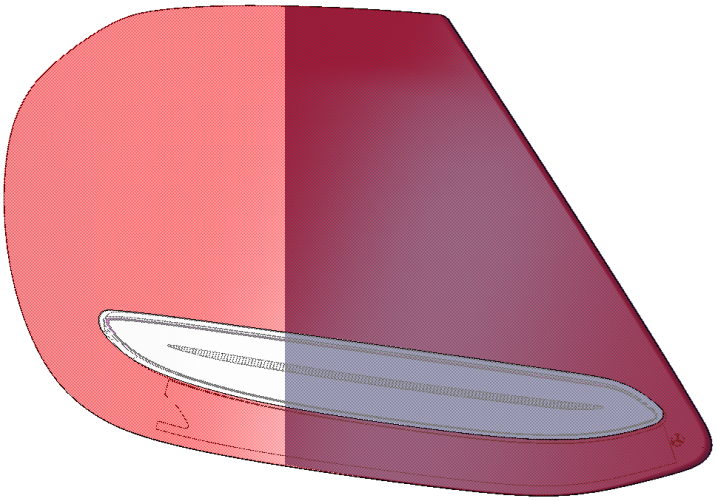

Fig. 1.2.8: example of outer lens.

1.2.6 Materials

Polycarbonate (PC), Acrylonitrile butadiene styrene (ABS) and Polymethylmethacrylate (PMMA) are the three main thermoplastic materials employed in the manufacture of the rear lamp components.

PC is a strong and durable material that can be easily worked and molded. Its glass transition temperature is 147°C and it has a high impact resistance, while its scratch resistance is quite low. Its colorless version is highly transparent to visible light, with better light transmission than many kinds of glass. It is stronger than PMMA and it holds up longer to extreme temperatures typical of LED light sources, making it the ideal material to manufacture internal components of the lamp.[42,43]

ABS's properties are very similar to the PC ones, making it a strong material which can be easily worked and molded. Its glass transition temperature is 105°C while its melting point can be located around 110°C (cannot be precise because ABS has an amorphous nature). The major differences regard the optical properties, because the ABS is more matte and opaquer with respect to the PC, which is brighter and shinier, as will be stressed in *Chapter 6*, and the price, because ABS is cheaper than PC.[44,45]

PMMA is characterized by a good tensile and flexural strength, solid abrasion resistance and outstanding UV tolerance. On the other side, impact strength and heat resistance are less developed than PC and ABS. The PMMA glass transition temperature ranges from 85 to 165°C depending on the precise composition of the plastic. It can transmit up to 92% of visible light and the reflection contribute from each of its surfaces is about 4% (depending on its refractive index which depends on the precise chemical composition). It filters light having wavelengths below 300 nm, preventing damages from UV radiations while passes infrared light up to 2800. Its high scratch resistance makes it the best material to be exploited for the outer lenses manufacture.[46,47]

The above presented materials are shaped, by means of injection molding process (presented in *Chapter 2*), to form the different components.

1.3 The physics and the optical design of rear lamps: photometry, colorimetry and luminance requirements

Photometry is the science concerning measurement of light and perception of brightness by the human eye; it refers to the electromagnetic radiation, restricted to the visible range (360-830 nm).[48]

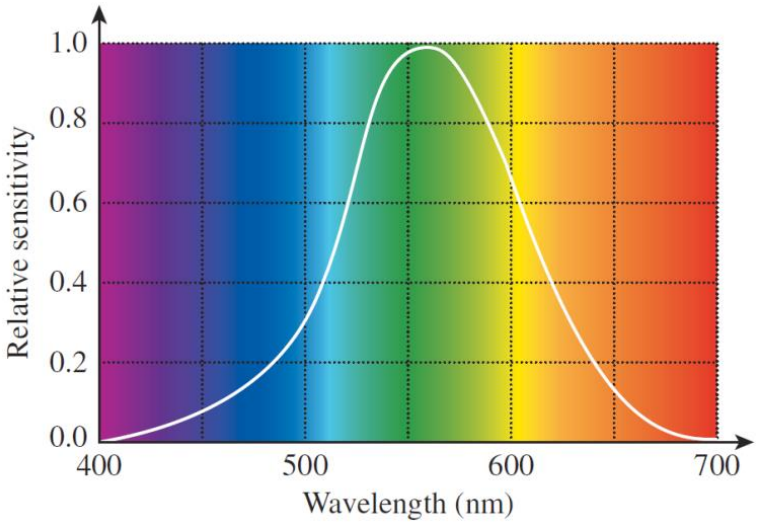

Fig. 1.3.1: representation of the photopic spectral luminous efficiency function (white).[49]

Before going into the details regarding the photometric quantities, it is important to introduce the photopic spectral luminous efficiency function $V(\lambda)$.[50] This function, represented by the white profile in *Fig. 1.3.1*, maps the specific sensitivity of the human eye to the electromagnetic radiation wavelengths and it has been determined experimentally by the CIE (Commission Internationale de l'Eclairage) in 1924.[51] Spectral radiant flux Φ_λ (energy emitted by a source per unit time), spectral irradiance E_λ (flux received by a surface per unit area), spectral radiance L_λ (flux emitted by a surface per unit solid angle per unit projected area), and spectral intensity I_λ (power emitted by a light source per unit solid angle) can be therefore convolved with the abovementioned $V(\lambda)$ to obtain the corresponding photometric quantities

$$\Phi_v = K_m \int_{360}^{830} \Phi_\lambda \, V(\lambda) d\lambda \tag{1.3.1}$$

$$E_v = K_m \int_{360}^{830} E_\lambda \, V(\lambda) d\lambda \tag{1.3.2}$$

$$L_v = K_m \int_{360}^{830} L_\lambda \, V(\lambda) d\lambda \tag{1.3.3}$$

$$I_v = K_m \int_{360}^{830} I_\lambda \, V(\lambda) d\lambda \tag{1.3.4}$$

that are all weighted by the human eye response. K_m is the maximum value of the luminous efficacy of radiation and has a value of $K_m = 683 \, \frac{lm}{W}$. The luminous flux (or luminous power)

represents the measure of the total perceived power of light in every direction and is measured in lumen (lm). However, the lamp performance is measured through the luminous intensity, in candela (cd) units, which takes into account, on the opposite of the luminous flux, the direction of the observer. The legislative constraints involving rear lamp evaluate the luminous intensity emitted by the lamp itself in different direction. *Fig. 1.3.2* shows the so called "photometrical grid" which represents the requirements a lamp must fulfill to be considered compliant. The luminous emission must accommodate precise intensity values in specific directions, considering several evaluation angles. Five numbers are shown inside each box: the blue and red (top and bottom right corners) represent the maximum and minimum intensities imposed by the regulations, the sky-blue and pink (top and bottom left corners) describe the maximum and minimum intensities accepted by the customer, while the black number in the middle of each box shows the measured luminous intensity.

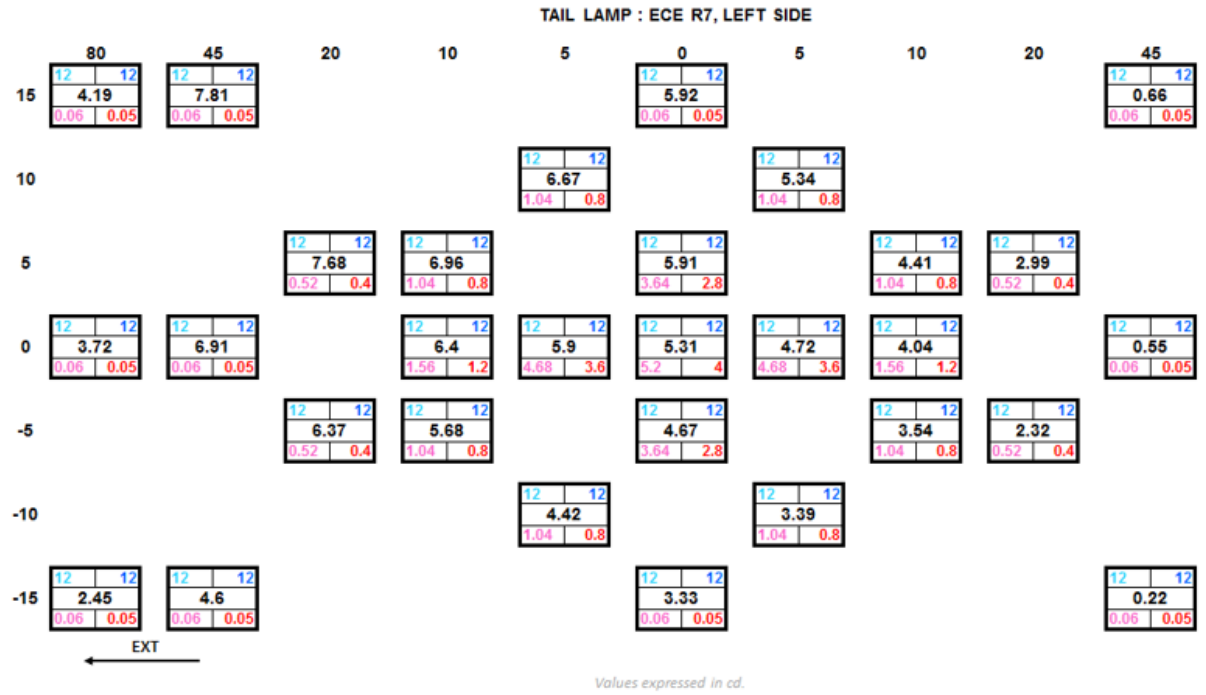

Fig. 1.3.2: example of photometrical grid.

Even if the legislative requirements are only related to the luminous intensity, sometimes it is possible that customers ask for a luminance and color evaluation too. Luminance is a measure of the luminous intensity per unit area of light travelling in a given direction. It describes the amount of light that passes through, is emitted or reflected from a particular area, and falls within a given solid angle as perceived by the human eye. The SI unit for luminance is the candela per square meter (cd/m^2) unit, which is denominated Nit. This quantity gives an idea of the "dazzling" produced by the source. In contrast to intensity, luminance is more related to the perception of light and better described the feeling drivers have while looking at a rear lamp, because it indicates the amount of luminous power that can be detected by the human eye while looking at a particular surface from a specific angle of view. Luminance can be thought as a way to physically describe the subjective perception of "brightness".

Finally, color is a characteristic of light which is related to the stimulation of different photoreceptors cells in the eye by photons having different energy. The high subjectivity of human color perception lead to the definition of conventions establishing a quantitative link

between the physical pure colors and the physiological perceived colors in human vision. The automotive sector adopts the 1931 CIE convention.[52] CIE's color matching functions are the numerical description of the chromatic response of the observer, as depicted in *Fig. 1.3.3*: in colorimetry, the light reaching the eye is called color stimulus, and this explains why the three signals represented in *Fig. 1.3.3* are called stimuli.

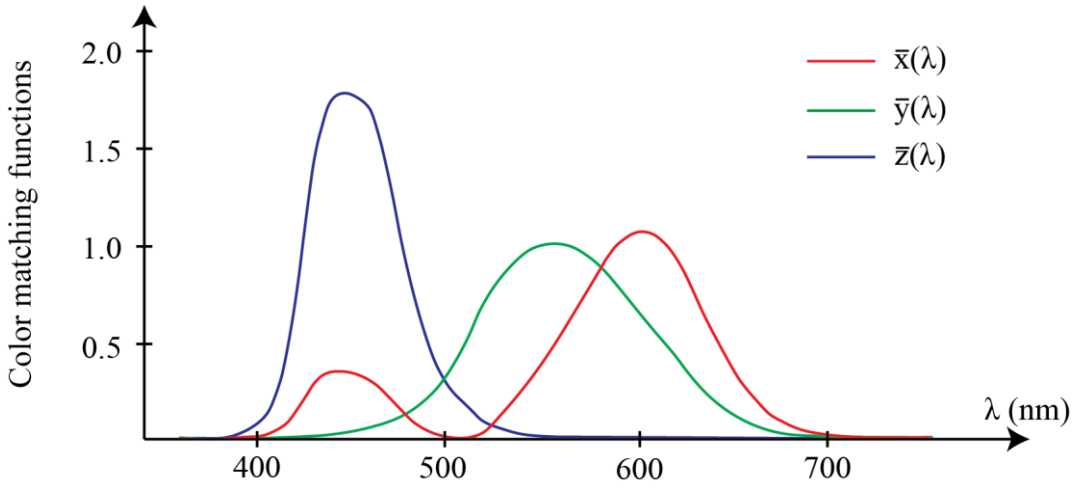

Figure 1.3.3: graphic representation of the standard observer $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ monochromatic test stimuli.

Collectively, these three functions are known as the CIE standard observer. The CIE XYZ color space was designed to give a simple, scientific, and reliable representation of color. In this description, the X, Y, and Z tristimulus values are defined as

$$X = k \int_{360}^{830} \Phi_\lambda(\lambda)\, \bar{x}(\lambda) d\lambda \tag{1.3.5a}$$

$$Y = k \int_{360}^{830} \Phi_\lambda(\lambda)\, \bar{y}(\lambda) d\lambda \tag{1.3.5b}$$

$$Z = k \int_{360}^{830} \Phi_\lambda(\lambda)\, \bar{z}(\lambda) d\lambda \tag{1.3.5c}$$

where $\Phi_\lambda(\lambda)$ is the color stimulus function of the light seen by the observer, k is a constant and $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ are the monochromatic test stimuli represented in *Fig. 1.3.3*. The chromaticity coordinates (x, y) can be introduced as

$$x = \frac{X}{X + Y + Z} \tag{1.3.6a}$$

$$y = \frac{Y}{X + Y + Z} \tag{1.3.6b}$$

and they are fundamental to describe the chromaticity with only two numbers. However, even if chromaticity can be expressed by means of two numbers, it is important to underline that the color stimulus can be correctly described only with 3 parameters (not 2). The usual CIE xyY chromaticity diagram, widely used to specify colors in practice, is a function of three variables (x, y, Y) is depicted in *Fig. 1.3.4a*.

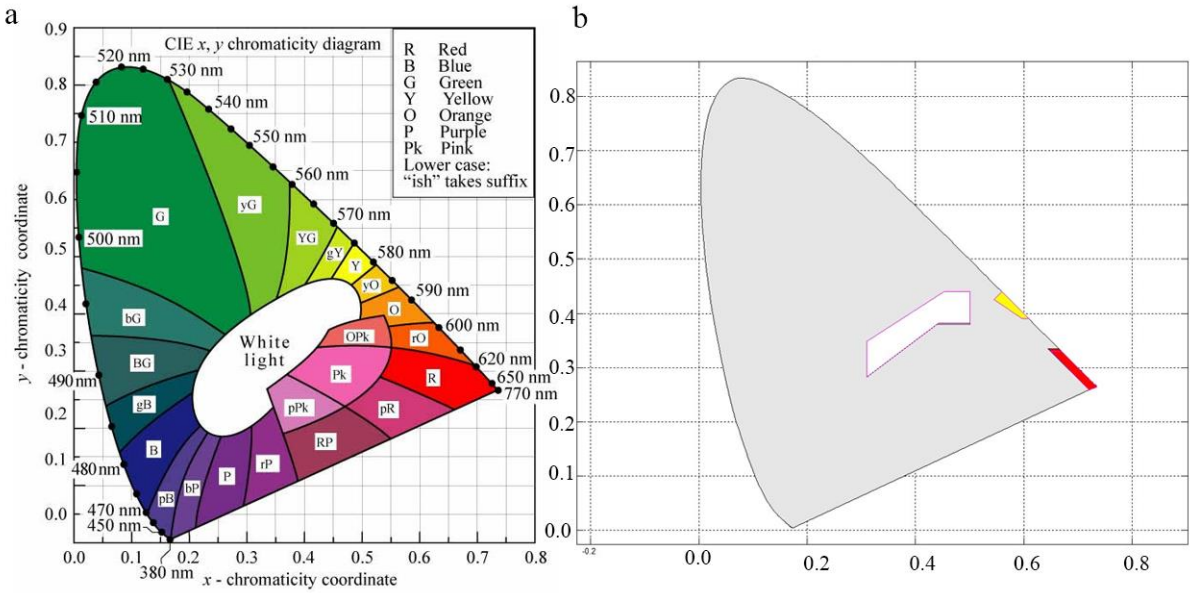

Fig. 1.3.4: CIE xyY chromaticity diagram (a);[53] *CIE xyY chromaticity diagram representing only the colors legislatively accepted on a rear lamp (b).*

This chart is a tool to specify how the human eye will experience light with a given spectrum: it cannot specify colors of objects, since the chromaticity observed while looking at an object depends on the light source as well. The regulation laws impose limits to the color owned by a rear lamp: it can be red, white or yellow, depending on the function considered and for each of those color different constraints must be fulfilled, as represented by the highlighted areas in *Fig. 1.3.4b*.

The final output of an optical simulation for a rear lamp consists of a report which contains all the information needed to describe the colorimetry and the photometry analysis. One fundamental tool exploited to evaluate the optical efficiency of a lamp is the isocandela map (*Fig. 1.3.5*): in this map regions having the same luminous intensity are represented with the same color. This can help visualizing the overall emitting properties of the lamp.

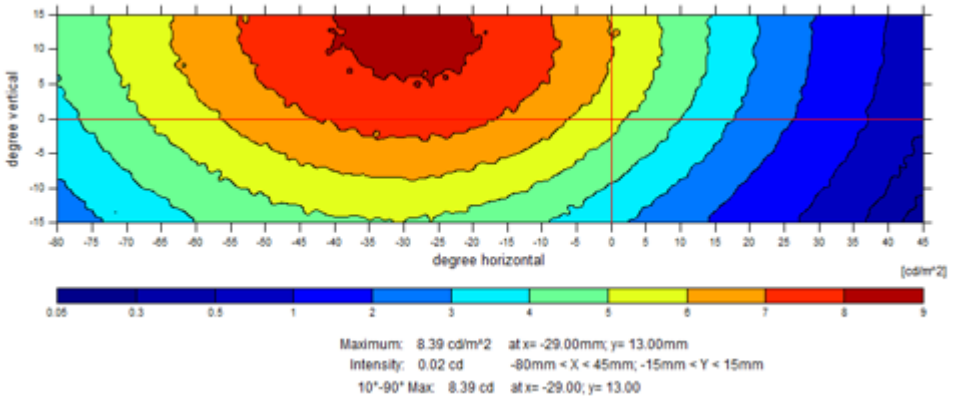

Fig. 1.3.5: example of isocandela map, which can supply an easy visualization of the luminous intensity of the portion of the lamp considered.

The aim of the optical and structural characterization of materials to be exploited in the rear lamps' manufacturing process, performed during this PhD project, is to provide more reliable tools to better simulate the different materials in the simulation environments. In fact, the correct classification of the diverse surfaces and interfaces of a rear lamp can return more precise simulations and therefore, a fast and reliable tool to check if the lamp is compliant to the photometry and colorimetry regulations.

2. Production of lamp optics: technical aspects and issues

2.1 Injection molding

Injection molding is a manufacturing process that produces components by injecting molten materials into a mold. The materials exploited during the process can be of different nature, from thermoplastic and thermosetting polymers to glasses and metals. In this thesis project I have concentrated my efforts on thermoplastic materials only. Injection molding will be synthetically described, but the attention will be mainly focused on the process-related parameters affecting the printing results.

The plastic granules, contained in a hopper, are inserted into a heated cylinder terminated with a nozzle extruder. An endless screw rotates within the cylinder and brings the granules inside the cavity, which is generated by the closure of the two steel mold halves. Here, pressure is increased allowing the plastic to take the mold shape without voids or gaps. Once the cavity is completely filled with fluid plastic, the mold is kept closed, so that the plastic can solidify by thermalization and maintaining the correct shape. After a couple of seconds, the mold halves open again and the component is ejected.[54] An example of injection molding structure is represented in *Fig. 2.1.1a*, while the visual description of the cylinder is displayed in *Fig. 2.1.1b*.

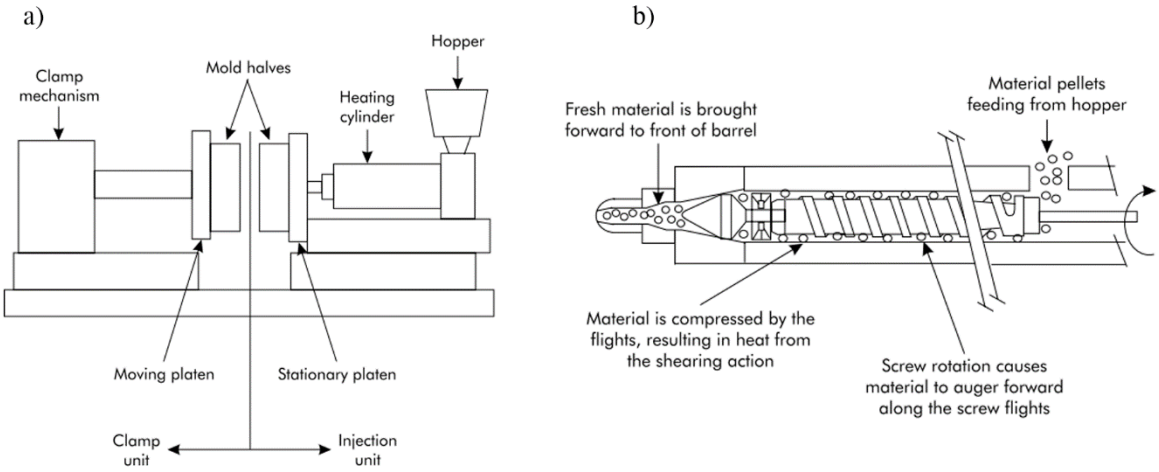

Fig. 2.1.1: (a), visual representation of the main components of an injection molding apparatus; (b), graphical description of the cylinder. Figure adapted from the literature.[55]

The above description is oversimplified, in fact the actual procedure is ruled by more than 100 parameters (i.e. temperature of the injected material, temperature of the mold, injection pressure, etc.) that must be fine-tuned to ensure production of a high-quality component.[56] However, in the next pages the three main parameters affecting the process will be introduced: temperature, pressure and time.[57]

2.1.1 Parameters

Temperature

Two different temperature values can be distinctly tuned in a standard injection molding process: temperature of the plastic material and temperature of the mold.[58]

Temperature of the material

All plastics have a most efficient injection temperature range. For amorphous phases, this range is rather broad, while for crystalline plastics it is much narrower (for the ABS the temperatures for a correct molding process range in the 210-270°C interval, while for PMMA the appropriate plastic temperature is in the range 210-230 °C). However, for both types of materials there is a specific temperature for which the plastic flows at best while maintaining its peculiar physical properties. Generally, the higher the temperature, the lower the viscosity.[59] This is true below a determined temperature limit, beyond which the plastic material degrades and its physical properties abruptly change (viscosity and phase diagrams can be found in the literature).[60–63] The cylinder containing the plastic granules is heated by electrical heater bands wrapped around it. Additional heat is generated by the turning action of the feed screw located within the cylinder: its main goal is to move fresh material into the cylinder from the hopper, but the turning action is also utilized to squeeze the plastic, thus creating friction.

Temperature of the mold

It is fundamental to cool the plastic at a rate that is slow enough to allow the material to reach proper physical strength and have the less thermal residual stress possible.[64] This is mainly obtained by controlling the mold temperature. The mold temperature ranges are quite wide, depending upon the different plastic used (PC: 80-120°C, ABS: 40-70°C). Specially designed channels are drilled into the most critical parts of the mold. These holes are then connected by means of hoses and water is flown inside them for proper thermalization.

Pressure

The hydraulic oil system within the molding machine and a series of control valves and regulators within that system provide the needed pressure. The pressure range is quite wide, spanning from approximately 3 MPa for fast flowing plastic, up to 135 MPa or more, for highly viscous materials. The specific requirements for the various pressure applications are discussed in the next lines.

Injection pressure

Injection pressure is the primary pressure used for the process and it is required to produce the initial filling of the free mold cavity, which represents approximately 95% of the total filling. The viscosity and flow rate of the injected plastic determine the amount of pressure that must be used, normally lying in the 70-135 MPa range. At the beginning of this phase the air trapped inside the mold is ejected by specifically designed duct.

Holding pressure

Holding pressure is applied at the very end of the primary injection stroke and it is used for the final 5% filling of the cavity image. It is called in this way because pressure is held against the cooling plastic in the mold while the material solidifies. This phase is fundamental to give the molded component its desired aspect and physical properties. Holding pressures are usually in the range of 50% of the primary injection pressure.

Back pressure

Back pressure is applied after the injection phases mentioned above. When the holding phase is completed, the machine starts turning the screw to bring in new material, in preparation for the next cycle. In this phase, the screw is not pulled back, but it only rotates and the fresh material itself begins to push the screw backward generating the so-called back pressure. Back pressure is small compared to the injection one, being approximately 10% of it. This parameter actually determines density and material appearance, minimizing (or even eliminates) voids in the molded product.[65]

Time

Time is the last, but not least, important parameter that will be introduced in the following. Two different time intervals can be distinguished: the injection time and the cooling time.

Injection time

The amount of time required for injection depends on how much material is being injected, the viscosity of the material, and the percentage of the machine's barrel capacity being utilized. During the initial injection time, the screw does not turn but only forces the material into the mold by pressing it. Initial injection is performed using the highest practical pressure for the specific application (normally between 70-135 MPa) in the fastest practical amount of time. In most cases it does not exceed 4-5 seconds. On the contrary, during the hold time the injection screw maintains pressure against the plastic that is already into the mold and no additional material is pushed inside. The molten plastic enters the mold cavity image through a gate. After all, the required material goes through the entrance and packs the cavity image, therefore the molten material is allowed to cool under hold pressure until it solidifies. The gate is the first portion to solidify, being normally the thinnest part of the cavity image. At this point, pressure is not applied because the plastic in the cavity lies beyond the solidified gate and the pressure from the injection unit no longer has any effect on it. The injection time can be defined as the interval elapsed between the beginning of the casting flowing in the mold and the moment when the gate solidifies. A gate with a thickness of 1.5 mm would take approximately 6 seconds to solidify.

Cooling time

The cooling time is the amount of time required for the plastic material to cool and solidify, allowing the plastic part to be rigid enough to withstand the ejection process. Ejection

occurs when the finished molded product is pushed out of the mold after the entire cycle is completed. If the outer skin of the plastic product is solidified to a sufficient depth, the remaining cooling will not have an appreciable effect on the molded part. If the skin is too thin, the remaining cooling will cause shrinkage stress to build up and the molded part may warp, twist, blister, or crack. The key to minimize these problems is to keep the part in the mold for a sufficiently long time. On average, a 1.5 mm-thick wall should take approximately 9-12 seconds to solidify (depending on material) to the point at which it can be ejected from the mold without undue distortion.

2.1.2 Defects

As described before, the injection molding process is controlled by a great number of parameters, which must be fine-tuned to produce a plastic item that is both aesthetically acceptable and with the correct structural properties. When the value of one single parameter is not correct, defects may arise on the final product. The three most common and frequent defects will be now introduced explaining their nature; for every defect, the possible causes ascribable to machine, mold, or material fault will be listed.

Black Spots or Streaks

If the plastic remains in the barrel longer than normal, it will begin to degrade. This degradation results in carbonized plastic, yielding small black clusters in the molded part, as shown in *Fig. 2.1.2*.

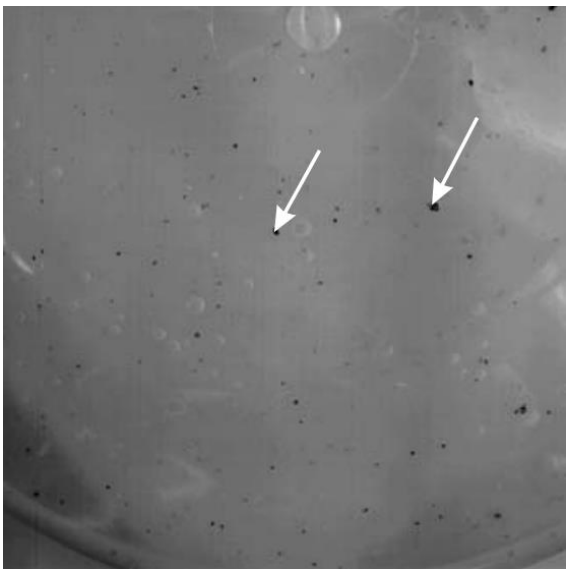

Fig. 2.1.2: visual representation of black specks. Figure adapted from the literature.[66]

These defects affect both opaque and transparent surfaces. The black spots could be also ascribed to sprue bushing cracked or broken. Any of these conditions will cause plastic to hang up in the mentioned cracks or holes where the material can overheat due to excessive residence time at that location, causing degradation or carbonizing.

Blisters or Bubbles

As the material is heated and transferred through the heating cylinder, air may be trapped within the molten material, as displayed in *Fig. 2.1.3*.

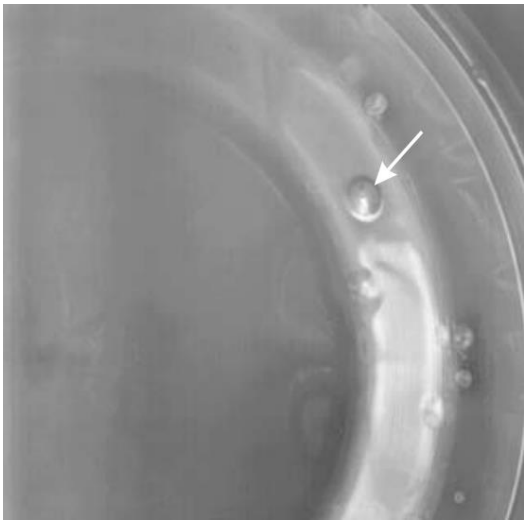

Fig. 2.1.3: example of an air bubble trapped inside the plastic (white arrow). Figure adapted from the literature.[66]

One of the uses of back pressure is to force this air out before it gets injected into the mold cavity image: if the back pressure is too low, air can remain trapped inside. Moreover, the injection temperature may be too high, which can cause the molten material to be extremely fluid, so air and gases remain trapped in the turbulent molten stream. As the material is injected into the mold, it starts to cool immediately, and a skin begins to form on the surface of the part. If this skin forms too quickly, any air that is mixed into the material will not be allowed to escape through the surface as intended, causing a blister. A cold mold will cause the skin to form too soon hence, to reduce blisters or bubbles, the mold must be heated.

Flow Lines

If the injection pressure is too low, the material entering the cavity is not packed together to form smooth layers against the molding surface, but it actually starts to wrinkle as one layer tries to crawl over the already cooling layer outside of it (*Fig. 2.1.4*). The cold mold could be another possible cause of flow lines: generally, a hot mold will allow the molten plastic to flow farther before cooling off and solidifying, minimizing the formation of flow lines. Eventually, a material that is too stiff may not flow fast enough to pack the mold before it solidifies, and the flow front may not be able to squeeze out the flow lines that will form.

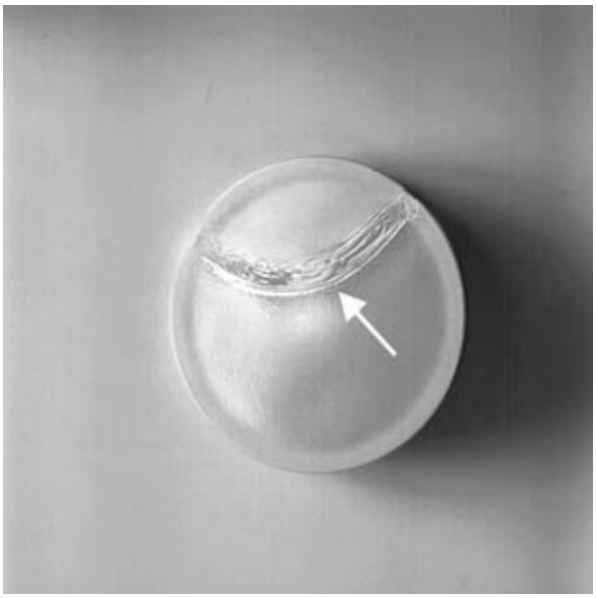

Fig. 2.1.4: example of flow lines on a plastic molded component. Figure adapted from the literature.[66]

2.2 Coatings deposition

Once the reflectors are molded, they need to be metallized in order to get their typical shiny and reflective aspect. The metallization process consists in the Physical Vapor Deposition (PVD) of aluminum (Al) on the bare reflector plastic substrate, and a following Plasma-Enhanced Chemical Vapor Deposition (PECVD) of Hexamethyldisiloxane (HMDSO) on top to prevent oxidation and damage of the metal layer. HMDSO is transparent to visible light and affects only slightly the optical properties of the aluminum film, as confirmed by the spectrophotometer data reported in *Fig. 2.2.1*. These data are taken on a glass slab and represent the optical properties of a 115 nm thick Al layer with (panel *a*) and without (panel *b*) a 35 nm thick HMDSO protective film on top.

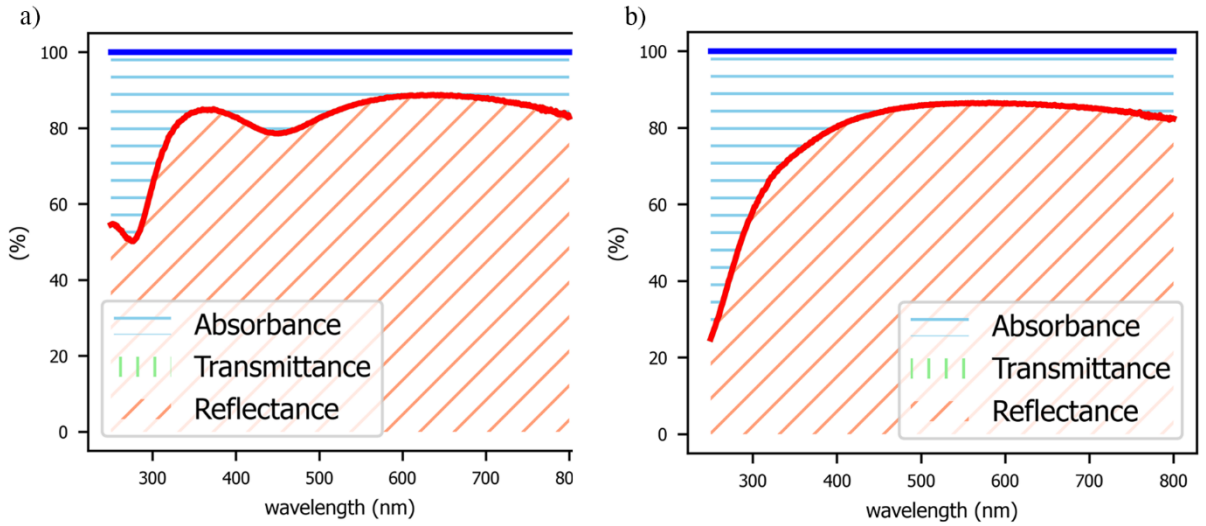

Fig. 2.2.1: spectrophotometer data acquired on 115 nm-thin film of Al and 35 nm-layer of HMDSO on top (a) and on 115 nm of bare Al (b) both grew on glass slabs. For both the samples, the transmittance is absent, meaning that the light is all absorbed or reflected.

With a standard Al coverage (115 nm is the standard thickness, as it will be shown in the next chapters) no light is transmitted through the sample. The HMDSO affects the optical performances of the Al as represented by *Fig. 2.2.1*: a decrement and modification of the reflectance in the 400-600 nm range is tolerated to ensure the physical protection of the metal layer.

2.2.1 Metallization setup

The metallization setup exploited in the manufacturing plant is a big cylinder (diameter = 190 cm, height = 180 cm) in which deposition of both Al and HMDSO are performed. Specifically, the ones exploited are the Tecnomet MV by Tecno Vacuum (*Fig. 2.2.2*). They are provided with two hemicylindrical doors to increase the productivity (*Fig. 2.2.2a*): while one door is closed (metallization cycle running), the other one is opened so that the operator can unload the metallized reflectors and load new unmetallized ones. The hemicylinder contains a mechanical carousel consisting in 8 support structures, which are filled with housing masks tailored for the specific reflector geometry (*Fig. 2.2.2b*).

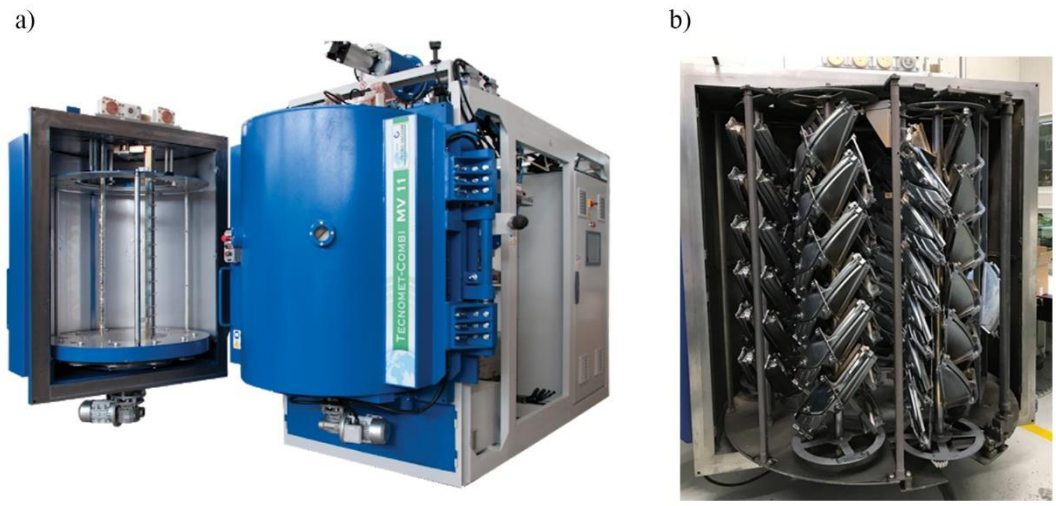

Fig. 2.2.2: (a), picture of a Tecno Vacuum Tecnomet MV, the metallizing machine used to grow the Al and HMDSO coatings on the reflectors.[67] *(b), metallizing machine filled with 8 support structures. Every structure contains the housing masks, which are specific for each different reflector type.*

Three fundamental elements are located inside the machine: an evaporator, a cathode, and a mechanical carousel. The evaporator is composed by 22 tungsten heating wires, each designed with a two-meanders-like shape (*Fig. 2.2.3a*), where the solid state Al spirals must be inserted (*Fig. 2.2.3b*). The geometrical properties of the wires are highlighted in *Fig. 2.2.3c*.

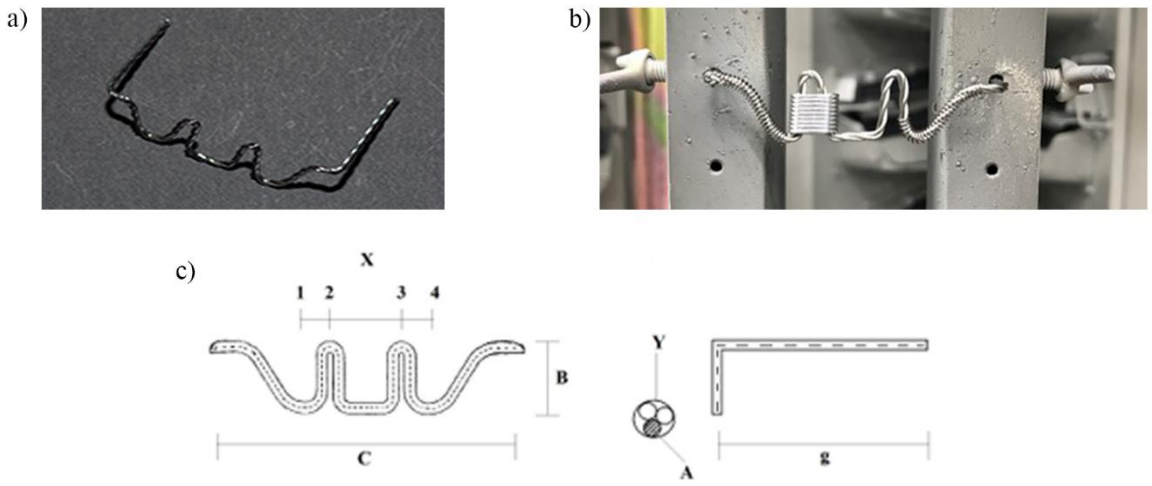

Fig. 2.2.3: picture of a tungsten wire (a). Al spiral inserted in one of the two meanders of the W wire (b). Schematic representation of the tungsten wire used for the Al evaporation (c). A = 0.75 mm (diameter of the Y circumference = 1 mm), B = 15 mm, C = 55 mm. Each tungsten wire is composed by three little wire weaved together, conferring it the possibility to be used for a longer time.

The bare Al source material has a nominal purity of 99.82% (composition listed in *Table 2.2.1*).

Element	Chemical composition weight (%)
Al	99.82
Si	0.04
Fe	0.12
Mn	0.01
Zn	0.01

Table 2.2.1: effective elemental composition of the Al spirals. The overall contamination concentration is less than 0.2%.

The cathode is a big cylinder (diameter = 10 cm, height = 130 cm) exploited to generate a potential difference inside the machine (the anode is represented by the carousel containing the housing masks). An intense electric field is fundamental to create a plasma inside the chamber and thus performing plasma cleaning and depositions, as will be described in the following. Up to 8 support structures can be inserted inside the mechanic carousel, each one provided with a variable number of housing masks for the specific reflector. Each structure rotates around its axis, while the 8 supports rotate around the center of the metallization machine, yielding both rotation and revolution motions.

2.2.2 Metallization process

The metallization process lasts approximately 22 minutes: during this time interval, a set of precise and well-defined steps are performed to give the reflectors its optimal reflective properties, thus following a customized and tuned recipe.

In the first part of the cycle the reflectors are loaded into the machine and the pumping system (composed by roots pump and scroll pumps) is turned on. When a pressure in the $8 \cdot 10^{-2}$-$1 \cdot 10^{-1}$ mbar range is reached, a potential difference of 2500 V (plasma current ~ 1 A) is applied for the plasma cleaning process, which lasts 300 s. This is fundamental to remove as much environmental contaminations as possible from the reflectors' surfaces. Plasma cleaning uses the interaction of ions, electrons, and radicals with the surface offering several advantages over conventional chemical cleaning methods. Plasmas provide a low temperature environment which does not exploit heat to promote chemical reactions, thus eliminating many of the drawbacks associated with wet chemistry. Moreover, plasmas are environmentally friendly, with no chemical wastes and hence no expensive disposal procedures. Argon is the element used to generate the plasma and it removes contaminants by exploiting ablation by energetic ions.[68–70]

Al is then deposited by means of PVD on the clean plastic surface.[71] PVD is an atomistic deposition process in which the material is sublimated from a solid source in the form of atoms or molecules and transported in the form of a vapor through a vacuum or low-pressure gaseous (or plasma) environment to the substrate where it condenses.[72] There are many ways for film depositions via PVD process, such as sputtering deposition, arc vapor deposition, and ion plating.[73] Thanks to this technique, it is possible to deposit a thin layer of material with controlled deposition rates in order to obtain, for example, nanostructures with well-defined growth direction and thickness.[74] In the present case, heating the Al spirals (physically in contact with the tungsten heating wires) is the method employed to evaporate aluminum and then grow the thin film. The PVD process starts when pressure in the chamber is in the 2.8-3.6

10^{-4} mbar range. The deposition in vacuo is mandatory to let the mean free path of the particles be long enough to reach the non-metallized reflectors and to avoid Al pre-oxidation by oxygen or water. During the deposition, the substrate is kept uniformly at room temperature. The uniform substrate temperature, granted by keeping a correct distance from the evaporator, is fundamental in order to avoid formation of agglomerates or heaps in the deposited layer. The evaporation process for the specific reflectors investigated in the present work lasts 45 seconds: high current (2000-2500 A) circulates in the W wires and heats the Al until it evaporates. The current ramp (300-350 A/s) is set to heat uniformly the Al spirals and not letting solid agglomerates reaching the substrate.

Over the aluminum layer, an HMDSO thin film is deposited by means of PECVD, without breaking the vacuum. Chemical vapor deposition (CVD) belongs to the class of vapor-transfer processes. It enables the deposition and growth of a solid phase structure at a surface starting from a chemical vapor phase that reacts and is catalytically converted at the solid-gas interface.[75] The material that needs to be deposited can be mixed with other more volatile precursors that work as carriers. The precursors chemically interact with the source material, generating volatile by-products. The PECVD arises from CVD and uses electrical energy supplied by the cathode to produce a plasma. The plasma activates the reaction by transferring the energy of its species to the precursors and induces free radical formation followed by radical polymerization.[76] The resulting chemistry of the thin films produced by PECVD is unique and cannot be obtained by common wet deposition techniques. Thin films grown in this way also demonstrate high solvent and corrosion resistance along with thermal and chemical stability, which makes PECVD the ideal technique to deposit HMDSO to protect the reflectors from external damages and oxidation. The PECVD process, characterized by a potential difference of 3700 V and a current of 0.5 A, lasts 300 s, allowing the growth of an adequate protective film onto the Al layer without strongly affecting the optical properties of the entire reflector.

In the course of all the processes mentioned above, the mechanical carousel rotates at a speed of 7 rpm, allowing the Al and HMDSO to be correctly deposited on every reflector placed inside the machine.

2.3 Photolithography

Photolithography (or optical lithography) is a process by which a light-sensitive polymer, denominated photoresist, is exposed to UV (in this case) light and developed to form a 3D relief pattern on the substrate. The ideal photoresist image has the exact shape of the designed mask exploited in the plane of the substrate, with vertical walls through the thickness of the resist. Thus, the final resist pattern should be binary: parts of the substrate are covered with resist while other parts are completely uncovered. This patterning is fundamental to obtain the desired geometry for the particular application. In the present case, this approach has been exploited to generate sharp edges in order to measure, by means of AFM, the thickness of the deposited films. The general steps carried out during a typical photolithographic process can be listed as: preparation of the substrate, spin coating application of the photoresist, post-apply bake, exposure to the (UV) light, post-exposure bake, and development (key-steps illustrated in *Fig. 2.3.1*). A brief discussion of each of them will be now given.

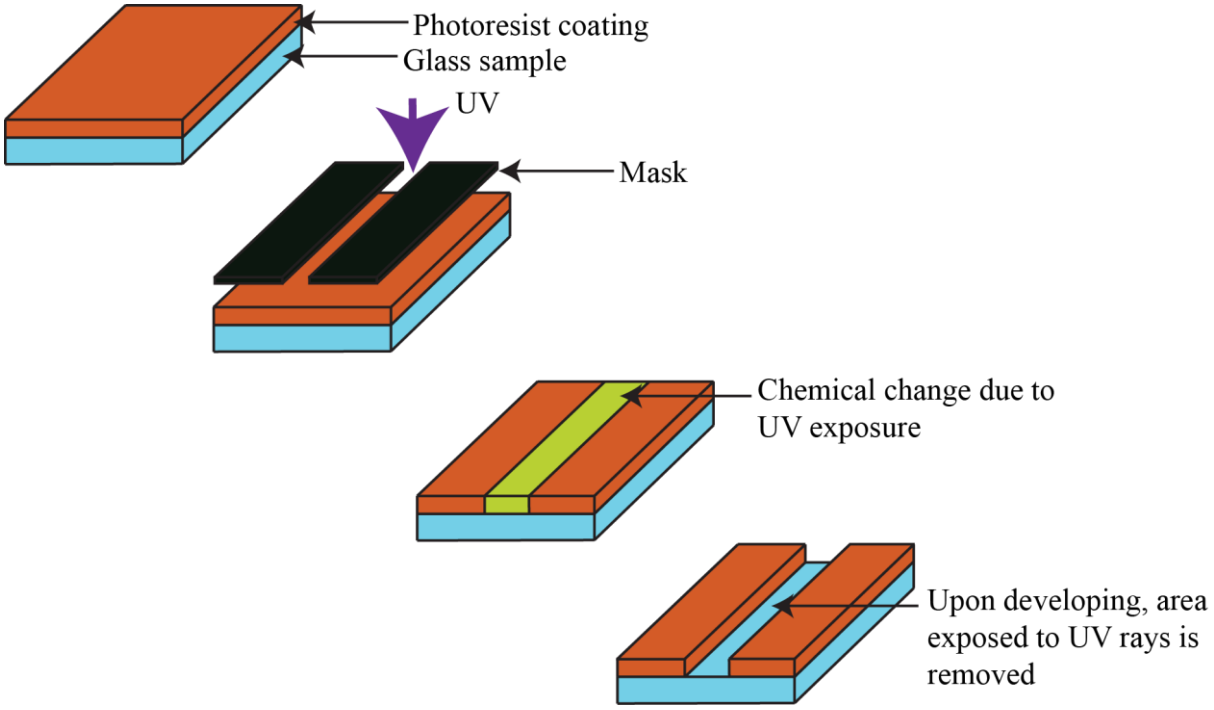

Fig. 2.3.1: key-steps characterizing the photolithographic process.

Substrate Preparation

The substrate must be prepared before the lithographic process, to improve the adhesion of the photoresist material and to yield the growth of a clean resist film without contaminations. Substrate contamination can take the form of particulates or the one of a film, which can be either organic or inorganic. The presence of particulates will result in spot-like defects in the final resist pattern, whereas film contamination may be the cause of poor adhesion. Particulates generally come from airborne particles and a chemical/mechanical cleaning is enough to remove them. Organic films, such as oils or polymers, can be originated from leftover due to the previous processing steps, while it is possible to encounter body oils and sweat, ascribable to human contamination. These films can be removed by chemical, ozone, or plasma stripping.

Similarly, inorganic films, such as native oxides and salts, can be removed by chemical or plasma stripping.

Photoresist Coating

A thin, uniform coating of photoresist is accomplished by the spin coating process. The photoresist is poured onto the substrate, which is then spun on a turntable at a high speed producing the desired film.[77] Two possible methods can be exploited: static dispense (stationary substrate while resist is dispensed) or dynamic dispense (spinning substrate while resist is dispensed), while spin speed, acceleration, and duration of the process can be tuned to fit the desired request. An important role in the resist thickness uniformity is played by the dispensed volume of the resist, its properties (such as solvent composition and viscosity) and the substrate characteristics (substrate material and topography). At the end of the spin-coating process, a thick, solvent-rich film of photoresist covers the substrate, ready for the post-apply bake. The centrifugal force applied by the spinning process pushes the liquid photoresist toward the edge of the sample, where excess resist is flung off; as the film thins, the centrifugal force (proportional to the resist's mass) decreases. The frictional force, supplied by the viscosity of the resist, opposes to this drift. Evaporation of solvent inside the photoresist (due to drying) leads to an increase in its viscosity thus, in the end, the increasing viscous force exceeds the decreasing centrifugal force and the resist stops flowing.[78] This generally occurs within the first seconds of the spin cycle; the remaining fraction of the spin cycle concerns the evaporation of the solvent without mass flow of the resist solids. The separation of the spin cycle into a very quick radial mass flow (coating stage) and the subsequent evaporation of solvent (drying stage) provides for some of the basic and important properties of spin coating. The spinning time is longer than the coating stage time, so the final thickness of resist is practically independent on the initial volume of resist poured onto the substrate above a certain threshold. For laminar flow of air above the spinning plate, the dried solvent thickness will be proportional to the square root of the angular spin speed. Since most of the thinning of the resist comes from the drying stage, the final thickness of the resist will be inversely proportional to the square root of the spin speed. The photoresist spin speed curve (*Fig. 2.3.2*) is an essential tool for setting the spin speed to obtain the desired resist thickness.

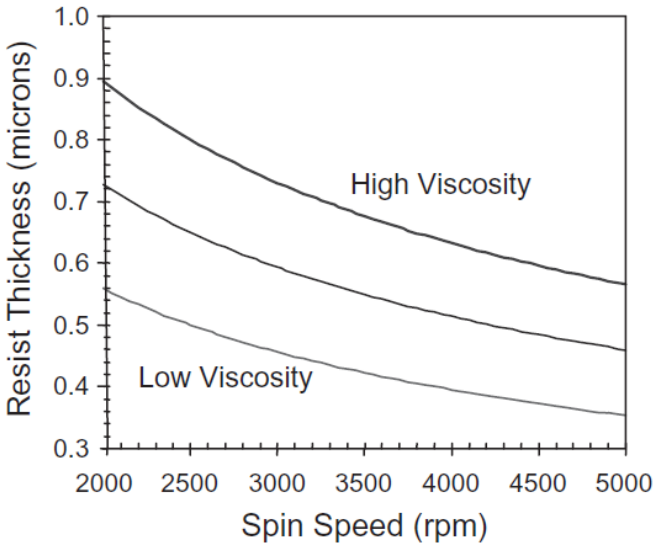

Fig. 2.3.2: spin speed curves for different resist viscosities. The inverse square root proportionality is visible from the graph. Figure adapted from the literature.[79]

As mentioned above, the final resist thickness varies as one over the square root of the spin speed (ω) and is roughly proportional to the liquid photoresist viscosity (ν) to the 0.4–0.6 power:

$$thickness \sim \frac{\nu^{0.4}}{\omega^{0.5}} \tag{2.3.1}$$

For a given desired resist thickness, the appropriate spin speed is chosen according to the spin curve.

Post-Apply Bake

After the spin coating procedure, the resist film will still contain between 20 and 40 % of solvent. It is fundamental to decrease the solvent percentage to obtain a more stable resist film. The post-apply bake (PAB) process, (sometimes denominated as prebake or softbake) involves drying the photoresist after spin coating by removing most of this solvent. Moreover, the adhesion properties are improved, and the film is less susceptible to particulate contamination. Typical PAB processes leave between 3 and 10 % residual solvent in the resist film, sufficiently small to keep the film stable during subsequent lithographic processing. Currently, the most popular method to bake photoresists is the hot plate. The substrate is brought either into intimate vacuum contact with or close proximity to a hot, high-mass metal plate. The photoresist is then heated to near the hot plate temperature quickly. The common exposure time is about 1 minute, which is the correct interval to improve uniformity of the photoresist.

Exposure

The basic principle behind the photolithography operation is the change in solubility of the resist in a developer upon exposure to light. The example of the diazonaphthoquinone positive photoresist will be presented.[80] This photoactive compound, which is not soluble in the

aqueous base developer, is converted to a carboxylic acid on exposure to UV light in the range of 350–450 nm. The carboxylic acid product is very soluble in the basic developer. Thus, a spatial variation in light energy incident on the photoresist will cause a spatial variation in solubility of the resist in developer. This is the principle exploited to change the solubility of the photoresist. We can distinguish two typologies of resists: positive, which are characterized by the solubility of the regions exposed to light, and negative, defined by an insolubility of the portion of photoresist exposed to light. Moreover, two types of lithography can be introduced: contact and proximity lithography, which are the simplest methods of exposing a photoresist through a master pattern called a photomask. Contact lithography offers reasonably high resolution (down to about the wavelength of the radiation) but introduces practical problems such as mask damage (or equivalently, the formation of mask defects). On the contrary, proximity printing reduces mask damage by keeping the mask at a finite distance above the sample (e.g. hundreds of μm),[81,82] but the resolution limit is increased significantly. Another important aspect of photoresist exposure is the standing wave effect. Monochromatic light, when projected onto a substrate, hits the photoresist surface over a range of angles, approximating plane waves. This light crosses the deposited material and, if the substrate is reflective for the exploited wavelength, is reflected up through the resist. The incoming and reflected light waves interfere, forming a standing wave pattern of high and low light intensity at different depths in the photoresist. This pattern is replicated in the material, causing ridges in the sidewalls. This kind of interference may cause a phenomenon denominated swing curves, consisting in a sinusoidal variation in the resist thickness.

Post-exposure bake

One method of reducing the standing wave effect is called the post-exposure bake (PEB). The high temperatures (100–130 °C) cause the diffusion of the photoactive compound, thus smoothing out the standing wave ridges. For a conventional resist, the main importance of the PEB is diffusion to remove standing waves.[83] On the contrary, for the chemically amplified resists the PEB is an essential step of the chemical reactions that create a solubility differential between exposed and unexposed regions of the resist. For these resists, exposure generates a small amount of a strong acid that does not itself change the solubility of the resin. During the PEB, this photogenerated acid catalyzes a reaction that changes the solubility of the polymer resin in the resist, that's why the control of the PEB is extremely critical for chemically amplified materials.

Development

Once exposed, the photoresist must be developed. The most used compounds employ aqueous bases as developers. The method of applying developer to the photoresist is important in controlling its uniformity and process latitude. It consists in dissolving the exposed or unexposed resist (depending if the resist used is positive or negative) in the developer. Several methods of development are available, as for example batch development, which consists in a boat of some 10–20 substrates developed simultaneously in a large beaker, usually with some form of agitation. In spin development the substrates are spun, using equipment similar to that used for spin coating, and developer is poured onto the rotating samples. The sample is also rinsed and dried while still spinning. Spray development uses a process identical to spin development except the developer is sprayed, rather than poured, on the substrate by a nozzle.

After the development, it is possible to perform a rinse in water or another bake on the hot plate to reduce the impurities on the surface.

In the present work, photolithography was applied on glass slabs, which have been cleaned by chemical processes in an ultrasound bath. The slabs have been then dried by means of gaseous nitrogen. The positive photoresist S1818 G2 from "Rohm and Haas" has been applied on the cleaned glass surfaces by static dispense. The spin coating process lasted 40 seconds at a speed of 3600 rpm, while the following PAB lasted 1 minute at 100°C. A 40 sec exposure to a 250 nm wavelength UV light was then applied, with a subsequent 1-minute PEB at 100°C (performed to reduce the standing wave effect). The developer bath exploited to develop the photoresist was composed by MF319 from "Rohm and Haas" and lasted 40 seconds. The agitation was performed by hand. For the experiments reported in this thesis, the photolithography was crucial to determine the thickness of the Al layer growth on a surface. In fact, it was exploited to create a sharp step between the bare substrate and the resist. The samples were then metallized and by means of a lift-off in acetone the resist was removed, leaving a sample exposing bare substrate and metallized regions, as represented in *Fig. 2.3.3*. This photolithographic process was not exploited to determine the HMDSO thickness, because the acetone react with HMDSO removing it from the substrate.

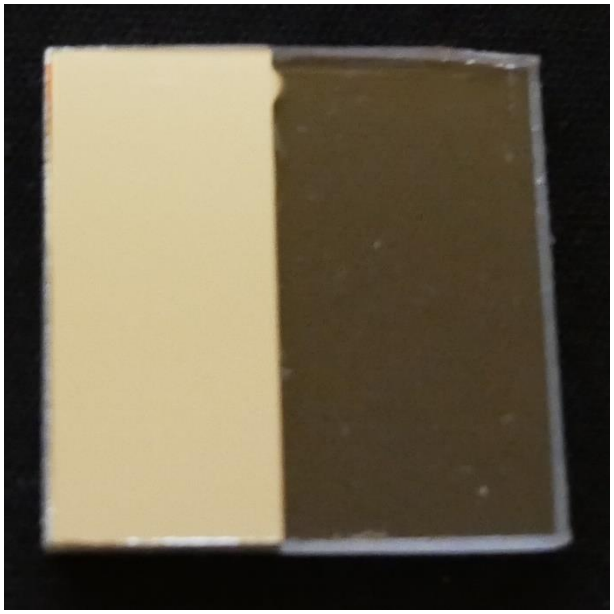

Fig. 2.3.3: sample produced by means of photolithographic process: the left part of the sample is covered with Al layer, while the right part is uncovered. By means of AFM is then possible to measure the thickness of the thin metal layer.

3. Introduction to the characterization methods

In this section, I will describe the experimental techniques exploited to study the optical and morphological properties of reflective coatings and heterostacks. The characterization has been performed by combining Scanning Electron Microscopy, Atomic Force Microscopy, Synchrotron-radiation X-Ray Computed Tomography, and Bidirectional Reflectance Distribution Function measurements. For each technique, the physics and the main features will be presented, including a description of the specific instrumentation and setup exploited to perform the experiments.

3.1 Scanning Electron Microscopy

Scanning Electron Microscopy (SEM) is an imaging technique routinely used to provide topographic and morphological information of conductive surfaces down to a few nanometers resolution by scanning on samples a focused electron beam.[84]

3.1.1 Electron-sample interactions

When the primary electrons (PEs) beam interacts with the surface, several signals can be produced, as represented in *Fig. 3.1.1*. In this chapter, only the description of the backscattered electrons (BSEs), secondary electrons (SEs), and characteristic X-rays will be addressed, which are the signals exploited to acquire the SEM images presented in this thesis.

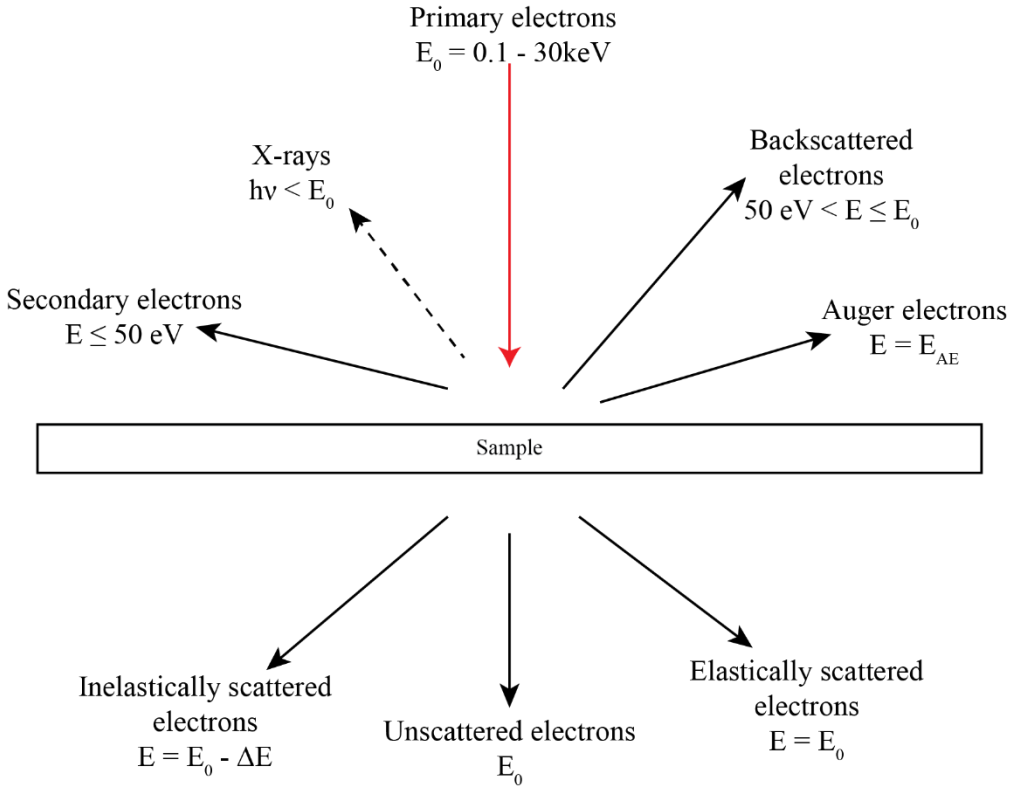

Fig. 3.1.1: drawing representing all the signals generated by the impinging electron beam. Only the BSEs, the SEs and characteristic X-rays have been exploited to acquire the SEM images shown in this thesis.

The electron absorption or scattering phenomena take place within a drop-shaped volume denominated interaction volume, which can extend from 100 nm to several μm inside the sample. The physical size of the interaction volume depends on the electron beam energy, on the elemental composition of the sample (higher atomic number materials absorb or stop more electrons, yielding a smaller interaction volume), and on the incident angle of the beam with respect to the surface (the further from normal, the smaller the volume) as represented in *Fig. 3.1.2*. The direct measurement of the interaction volume is rarely possible, and for this reason calculations and simulations are usually exploited to evaluate the electron-material interaction. Another important physical quantity is the escape depth (*t*), which represents the maximum depth under a surface from which secondary, backscattered, Auger electrons, and radiation can escape. Clearly, every kind of electron/radiation cited above has a proper escape depth, which will be labeled with a subscript in the next pages of the chapter (i.e., SE escape depth = t_{SE} etc.).

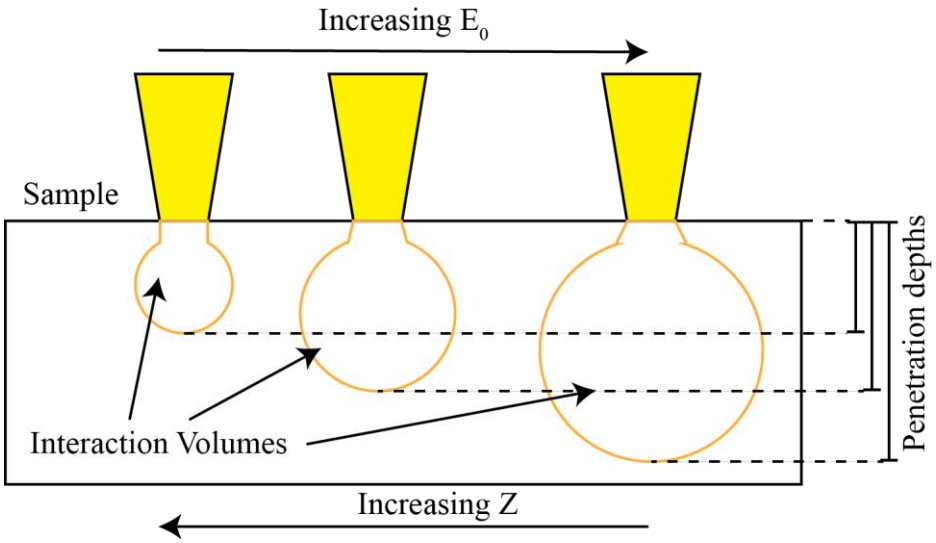

Fig. 3.1.2: examples of different interaction volumes. The energy of the electron beam E_0 (deriving from the accelerating potential) and atomic number (Z) dependences are highlighted, while the incident angle is taken always perpendicular to the surface.

3.1.2 Backscattered Electrons

BSEs are reflected by the sample's surface due to elastic scattering. Reflection is caused by the interaction with the electrical field of the positively charged nucleus and results only in a trajectory change of the beam, with small or no loss of kinetic energy. This type of scattering can be described semi-quantitatively by the elastic scattering cross section σ_{el}, using the principles of the Rutherford scattering:

$$d\sigma_{el} = \frac{1}{4}\left(\frac{Ze^2}{8\pi\epsilon_0 E_0}\right)^2 \frac{d\Omega}{sin^4(\frac{\theta}{2})} \qquad (3.1.1)$$

The equation above defines the differential scattering cross section as a function of the atomic number of the scatterer *Z*, the elementary charge e, the vacuum permittivity ϵ_0, the kinetic energy of the impinging electron E_0, and the scattering angle θ. The scattering cross section is proportional to Z^2 and inversely proportional to E_0^2, i.e., the probability for elastic

scattering increases with the atomic number and decreases for increasing electron energy E_0.[85] By definition, the energy of BSE is in the range $50\ eV < E_{BSE} \leq E_0$.

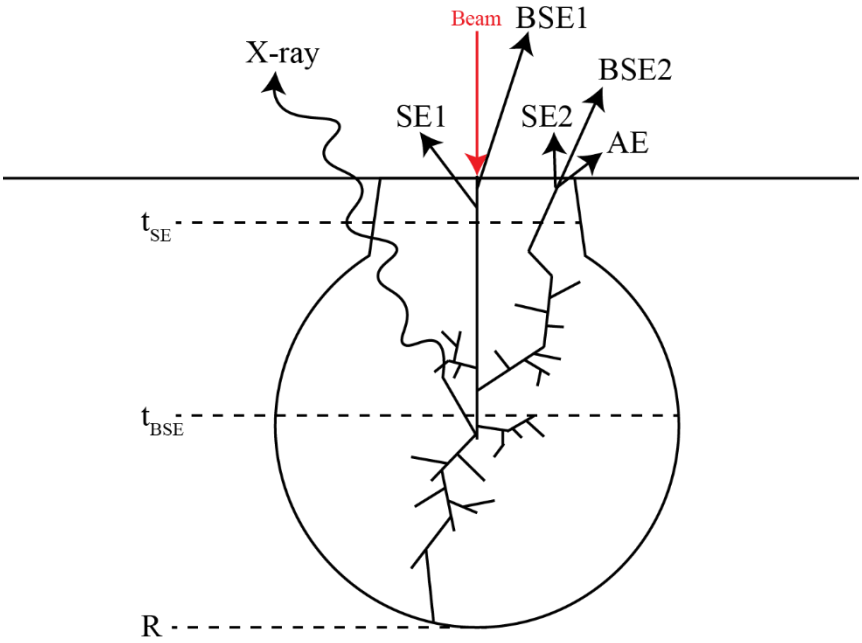

Fig. 3.1.3: generation of secondary electrons (SE1 and SE2) and backscattered electrons (BSE1 and BSE2) from a beam impinging on a sample. The escape depth for the SE and BSE are indicated with t_{SE} and t_{SE}, while R is the penetration depth.

As shown in *Fig. 3.1.3*, BSEs can originate either from the small portion of the sample directly irradiated by the electron beam (denoted BSE1) or after multiple elastic scattering events from a significantly larger area around the impact point (denominated BSE2). The lateral distribution of BSE2 has been calculated by Monte Carlo simulations for different materials.[86] Simulations show that the BSE emitting surface area increases with electron energy E_0 and decreases with the atomic number. The BSE1 carry local information about the small volume and deliver high-resolution information for a beam diameter of about 1 nm. As a consequence of the lateral spreading, the BSE2 carry information about a larger region, thus fine structural details on the scale of the beam diameter cannot be resolved.

Fig. 3.1.3 also shows that the primary electrons (PEs) travel in a small subsurface volume before they return to the surface to escape as BSE2. Knowledge on the angular distribution of BSEs is of great importance for understanding and optimizing the BSE detection geometry. For normal beam incidence the angular distribution can be approximated by a cosφ distribution,[87] where φ represents the BSE emission angle referred to the surface normal. BSEs move on nearly straight trajectories, therefore the angular detector position has a strong influence on the collection efficiency of the detector; for non-normal beam incidence, the distribution is asymmetric, and a reflection-like emission maximum is observed. The BSE coefficient η, a parameter quantifying the number of backscattered electrons, is defined as

$$\eta = \frac{n_{BSE}}{n_b} \tag{3.1.2}$$

where n_b is the number of incident electrons and n_{BSE} is the number of BSEs.[88] η is approximately independent of the electron energy E_0 in the range of about 10-30 keV. However, at low energies it depends in a complex manner on the atomic number.[89] At a fixed

energy, η increases monotonically with the atomic number, as depicted in *Fig. 3.1.4*, so BSE images provide a qualitative (but not quantitative) compositional contrast information.

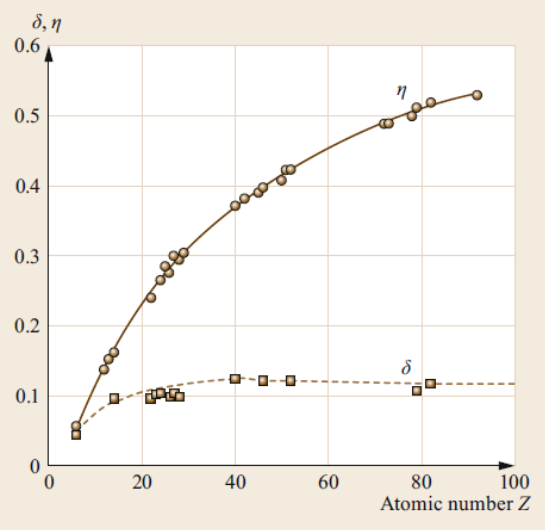

Fig. 3.1.4: trend of the BSE coefficient η and SE yield (δ) with respect to the atomic number Z of the specimen. $E_0 = 30\ keV$ *and* $\theta = 0°$. *Figure adapted from the literature.*[90]

3.1.3 Secondary Electrons

The inelastic scattering of the electrons is caused by their interactions with the electrical field of the electrons in the solid, i.e., either with the electrons in the valence or conduction band and with atomic electrons of inner shells, respectively. After an inelastic scattering event, the electron trajectory changes direction slightly (typically the inelastic scattering angles are of the order of a few milliradians only). The differential inelastic electron-scattering cross section with a free electron (which is an approximation for an electron in the valence or conduction band) is given by *Eq. 3.1.3*

$$d\sigma_{in} = \frac{\pi e^4 dW}{[(4\pi\varepsilon_0)^2 E_0 W^2]} \tag{3.1.3}$$

where W is the energy loss in the inelastic collision. *Eq. 3.1.3* shows that the differential inelastic scattering cross section is inversely proportional to electron energy E_0 and to W^2, and that small energy losses occur with a larger probability.

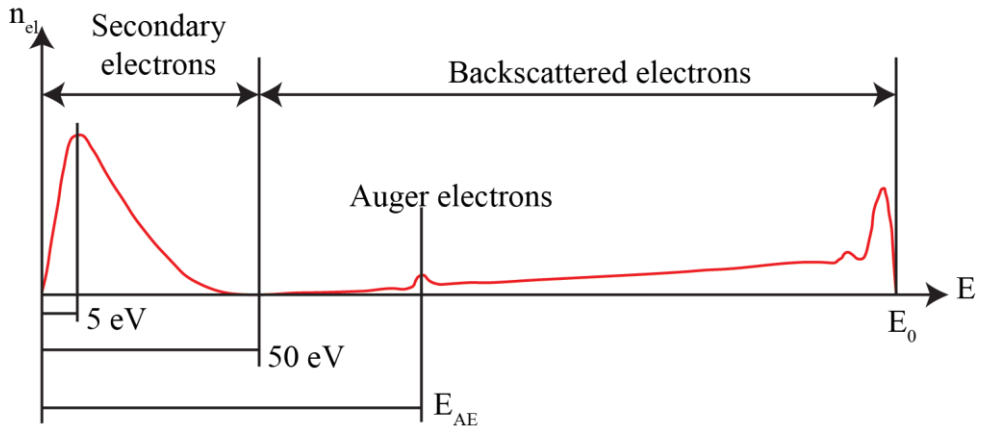

Fig. 3.1.5: energy distribution of the different types of electrons emitted from a surface.

SEs are electrons from within the sample which are emitted due to the inelastic scattering of the beam along its trajectory inside the material. Due to the low energies involved in the process, an electron can be extracted from the sample only if it is located close to the surface. This can be viewed in *Fig. 3.1.3*, where SEs are produced even in the bulk, but they do not have enough energy to escape from the sample (there is a big difference between t_{SE} and t_{BSE}). The SE distribution shows a peak at low energies with a most probable energy of 5 eV, as it is possible to see from *Fig. 3.1.5*, while the maximum energy of SE amounts to 50 eV. SEs generated by the primary incident beam electrons are designated SE1. The SE1 carry local information about the small cylindrical volume that is given approximately by the cross section of the beam $\left(\frac{\pi}{4}\right) d_{PE}^2$ and the escape depth t_{SE}. For a beam diameter of about 1 nm the SE1 deliver high lateral resolution information.[87] The beam electrons which are scattered multiple times and emerge as BSE, also generate secondary electrons within the escape depth: the latter are designated as SE2. Their origin is far from the point of incidence of the beam caused by the spatial dispersion of BSE. Changes of the amount of SE2 correlate with corresponding changes of BSE, thus SE2 carry information about the volume from which the BSE originate. The size of the volume depends on the penetration depth (or electron range) R and is much larger than the excitation volume of the SE1 for electron energies E_0 > 1keV (*Fig. 3.1.5*). Thus, SE2 deliver low lateral resolution information. Analogously to the BSE case, the SE yield (δ) can be introduced, which represents the number of SE generated by each impinging electron of the main beam. *Fig 3.1.4* shows that there is no monotonic relation between Z and δ. Instead, δ increases for increasing incidence angle, according to

$$\delta(\theta) = \frac{\delta(\theta = 0)}{cos\theta} \tag{3.1.4}$$

The relation presented in *Eq. 3.1.4* is valid for $E_0 \geq 5$ keV, and θ up to a few degrees below 90°. The angular dependence of the SE yield on θ is therefore at the origin of the topographic contrast in secondary electron images.

3.1.4 X-rays

X-rays are generated by the deceleration of the impinging electrons (X-ray continuum or Bremsstrahlung) or by electron transitions from a filled higher state to a vacancy in a lower

electron shell (characteristic X-ray lines; a typical X-ray spectrum is represented in *Fig. 3.1.6*). The former contribution yields a continuous distribution of intensity as a function of the photon energy, while the latter transitions generate a series of peaks at discrete and well defined element-specific energies. These features can be thus exploited as elemental fingerprints for a chemical characterization of the sample. Concerning the continuum, maximum of the distribution shifts towards higher energies with the impinging electron energy. According to Kramers,[91] the intensity of the continuous spectrum I_C emitted in an energy interval with the width dE_x is given by

$$I_C(E_x)dE_x = \frac{kZ(E_0 - E_x)}{E_x}dE_x \qquad (3.1.5)$$

where *k* represents the Kramers constant, which slightly varies with the atomic number.[92] A detailed treatment of the continuum X-rays emission is out of the scope of this thesis and can be found in the literature.[93]

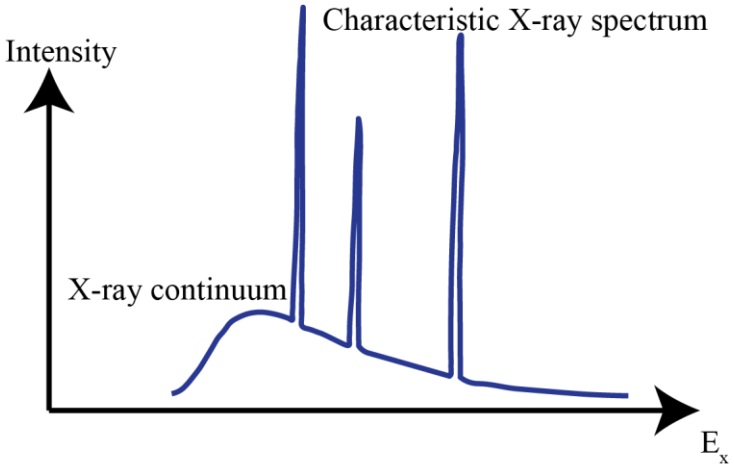

Fig. 3.1.6: X-ray spectrum generated by impinging electrons. The continuum (lower E_x) can be distinguished from the characteristic peaks (well defined E_x).

The line positions are instead independent on the energy of the incident electrons, being defined by the internal electronic structure of the sample. The generation of characteristic X-rays can be modeled in three different steps (*Fig. 3.1.7*). First, a beam electron interacts with an inner shell electron of an atom inducing its ejection, thus leaving the atom in an excited state. Second, the excited atom relaxes to the ground state through a transition of an electron from an outer shell to the inner shell vacancy. The characteristic energy difference ΔE_{ch} between the involved shells is specific for the atomic species. Third, the excess energy is released either by the emission of an electron from an outer shell (Auger electron) or by the emission of a characteristic X-ray photon with energy $E_x = \Delta E_{ch}$. The fraction of characteristic X-rays emitted when an electron transition occurs is given by the fluorescence yield ω. The latter increases with the atomic number and depends on the involved inner electron shell. The complementary yield, 1 - ω, defines the fraction of transitions that generate Auger electrons. The yields for the different shells and subshells can be calculated.[94] The characteristic X-ray lines are classified on the basis of the involved transitions.

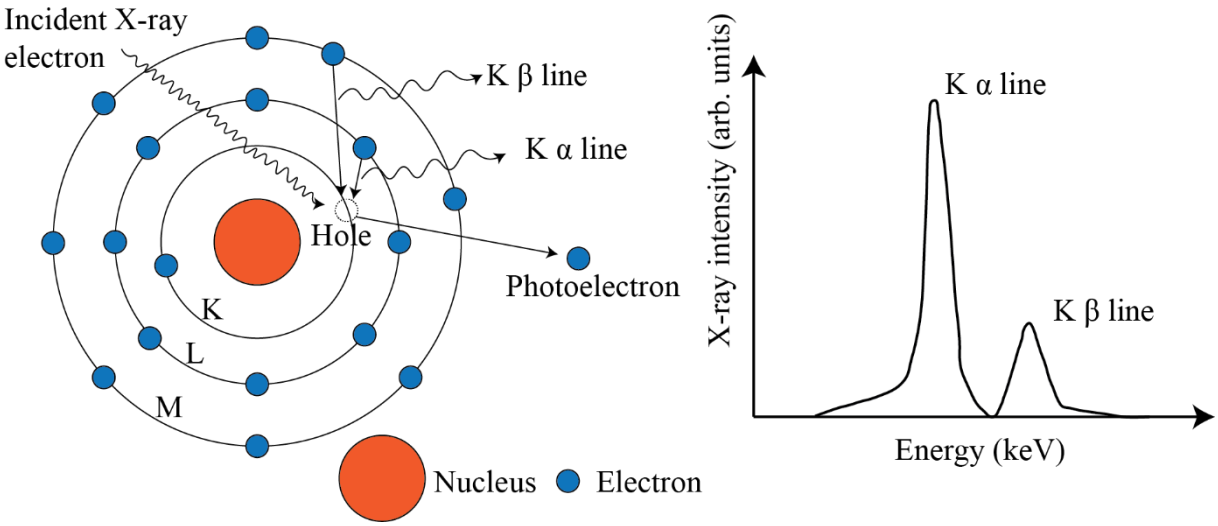

Fig. 3.1.7: example of characteristic X-ray production from peculiar electron transition.

3.1.5 Instrumentation

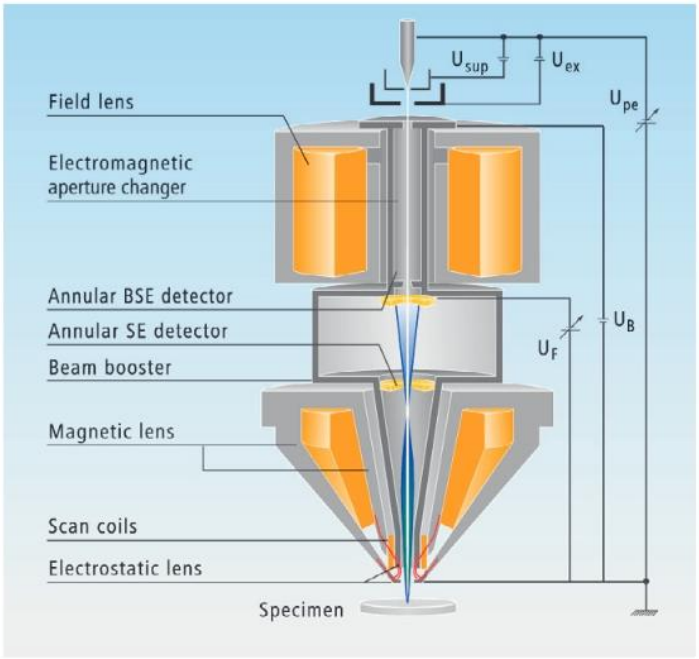

Fig. 3.1.8: schematic representation of a Scanning Electron Microscope. Figure adapted from the literature.[95]

The main components of the Scanning Electron Microscope are the electron column, the sample stage, and the detectors.[96] A schematic representation of a SEM is shown in *Fig. 3.1.8*. The column contains the electron gun, the lenses and the scanning coils; here the PE beam is created and focused by thermionic or field emission. By means of field and magnetic lenses it is possible to control and manipulate the physical size of the electron beam: the diameter of the beam hitting the sample's surface defines the lateral resolution achievable during the measurements. SEM images are produced by rastering the electron beam on the sample: its sweep is obtained thanks to the scanning coils. The objective lens is the last and most important lens in the SEM, allowing a finer control of the electron beam and defining the final crossover

of the impinging electron trajectories. Moreover, it is designed to screen the magnetic field in the measurement chamber. Nominal operating parameters are 0.5 – 5 nm for the beam diameter with an electron current in the 10^{-12} – 10^{-10} A range.[97]

The sample is mounted on a metallic holder, sitting on a positioning and alignment stage. The working distance (WD) can be modified to adjust the focus. Two pumping systems maintain optimal vacuum conditions in the column (UHV, 10^{-10} mbar) and in the experimental chamber (HV, 10^{-7} mbar).

3.1.6 Experimental Setup

All the images presented in this thesis were acquired using a ZEISS SUPRA 40 high resolution Field Emission Gun (FEG) SEM, located in the TASC laboratories of the CNR-IOM in Trieste (*Fig. 3.1.9*). The microscope can operate with a 4 pA – 10 nA probe current and 0.1 – 30 kV Electron High Tension (EHT), with a nominal resolution of 1.3 nm at 15 kV. It is equipped with a Back-scattered Electrons Detector, an Everhart-Thornley Secondary Electrons Detector, and a High efficiency In-lens detector, the latter providing an excellent signal-to-noise ratio in image acquisition. The microscope is also equipped with an Oxford Aztec Energy Dispersive X-ray Spectroscopy (EDS) system and an X-act 10 mm Silicon Drift Detector (SDD) for compositional analysis.

Fig. 3.1.9: experimental setup of the CNR-IOM ZEISS SUPRA 40 SEM used in this thesis.

3.2 Atomic Force Microscopy

Atomic Force Microscopy (AFM) is a technique based on a mechanical probe, yielding information on the topography of a surface with nanometer resolution. The microscope can operate in different environments like air, liquid, or vacuum, and can scan materials surfaces irrespectively of conductivity or opaqueness issues.[98] It was invented in 1986 by Binnig and his coworkers and it works by monitoring the force between the sample surface and a probe made by an elastic cantilever with a sharp tip at the end (schematic representation in *Fig. 3.2.1*).[99]

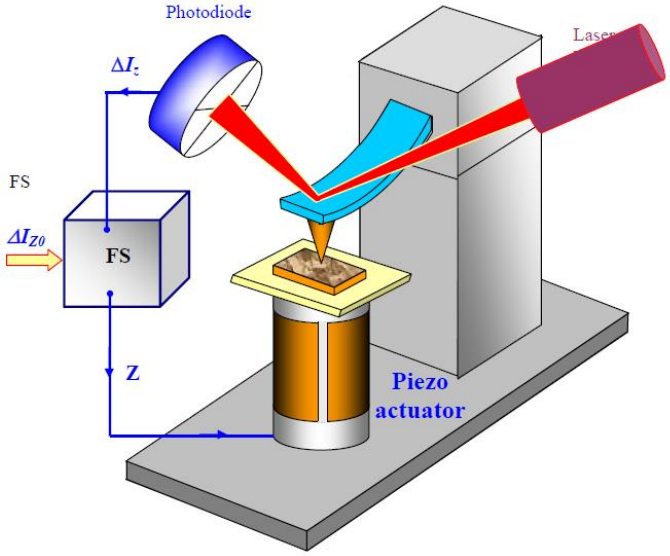

Fig. 3.2.1: schematic representation of the Atomic Force Microscope. Here the laser, the cantilever, the tip, the photodiode, the feedback loop (or feedback system FS), and the piezo actuator are visible. Figure adapted from the literature.[98]

By measuring the cantilever deflection caused by the tip-sample interaction, it is possible to get important information on the topography of the surface. All the details regarding this technique will be presented in the following subsections, starting from a description of the main components of the microscope and then introducing its operating modalities.

3.2.1 Cantilever and tip

The force between the tip and the sample surface is monitored through the small deflections of the microcantilever, whose typical dimensions are ~100 μm in length, tens of μm in width and few μm in thickness. The cantilever is mounted on a holder (typical size ~ mm), while on the opposite side a sharp tip (tip radius ~ nm)[100] is installed. Typical materials used for cantilever manufacturing are silicon and silicon nitride (Si_3N_4), with its rear side covered by a thin layer of a highly reflective material (for example aluminum or gold) to reflect efficiently the laser beam directed at it, as it will be discussed in the following. The response of the cantilever to the tip-sample interaction can be described assuming the linear elasticity regime, which holds for small deformations only. With this assumption, the stress σ depends linearly on the strain ε:

$$\sigma = E \cdot \epsilon \qquad (3.2.1)$$

where E is the Young's modulus. The cantilever motion can be well approximated in terms of Hooke's law:

$$F = -k_c \cdot d \tag{3.2.2}$$

where F is the loading force, k_c is the cantilever spring constant and d is the deflection. Knowing the cantilever's Young's modulus and its dimensions, k_c can be obtained by the following expression, used to describe rectangular cantilevers:

$$k_c = \frac{\omega E}{4}\left(\frac{t}{l}\right)^3 \tag{3.2.3}$$

with l, ω, t being the cantilever's length, width, and thickness, respectively. To be precise, the above mentioned expression gives the vertical force constant of the cantilever, i.e. k_{ver} , with the acting force and cantilever deformation both being in the vertical direction too; a more general treatment is reported in the literature.[101]

Sometimes the shape of the cantilever is complex or one of the physical parameters to determine k_c using the formula presented above is unknown. However, it is possible to obtain the cantilever spring constant k_c from the total cantilever deflection (Δz) caused by thermal fluctuations.[102] If the cantilever is modeled as a harmonic oscillator in thermodynamic equilibrium, the mean-square displacement of the cantilever tip from its neutral position is described by *Eq. 3.2.4*,[103]

$$< \Delta z^2 > = \frac{k_B T}{k_c} \tag{3.2.4}$$

with k_B being the Boltzmann constant and T the temperature. Due to the cantilever's geometry, several vibrational modes are possible and can thus be excited (*Fig. 3.2.2*).

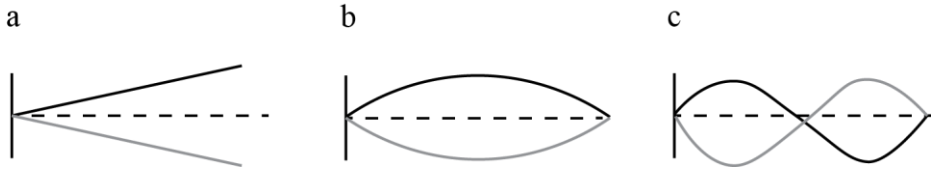

Fig. 3.2.2: example of the first three vibration modes for a rectangular cantilever.

Therefore, to obtain the mean square deflection of one mode (indicated as $< \Delta z^2 >$), a correction factor β_n must be employed

$$< \Delta z_n^2 > = \frac{k_B T}{k_c}\beta_n. \tag{3.2.5}$$

The deflection is usually sensed using the optical lever technique, which is in fact a measure of the inclination and not of the deflection. Therefore, if we are interested in the forces acting at the end of the cantilever, another correction factor β_n^* (depending on the shape and on the material of the tip, tabulated in literature)[102] should be introduced. The values of k_c is finally given by *Eq. 3.2.6*

$$k_c = \frac{\beta_n^* k_B T}{< \Delta z_n^{*2} >}. \tag{3.2.6}$$

In practice, the effective cantilever deflection when it is oscillating far from the surface is recorded, and the power spectral density (PSD) of the temporal signal is computed, thus obtaining the resonant modes of the thermally excited cantilever. Then, both the first resonant peak and the area under the curve (which is equivalent to $< \Delta z_n^{*2} >$) are fitted using a lorentzian function. At this point the cantilever spring constant k_c can be assessed using the equation above, with the correction factor β_n^* corresponding to the cantilever geometry. A great number of different tip shapes is available (examples shown in *Fig. 3.2.3*), and the choice depends on the sample's properties and on the physical quantity to be measured. The same holds for the choice of the k_c. Tipless cantilevers are used as templates to attach desired functional probes (molecules, nanotubes, cells), spherical or semispherical tips are usually exploited in force spectroscopy studies, while tetrahedral and sharp tips are employed in topography measurements, force spectroscopy, and nanolithography studies.[104]

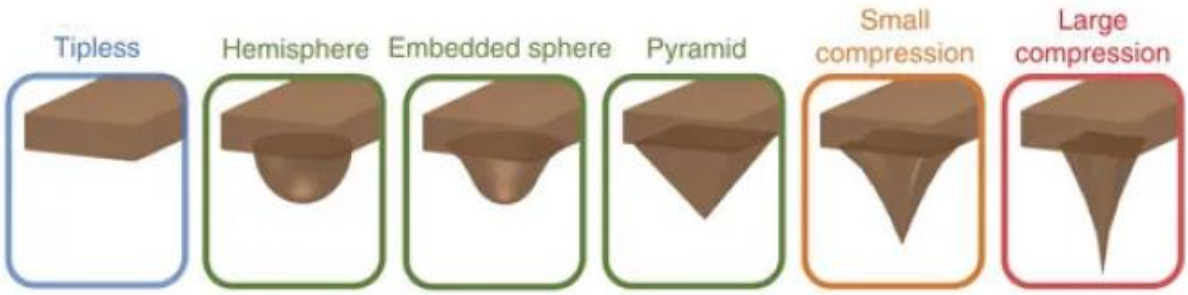

Fig. 3.2.3: examples of different shapes for the Atomic Force Microscope tip. Figure adapted from the literature.[105]

3.2.2 Laser and photodetector

Two other main components of an Atomic Force Microscope are the laser and the photodetector. The laser is used to track and magnify the nanometric motions of the cantilever, exploiting the optical lever principle. The laser beam is reflected from the back side of the cantilever onto the four quadrants photodetector, as represented in *Fig. 3.2.4*.[106]

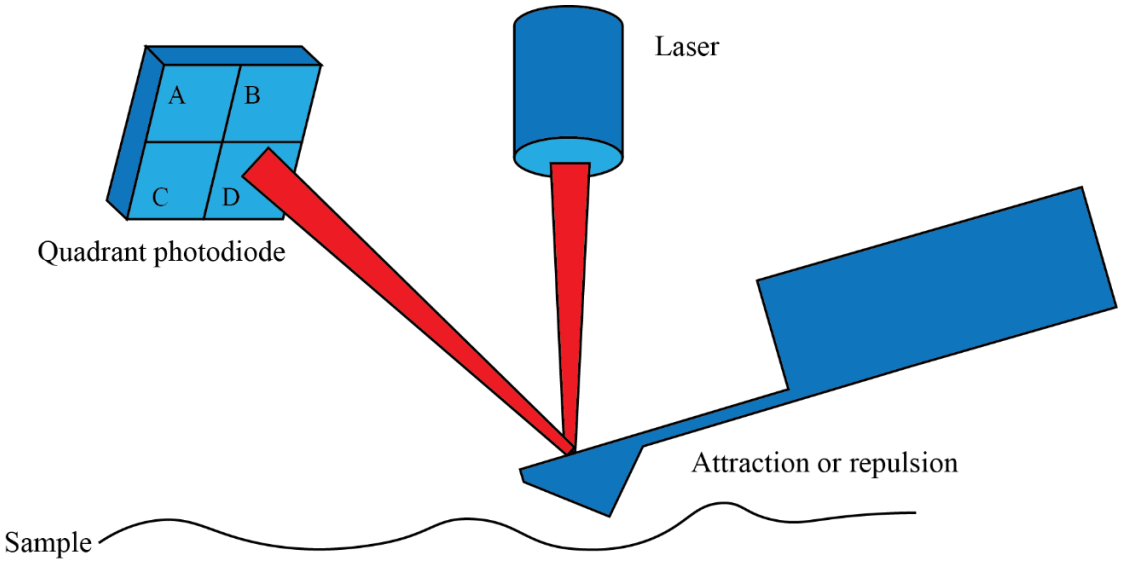

Fig. 3.2.4: scheme of the AFM working principle: the surface is scanned with the tip placed at the end of the cantilever rastering the selected region. The laser light is reflected by the back of the cantilever onto a four quadrants photodiode, capable to obtain the deflection signal and thus reconstruct the surface topography.

When the probe interacts with the sample, the change in the light's path is detected by the photodetector and thus the tip-surface interaction can be indirectly measured. The photodetector's four quadrants allow the measurement of both the reflected beam intensity and position. The voltage detected in each region of the photodetector is proportional to the reflected intensity of the laser; the horizontal and vertical displacement of the beam due to cantilever deflection are obtainable as:

$$X = \frac{(V_A + V_C) - (V_B + V_D)}{V_A + V_B + V_C + V_D} \tag{3.2.7a}$$

$$Y = \frac{(V_A + V_B) - (V_C + V_D)}{V_A + V_B + V_C + V_D} \tag{3.2.7b}$$

with the sum $(V_A + V_B + V_C + V_D)$ giving the voltage corresponding to the total reflected intensity. Before every measurement, the null condition for both horizontal and vertical motion detection must be set: the null condition states that left and right, and top and bottom halves of the photodiode must read the same voltage values, meaning the central positioning of the laser spot on the photodiode is reached.[107]

Another key parameter associated with the optical lever sensor is its sensitivity. The need for such parameter comes from the fact that the photodiode's output is the voltage *V*, not the relative deflection of the cantilever *d*, which is of greater scientific interest. *d* is proportional to the output voltage *V*, with the proportionality factor *s*, called the optical lever sensitivity and measured in μm/V:

$$d = s \cdot V \tag{3.2.8}$$

The sensitivity calibration is obtained by performing the force curves on a clean, mechanically rigid substrate (e.g. glass).

3.2.3 Piezoelectric scanner

The piezoelectric scanner is the part of the AFM allowing the nanometric controlled displacements of the cantilever and thus performing the topographic analysis of the surface. Piezoelectric materials exhibit a measurable voltage difference when applying an external force in a well-defined direction (direct piezoelectric effect), because the electric dipole moments in a crystallographic unit cell are aligned and produce a net electric polarization. The atomic dipole moment arises due to an asymmetric shift in the atomic positions in the unit cell of a crystal; thus, a symmetric load from all directions would result in no voltage. The direct piezoelectric effect is described by:

$$P_i = d_{ijk}\sigma_{jk} \tag{3.2.9}$$

where P_i is the polarization vector and σ_{jk} is the the stress tensor. The coefficient d_{ijk} is called the piezoelectric coefficient. Conversely, if an external electric field E_k is applied, a strain ϵ_{ij} is produced (indirect piezo effect):[108]

$$\epsilon_{ij} = \tilde{d}_{ijk}E_k \tag{3.2.10}$$

Typical piezo elements are made of piezoceramics containing lead barium titanate, $PdBdTiO_3$ or lead zirconate titanate, $Pb[Zr_xTi_{1-x}]O_3$, $0 < x < 1$.

3.2.4 Feedback loop

The feedback loop is fundamental for the data acquisition, since it permits to monitor and control the output of the system. The key parameter is represented by the error signal $E(t)$, which is the difference between the measured signal coming from the detector $Q_{measured}(t)$ and a set-point value $S_{set\ point}(t)$ set by the operator,

$$E(t) = S_{set\ point}(t) - Q_{measured}(t). \tag{3.2.11}$$

Based on $E(t)$, the feedback loop supplies the correct voltage to adjust the piezoelectric scanner in order to reduce $|E(t)|$, ideally reaching the condition $E(t) = 0$. The type of feedback control used in AFM is called a proportional-integral-derivative controller (PID), described as:

$$E_{adjusted}(t) = P \cdot E(t) + I \cdot \int E(t')dt' + D \cdot \frac{dE(t)}{dt} \tag{3.2.12}$$

where $E_{adjusted}(t)$ is the adjusted error signal, while P, I, D are the proportional, integral and derivative parameters, respectively. With the appropriate choice of the parameters P, I, D the probe will follow the sample's topography, keeping $E(t)$ at minimum. The integral I term facilitates the probe moving over large surface features and the P and D terms allow the probe to follow the smaller, high-frequency features on a surface. Any feedback control system has its inherent bandwidth, which means that due to its nonzero response time the feedback loop might fail to make adjustments fast enough if the measured quantity $Q_{measured}(t)$ is varying rapidly with time. [109]

3.2.5 Operation modes

The AFM has three main operation modes, denominated contact, non-contact and tapping mode. They will be introduced in the following, but particular attention will be given to the tapping mode, being the one used to acquire the AFM data presented in this thesis.

In the *contact mode* of operation, the cantilever is moved towards the sample until it physically touches it. In this case, the cantilever-sample interaction leads to a bending of the cantilever due to repulsion forces. The height position of the translation stage is controlled by a feedback loop in order to maintain the force between the tip and the sample constant while scanning the surface. Typically, the cantilevers exploited in the contact mode are made of silicon nitride, thus making them extremely flexible. This operating mode creates high lateral forces, capable of altering the surface and perturbing the measurements.

In the *non-contact mode*, the tip never physically touches the surface of the sample. In this case, the cantilever oscillates at a frequency close (or equal) to the resonant one. Forces from the surface of the samples (e.g. Van der Waals) interfere with the cantilever trying to modify its oscillation frequency, which is kept constant by the feedback loop system by adjusting the tip-sample distance. In this case, stiffer (with respect to the contact mode) tips have to be exploited, because there is no risk in damaging the sample. Lateral forces are avoided, while as a drawback an instability caused by working in the attractive regime is present.

The *tapping mode* (usually denoted as semi-contact mode or intermittent contact mode) is a compromise between contact and non-contact modes. In this technique, the cantilever is forced to oscillate near resonance frequency with an amplitude about 1 – 100 nanometers.

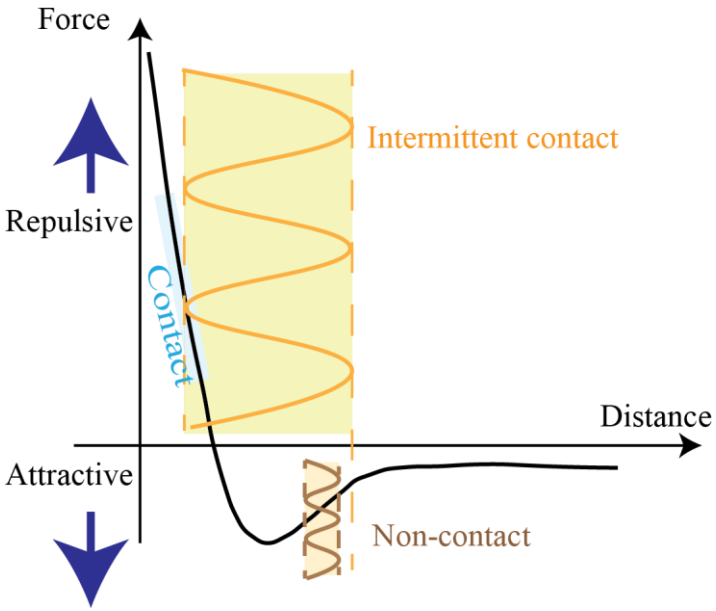

Fig. 3.2.5: Graph representing the force between the microscope tip and the sample surface. For small distances, the tip physically touches the sample (contact regime) experiencing a repulsive force, while for increasing distances the regime shifts to the non-contact and the tip experiences an attractive force.

These oscillations cause the cantilever to tap along the surface, getting in contact with the sample surface in the lower semi-oscillation, which corresponds to the repulsive region in the force-distance diagram (*Fig. 3.2.5*). The tip-sample interaction influences the oscillation

amplitude of the tip, which is kept constant by means of the feedback loop acting on the position of the cantilever above the sample. The operating mode just described prevents the tip from damaging the sample surface. Due to the nature of this approach and to the high frequency interactions with the sample, the tips employed must be stiffer (usually made of silicon) than the ones used for the contact mode. Similarly to the contact mode, tapping mode exploits the contact region of the force potential curve, but lateral forces are much reduced, decreasing the risk of sample damage.[106] During scanning, the changes of amplitude and phase of cantilever oscillations are recorded.

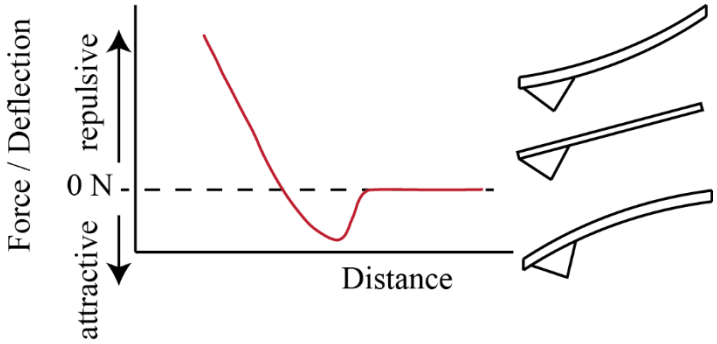

Fig. 3.2.6: scheme representing the tip-sample interaction in the tapping mode: the range goes from the repulsive region, where the tip is in contact with the surface, to the attractive region where the tip is not touching the sample. These interactions cause different cantilever bending highlighted in the right side of the figure.

In the tapping process, tip-sample interaction passes from the attractive regime, through the 'zero-force' regime, and into the repulsive regime, as shown in *Fig. 3.2.6*, leading some important advantages: the first is that due to the movement of the tip (perpendicular to the surface) lateral forces, which can cause great problems in contact-mode, are almost eliminated, while the fundamental instability of non-contact AFM (operating in attractive regime) is overcome. The second advantage is represented by the absence of the influence of the contamination layer, which is a thin film present on most surfaces in air due to the direct exposure of the sample to the environment condition. This region is not representative of the sample (see *Fig. 3.2.7*) and by means of tapping mode it is possible to avoid its influence on the data acquisition. In fact, the restoring force of the cantilever withdraws the tip from the contamination layer in each cycle, thus reducing the effect of capillary forces on the image.

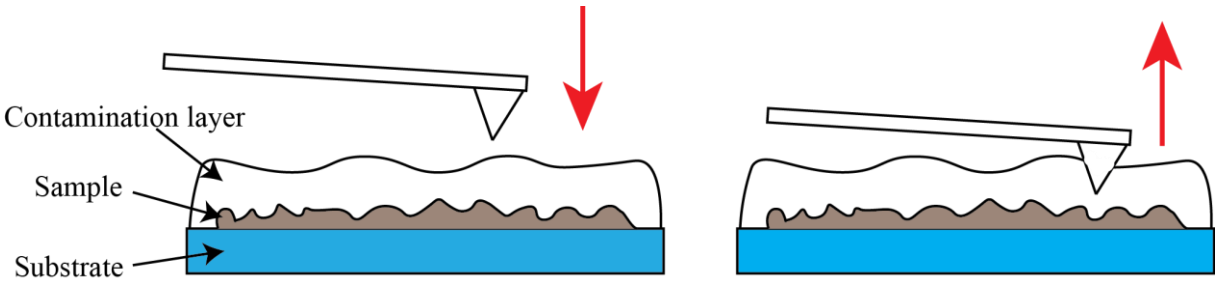

Fig. 3.2.7: representation of the measurement process of the tapping mode AFM: during the scan, the tip reaches the surface of the sample with vertical movements being not sensitive to the contamination layer.

In most cases, the probe is oscillated by an additional piezoelectric element attached to the probe holder. Often, in addition to the amplitude signal (main signal acquired in tapping mode), the delay in the phase of the probe oscillation is recorded (*Fig. 3.2.8*). In a similar way to deflection in contact mode, the amplitude signal in intermittent contact may be used as an illustration of the shape of the sample.

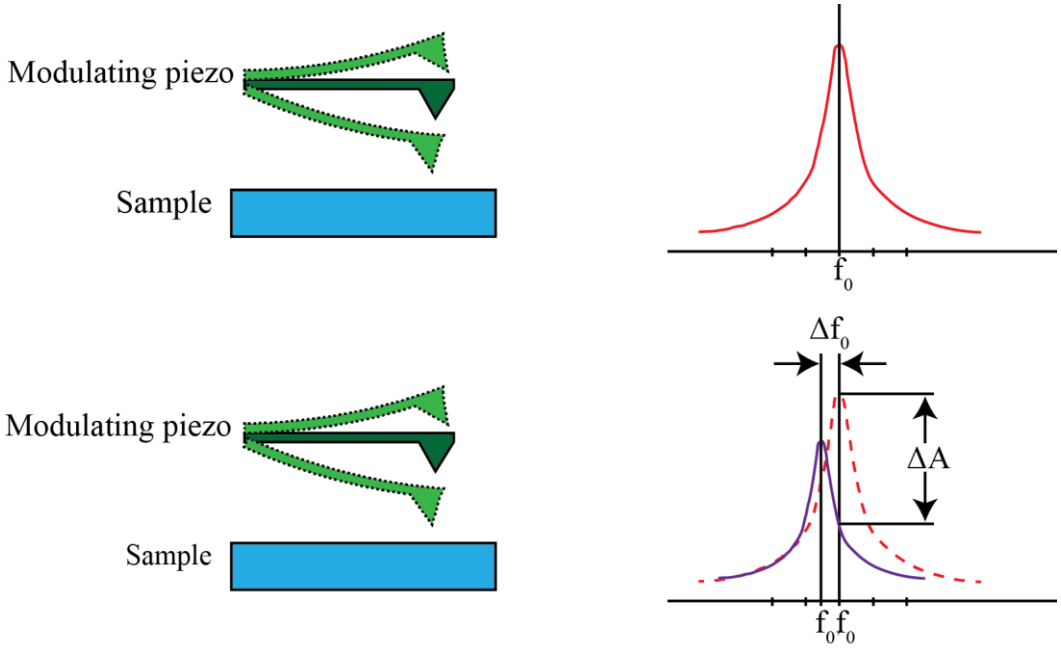

Fig. 3.2.8: differences in the amplitude and phase shift signals characterizing the tapping mode.

In order to describe quantitatively the tapping mode, at a first glance the AFM cantilever can be interpreted as a point-mass spring acting like a harmonic oscillator. In the absence of tip-surface forces, i.e. $F_{ts}(z) = 0$, the general equation is the one of a forced harmonic oscillator with damping. The dependence of the amplitude with the excitation frequency can be calculated by the Lorentzian expression

$$A(\omega) = \frac{F_0/m}{\left[(\omega_0^2 - \omega^2)^2 + (\frac{\omega \cdot \omega_0}{Q})^2\right]^{1/2}} \qquad (3.2.13)$$

where ω_0 is the natural frequency of the oscillation and F_0 its related elastic force, Q is the quality factor for the system ($Q = \frac{\omega_0 m}{\gamma}$) with γ taking into account the viscous force in air (damping). The phase shift can be expressed as

$$\varphi = atan\left(\frac{\frac{\omega \cdot \omega_0}{Q}}{\omega_0^2 - \omega^2}\right). \qquad (3.2.14)$$

The damping modifies the resonance frequency of the cantilever. Resonant and natural (free resonance) frequencies are related by *Eq. 3.2.15*

$$\omega_r = \omega_0\left(1 - \frac{1}{2Q^2}\right)^{1/2}. \qquad (3.2.15)$$

Assuming that the cantilever (harmonic oscillator) is under the influence of a parabolic tip-surface interaction potential, then the total force F acting on the tip includes the elastic response $k_c \cdot z$ and the tip-sample interaction force F_{ts}. For small displacements with respect to the equilibrium position, the total force can be expressed by

$$F = F_0 + \left(\frac{dF}{dz}\right)_{z_0} (z - z_0) \tag{3.2.16}$$

with

$$k_e = -\frac{dF}{dz} = \left(k_c - \frac{dF_{ts}}{dz}\right)_{z0} \tag{3.2.17}$$

then, the new effective resonance frequency is calculated by

$$\omega_e = \left(\frac{k_c - (dF_{ts}/dz)}{m}\right)^{1/2}. \tag{3.2.18}$$

This shows that the resonance frequency of a weakly perturbed harmonic oscillator depends on the gradient of the interaction (in the AFM case, the tip-sample interaction). The dependence of the amplitude on the excitation and effective resonance frequencies introduces a possible mechanism to explain the dependence of the oscillation amplitude with the strength of the interaction force. Assuming the tip is excited at the natural frequency, the approach of the tip towards the surface will modify the resonance frequency, which implies a modification of the oscillation amplitude. The actual amplitude will be given by the value of the new resonance curve at the excitation frequency ω_e of the oscillator. As a consequence, the new oscillation amplitude would be smaller than the free amplitude. Harmonic models are useful to provide some of the concepts that appear in AFM experiments, however in most cases they fail to provide quantitative agreement with the experiment. A quantitative discussion on the limitations of the harmonic approach to interpret dynamic AFM experiments is provided in the literature.[110]

One important approach was proposed by San Paulo and Garcia.[111] They have applied the virial theorem $< K >= -\frac{1}{2} < Fz >$ and the consideration that at the steady-state the average rate at which energy is supplied to the tip must be equal to the average rate at which energy is dissipated by hydrodynamic and tip-surface forces. Those assumptions have allowed to derive a general relationship for the oscillation amplitude

$$A \approx \frac{A_0}{\sqrt{2}}\left(1 - \frac{2P_{ts}}{P_{med}} \pm \sqrt{1 - \frac{4P_{ts}}{P_{med}} - 16\left(\frac{< F_{ts}z >}{F_0 A_0}\right)^2}\right)^{1/2} \tag{3.2.19}$$

where P_{ts} and P_{med} are the average dissipated power in a cycle by tip-surface (P_{ts}) and hydrodynamic (P_{med}) forces (where the label *med* indicates the medium in which the measurement is performed). The formula states, without placing any restrictions on the nature of the tip-sample forces, that the amplitude reduction is controlled by the average interaction

force times the deflection and the average power dissipated by the tip-sample forces. The dependence on average quantities is a direct consequence of a tip motion that experiences different values of the tip-surface force per cycle.

Cleveland and co-workers have deduced an analytical relationship between the phase angle of the tip motion and the energy dissipated by the tip-surface forces.[112,113] Similar to the previous case, this model is based on the assumption that in the steady-state the average energy supplied to the cantilever per period must be equal to the average energy dissipated via hydrodynamic viscous interactions with the environment (E_{med}) and by inelastic interactions at the tip-sample interface (E_{ts}); E_{ext} is the energy supplied by piezo-vibrator during one oscillation period.

$$E_{ext} = \int_{t}^{t+2\pi/\omega} F_0 \cos(\omega t) \frac{dz}{dt} dt \tag{3.2.20a}$$

$$E_{med} = \int_{t}^{t+2\pi/\omega} \frac{m\omega_0}{Q} \left(\frac{dz}{dt}\right)^2 dt \tag{3.2.20b}$$

$$E_{ts} = \int_{t}^{t+2\pi/\omega} F_{ts}(z) \frac{dz}{dt} dt \tag{3.2.20c}$$

From the equilibrium condition it follows that

$$E_{ext} = E_{med} + E_{ts} \tag{3.2.21}$$

and assuming stationary oscillations $z = u_0 cos(\omega t + \varphi)$

$$E_{ts} = E_{ext} - E_{med} = \frac{\pi k A u_0}{Q} \sin(\varphi) - \frac{\pi k \omega A^2}{\omega_0 Q} \tag{3.2.22}$$

the phase shift can be obtained from the following equation:

$$sin\varphi = \frac{\omega A}{\omega_0 u_0} + \frac{Q E_{ts}}{\pi k A u_0} \tag{3.2.23}$$

Thus, the cantilever oscillations phase shift in "semi-contact mode" is determined by the amount of dissipative tip-sample interaction. A closer examination of the previous equation reveals the presence of two components, the elastic $\left(\frac{A}{u_0}\right)$ and the inelastic contributions (Q and E_{ts}). The phase imaging in tapping mode is performed at fixed set point amplitude $A = const$. This means that the elastic term remains constant while imaging. Consequently, changes in the phase shift could only come from changes in the energy dissipated by the tip-sample forces.

3.2.6 Experimental setup

In this thesis work, an AFM Solver Pro-M (NT-MDT, Russia, *Fig. 3.2.9*) was used to characterize the films in terms of thickness and surface roughness. The AFM was equipped with a NSG01 tetrahedral Si tip with nominal resonance frequency in the 87 – 230 kHz range. The microscope was operated in tapping mode and all measurements were carried out in air at room temperature. Regarding the experimental use of the AFM, in this thesis two different type of data have been collected: information about the thickness of the layers and information regarding the surface roughness, following the process illustrated in *Section 6.2*.

Fig. 3.2.9: image of the AFM Solver Pro-M used to acquire the AFM data.[114]

3.3 Synchrotron-radiation X-Ray Computed Tomography

Generally speaking, CT is an imaging technique consisting on a computer-processed acquisition of a 3D reconstruction of any object. Many typologies of CT are known, as for example computed axial tomography (CAT), single-photon emission computed tomography (SPECT) or positron emission tomography (PET), but in this thesis Synchrotron-radiation X-Ray Computed Tomography has been specifically exploited. CT data presented in this work were collected at the SYRMEP beamline of the Elettra Synchrotron light source in Trieste, using a Phase-contrast Imaging approach, and in the specific by means of Propagation-based Phase-Contrast Imaging (PPCI). A full overview of Phase-contrast Imaging can be found in the literature.[115] To understand better all the physics behind the CT technique, it is important to introduce a set of concepts regarding X-ray interaction with matter and to explain in detail the PPCI.

3.3.1 Coherence

An imaging system that exploits the coherence of the illuminating radiation can produce observable effects of interference and diffraction. In particular, temporal (or longitudinal) coherence is related to the energy spectrum of the source (the narrower the spectral bandwidth, the higher the coherence) and spatial (or lateral) coherence is related with the source size and geometry (the smaller the source size and the larger the source-to-sample distance, the higher the coherence). A hypothetical system with a monochromatic point-like source is perfectly spatially coherent, while the degree of coherence of a real source depends on how well it approximates monochromaticity and point-like size. When dealing with partially coherent sources, it may be helpful to introduce the coherence lengths. The temporal (or longitudinal) coherence length $L_t = \frac{\lambda^2}{\Delta\lambda}$, with $\Delta\lambda$ being the spectral bandwidth, defines the length up to which the beam can be considered coherent along the propagation direction. On the other hand, the spatial coherence length $L_s = \frac{\lambda z_1}{S}$, where z_1 is the source to sample distance and S is the source size, defines a scale for the transverse coherence perpendicular to the beam propagation direction.[116] Synchrotron radiation facilities are examples of moderately coherent X-ray sources, so that this property of their radiation can be exploited, at least practically, for imaging purposes, while Free-Electron Laser (FEL) sources are largely coherent, yielding phase-sensitive diffraction-based imaging capabilities (speckles).

3.3.2 X-ray imaging

In X-ray absorption-based imaging, the contrast between different materials is generated by differences in density, thickness of the sample and X-ray absorption coefficient. The interaction between X-rays and matter can be described exploiting the complex refractive index, whose dispersion corrections (δ and β) are related to scattering and absorption of X-rays, respectively. The complex refractive index for X-rays can be thus written as [117]

$$n = 1 - \delta + i\beta \tag{3.3.1}$$

Suppose now X-rays going through a sample, as depicted in *Fig. 3.3.1*.

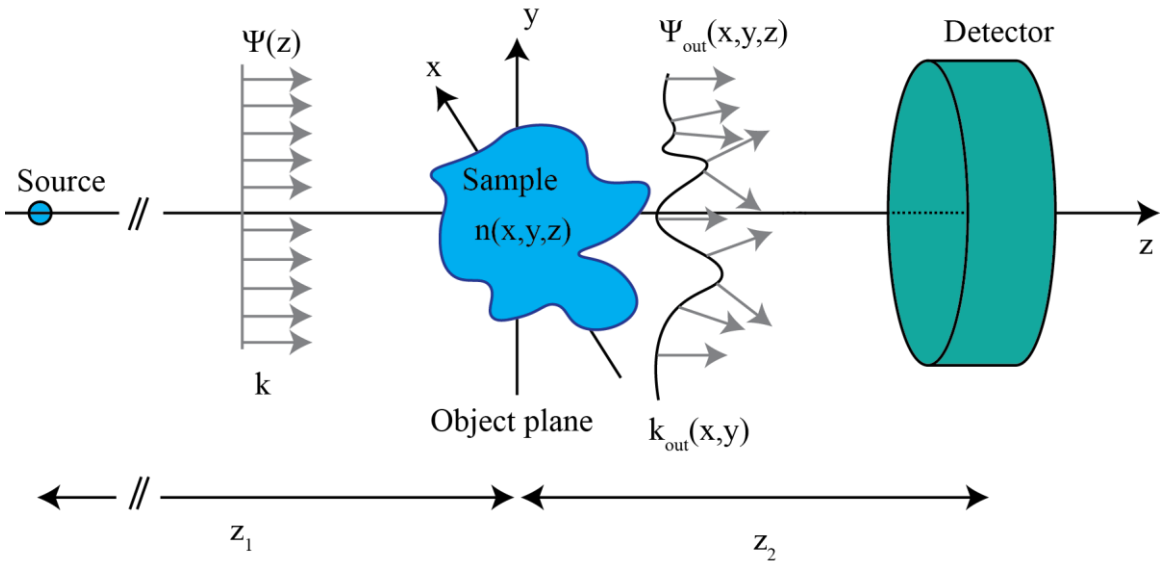

Fig. 3.3.1: schematic description of the X-ray wave interaction with the sample. The plane wave front is modified by the specimen and this modification gives important information regarding the transmission function of the sample itself.

The object can be described by means of a three-dimensional (3D) distribution of the complex refractive index

$$n(x,y,z) = 1 - \delta(x,y,z) + i\beta(x,y,z) \tag{3.3.2}$$

The X-ray wave $\Psi_{out}(x,y,z)$ emerging from the point (x,y) of the object can be written as a product of the incident plane wave $\Psi(z)$ (parallel and monochromatic X-ray beam traveling in vacuum along the z axis) and the sample's transmission function $T(x,y)$

$$\Psi_{out}(x,y,z) = T(x,y)\Psi(z) = T(x,y)\Psi_0 e^{ikz} \tag{3.3.3}$$

where $\Psi(z)$ has been splitted in amplitude (Ψ_0) and phase (e^{ikz}) contributions and $T(x,y)$is a complex function that can be written as

$$T(x,y) = A(x,y)e^{i\Phi(x,y)} \tag{3.3.4}$$

with

$$A(x,y) = e^{-k\int \beta(x,y,z)dz} \tag{3.3.5}$$

and

$$\Phi(x,y) = -k\int \delta(x,y,z)dz \cong -r_0\lambda\int \rho_e(x,y,z)dz \tag{3.3.6}$$

where k represents the wavenumber of the impinging light and the integrals are evaluated over the full extension of the object along the z direction.[118] The last expression was evaluated assuming the X-ray energy is higher than a given elemental absorption edge. In this case, the classical electrodynamics equations can be exploited, obtaining

$$\delta \cong \frac{r_0 \rho_e \lambda^2}{2\pi} \tag{3.3.7}$$

being r_0 the classical electron radius and ρ_e the electron density. Thus, the object affects the X-ray beam both in amplitude and in phase. The amplitude modulation can be linked to the so-called Lambert-Beer law

$$I_{out} = I_0 e^{-\int \mu(x,y,z)dz} \tag{3.3.8}$$

where I_0 is the incident beam intensity, I_{out} the intensity of the outgoing beam and $\mu(x,y,z)$ the linear attenuation coefficient. The amplitude attenuation given by *Eq. 3.3.5* corresponds to a reduction in intensity of

$$A^2(x,y) = e^{-2k\int \beta(x,y,z)dz} \tag{3.3.9}$$

which is equivalent to the Lambert-Beer law. Linking the two equations together, remembering $k = \frac{2\pi}{\lambda}$, gives the following relation between the linear attenuation coefficient μ and the imaginary part of the refractive index β:

$$\mu(x,y,z) = \frac{4\pi\beta(x,y,z)}{\lambda} \tag{3.3.10}$$

Alongside absorption, also a phase shift $\Phi(x,y)$ is introduced. As seen before, the phase shift $\Phi(x,y)$ is proportional to the refractive index decrement δ (*Eq. 3.3.6*), while absorption is related to the imaginary part β (*Eq. 3.3.5*). The wave coming out from the object will no longer be a plane wave traveling along the z direction due to the interaction with the object itself. Supposing the wave number is larger than the absolute values of the spatial derivatives, which means

$$\left[\frac{\partial\Phi(x,y)}{\partial x}\right], \left[\frac{\partial\Phi(x,y)}{\partial y}\right] \ll k \tag{3.3.11}$$

the wave vector $\boldsymbol{k}_{out}$ of the outgoing wave at the point (x,y) can be written as

$$\boldsymbol{k}_{out}(x,y) = \left[\frac{\partial\Phi(x,y)}{\partial x}\right]\hat{\boldsymbol{x}} + \left[\frac{\partial\Phi(x,y)}{\partial y}\right]\hat{\boldsymbol{y}} + k\hat{\boldsymbol{z}} = \boldsymbol{\nabla}_{xy}\Phi(x,y) + k\hat{\boldsymbol{z}}. \tag{3.3.12}$$

In other words, the wave leaving the object plane at the point (x,y) will be deviated by a refraction angle α with respect to the original direction z, where

$$\alpha \cong \frac{1}{k}\sqrt{\left(\frac{\partial\Phi(x,y)}{\partial x}\right)^2 + \left(\frac{\partial\Phi(x,y)}{\partial y}\right)^2} = \frac{1}{k}\left|\nabla_{xy}\Phi(x,y)\right|. \tag{3.3.13}$$

Conventional X-ray imaging is not capable of acquiring the phase shift $\Phi(x,y)$ information because X-ray detectors are sensitive to the intensity of the radiation only, and not to its phase. In order to highlight the phase shift and to use it as an additional information, it must be first converted in an intensity modulation. Several methods, typically based on coherent

X-ray sources, can be used to reveal the contrast hidden in the phase shift term $\Phi(x,y)$: they are the so called phase-sensitive X-ray imaging techniques. Among all the phase-sensitive techniques, in this thesis the attention will be focused only on the propagation-based phase-contrast imaging (PPCI) approach.

3.3.3 Propagation-based phase-contrast imaging

In PPCI technique, an X-ray beam irradiates the sample, giving rise to a spatially varying phase shift. As the beam passes through the object, parts of the wave front undergoing deflections interfere, producing a characteristic pattern, which is registered by a detector placed at a suitable distance.[119] Due to the Fresnel diffraction, the phase shifts are then converted into detectable intensity modulations. In this technique no optical elements between the sample and detector are required, avoiding aberration problems and decrements of spatial resolution.[120,121] A crucial dependence of the phase contrast signal on the effective propagation distance can be defined as

$$D = \frac{z_1 z_2}{(z_1 + z_2)} \tag{3.3.14}$$

where z_1 the source-sample distance and z_2 is the sample-detector distance, as reported in *Fig. 3.3.2*. For $z_1 >> z_2$, which is the normal operating condition for synchrotron facilities, the effective distance is determined by the propagation distance z_2. Taking into account the transverse size *h* of the sample, four different regimes can be defined:[122]

- Absorption regime: sample-detector distance is close to zero ($z_2 \approx 0$). It is also called contact region.
- Near-field diffraction regime: the effective propagation distance is such that $\lambda D \ll h^2$. In these conditions, the contrast is formed locally around specific object features. The boundaries of the object are strongly enhanced, and a distinct interference pattern corresponds to every edge, giving reliable information on the geometry of the object. To express the above condition for the near-field regime, a Fresnel number must be defined as $N_F = \frac{h^2}{\lambda D}$ so that $N_F \gg 1$.
- Fresnel regime: here $\lambda \approx h^2$ meaning $N_F \approx 1$
- Fraunhofer regime: the effective propagation distance is quite large, $\lambda D \gg h^2$ and so $N_F \ll 1$. In this condition, the interference fringes can be well detected but they cannot be related to a specific edge of the sample, so that the shape of the sample is difficult to be recognized.

In this thesis, all CT images have been acquired in a near-field diffraction regime. An example of the PPCI principle is shown in *Fig. 3.3.2*.

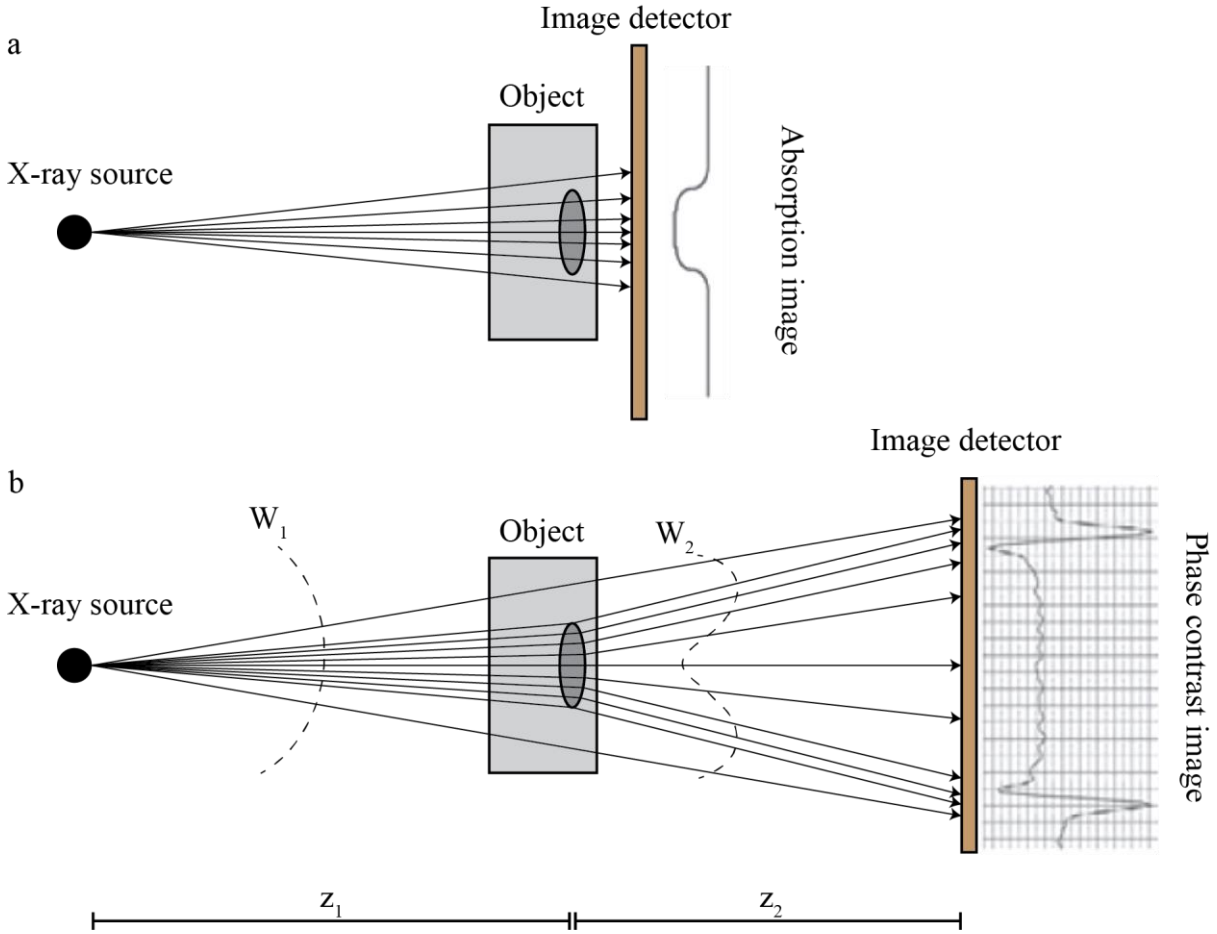

Fig. 3.3.2: example of PPCI principle: in panel (a) the detector is too close to the object and only the absorption component can be acquired, while in panel (b) the distance is large enough to acquire both the absorption and phase signals.

Fig. 3.3.2a shows the conventional configuration, in which the detector is placed close behind the object (contact region); the image is recorded where the contrast emerges from the selective absorption (or scattering at large angles due to the Compton effect) of structures inside the sample. In this case, the detail (represented as a dark oval-shaped) absorbs more X-rays than the background (grey rectangle), resulting in an intensity peak on the profile of the standard absorption image. On the contrary, if the source is sufficiently spatially coherent, PPCI is achievable by simply moving the detector at an adequate distance away from the sample (*Fig. 3.3.2b*). In this configuration, the perturbed wave front W_2, generated by the object from the impinging unperturbed wave front W_1, has adequate space to evolve before being captured by the detector. Therefore, a typical intensity modulation profile corresponding to the edges is recorded in the phase-contrast image. These peaks generate a strong edge enhancement that can increase the possibility of structures identification, especially when the absorption contrast is not sufficient.

3.3.4 Computed Tomography

After discussing the origins of the contrast, it is time to introduce the Computed Tomography (CT). It is possible to describe the working principle of CT saying that a series of angular projection data are gathered together and processed by a tomographic reconstruction

algorithm, producing an image of the section of the sample, called "slice". The obtained image contains information about the internal 3D distribution of the refractive index. The process of data acquisition is the same as the one explained in the sections above, in fact the PPCI technique can be readily implemented also in CT mode provided that reconstruction algorithms are exploited to reproduce the 3D geometry of the sample. Phase retrieval is a fundamental process to reconstruct the image, therefore the general idea will be now presented. The action of the sample on the X-ray beam can be represented, as described before, by the transmission complex function

$$T(x,y) = A(x,y)e^{i\Phi(x,y)}. \tag{3.3.15}$$

The absorption contrast, proportional to $A^2(x,y)$ can be mixed with phase contrast, proportional to $\Phi(x,y)$. In this context, phase retrieval describes the procedure of carrying out the phase-shift map $\Phi(x,y)$ and more generally the transmission function $T(x,y)$ of an object from its phase-contrast images. While boundaries of the different details are greatly enhanced by PPCI, the bulk signal could be weak and the phase shift $\Phi(x,y)$ may vary slowly; phase retrieval can then highlight possible important differences in the phase-shift value between adjacent voxels. Normally, phase retrieval in PPCI requires at least two images, acquired either at two different distances from the sample or at two different beam energies. Supposing that only one image is acquired, there is no way to determine which contribution in the recorded intensity pattern is due to absorption and which one must be ascribed to the phase shift term. This uncertainty is due to the fact that image detectors are sensitive to intensity only, and that consequently both absorption and phase contrast are recorded as intensity fluctuations; once the intensity pattern is collected, there is no way to disentangle the contributions due to the different physical effects. However, the two contributions can be distinguished by exploiting distance or energy dependence. Several different approaches to phase retrieval are reported in the literature, but their description is out of the scope of this thesis.[123–126]

Once the set of projections is collected, the problem consists in reconstructing the 3D $n(x,y,z)$ distribution. The reconstruction of the CT images can be performed by means of several algorithms based on the analytical Fourier-based method or on iterative methods (algebraic and analytical). The following sections will address the physical and mathematical framework that enables such a reconstruction, followed by an assessment of typical artifacts that may occur and limiting the attention to the use of synchrotron radiation. All the basic steps of the image processing, such as flat fielding normalization and ring removal algorithm, will be described as they have been applied to obtain the reconstructions presented in this thesis work. The Radon transform and the Fourier slice theorem will be addressed, as each applies to CT image reconstruction. These techniques require reconstruction of a density function representing the internal structure of an object. This is typically accomplished by calculating a series of two-dimensional density functions (or slices) through the object on a set of planes and reconstructing the three-dimensional image from those images. Thus, the fundamental problem in both of these techniques is the calculation of the two-dimensional density function starting from angular-dependent projections. The following derivations use Fourier analysis to relate a filtered version of this measured signal to the density function of an object within the measured region.

3.3.5 Physical background

Radon Transform

Johann Radon in 1917 derived the fundamental mathematics for reconstructing a function, when its integral values are given.[127] Consider *Fig. 3.3.3*.

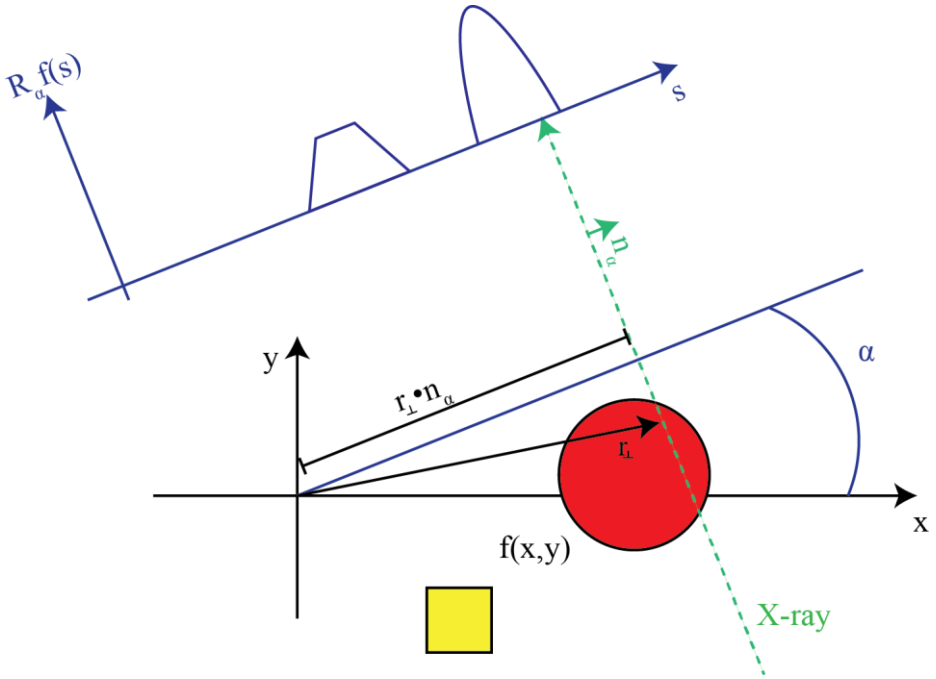

Fig. 3.3.3: illustration of the Radon transform. A 2D object (red circle and yellow square) described by a function $f(x,y)$ is hit by parallel X-rays (dashed green line). Interacting with the 2D object, the projected values form a one-dimensional profile $R_\alpha f(s)$ depending on the position s along a line, perpendicular to the projection direction.

A function $f(x,y)$, maps each point $r_\perp \in \mathbb{R}^2$ of an object to a value $g \in \mathbb{R}$. In the example represented in *Fig. 3.3.3*, the object is composed by two regions of constant density: a red sphere with $f(x,y) = 1$ and a yellow rectangle with $f(x,y) = 0.8$. The X-ray beam, represented by the green dashed arrow, is illuminating the object with a certain angle α, and is subject to attenuation and phase shift. Behind the object, a projection (represented by blue lines), given by all the values corresponding to each point (x,y) over the whole distance can be measured. The calculation, for a given angle, is made possible by means of the 2D Radon transform $R_\alpha f$ mathematically expressed as follows:[128]

$$R_\alpha f(s) = \int f(\boldsymbol{r}_\perp)\delta(\boldsymbol{r}_\perp \cdot \boldsymbol{n_\alpha} - s)d\boldsymbol{r}_\perp \tag{3.3.16}$$

where $\boldsymbol{r}_\perp = (x,y)^T$ is the vector defining the coordinate in the image plane that is orthogonal to the rotation axis, s is the coordinate of the projected profile, $\delta(\boldsymbol{r}_\perp \cdot \boldsymbol{n_\alpha} - s)$ is the one-dimensional Dirac delta distribution and $\boldsymbol{n}_\alpha = (cos\alpha, sin\alpha)^T$ is the normal vector perpendicular to the projection direction. Doing the scalar product $\boldsymbol{r}_\perp \cdot \boldsymbol{n_\alpha}$, the coordinate vector $\boldsymbol{r}_\perp$ is projected to a line forming the angle α with the x-axis (black line in *Fig. 3.3.3*). Accordingly, the argument of the delta distribution goes to zero for points forming a straight

line of angle α with the y-axis with distance s to the origin. In such a way, only points on the X-rays propagation direction contribute to the $R_\alpha f(s)$. Thus, the 2D integral in *Eq. 3.3.16* becomes a one-dimensional integral over the line.

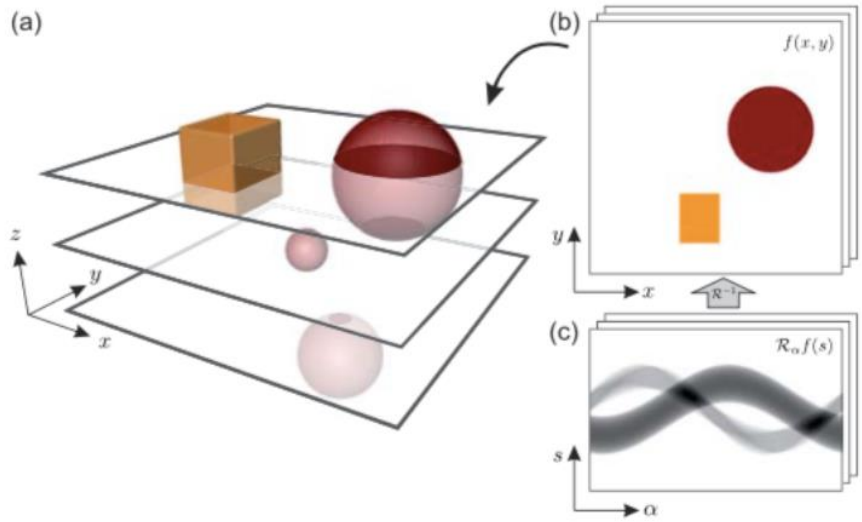

Fig. 3.3.4: illustration of the process from 2D slices to a 3D object. In a tomographic scan, each vertical position in a projection image (z = const) forms (c) a sinogram, i.e., the projected profiles$R_\alpha f(s)$ plotted as a function of the projection angle α. These can be inverted to obtain (b) a reconstructed slice f(x, y) that can be virtually stacked to yield (a) a 3D volume. Figure adapted from the literature.[129]

The generalization of the Radon transform to higher dimensions n is easily possible.[129] For example, in the case of a 3D Radon transform as a function R: $\mathbb{R}^3 \rightarrow \mathbb{R}$, in which the integral is calculated over a plane instead of a line and the vector $\boldsymbol{n}$ will be the normal vector to this plane. Taking into account the Radon transform of a single point at a certain distance from the origin, the 2D plot of all the projected values as a function of the angle, is a sine curve; such a data representation is the so-called sinogram (see *Fig. 3.3.4c*). For a given sinogram $p(\alpha, s) = p_\alpha(s) = R_\alpha f(s)$, therefore it is possible to obtain a 2D image $f(x, y)$ from a set of 1D projections taken at different angles by mean of an inverse Radon transform formally written as:

$$f(x,y) = R^{-1}\big(p(\alpha,s)\big) \tag{3.3.17}$$

thus, to obtain a 3D reconstruction of an entire object, all the collected $f(x, y)$ can be stacked in the z direction, yielding a 3D dataset $f(x, y, z)$ as shown in *Fig. 3.3.4*.

Fourier slice Theorem

The reconstruction of the 2D image of the object from the Radon transform can be accomplished exploiting the Fourier analysis technique, using the Fourier slice theorem. To explain the Fourier slice theorem, consider a general 2D function $f(x, y)$ which is projected (or more precisely integrated) along the y axis to generate the new function of x only, defined by

$$p(x) = \int f(x,y)dy. \tag{3.3.18}$$

The Fourier transform of $p(x)$ is

$$P(q_x) = \int p(x)e^{iq_x x}dx \tag{3.3.19}$$

while the Fourier transform of the function $f(x,y)$ is

$$F(q_x, q_y) = \iint f(x,y)e^{i(q_x x + q_y y)}dxdy. \tag{3.3.20}$$

Setting $q_y = 0$ defines a slice through $F(q_x, q_y)$ given by

$$F(q_x, q_y = 0) = \int \left[\int f(x,y)\,dy\right] e^{iq_x x}dx \tag{3.3.21}$$

which can be written as

$$F(q_x, q_y = 0) = \int p(x)e^{iq_x x}dx = P(q_x) \tag{3.3.22}$$

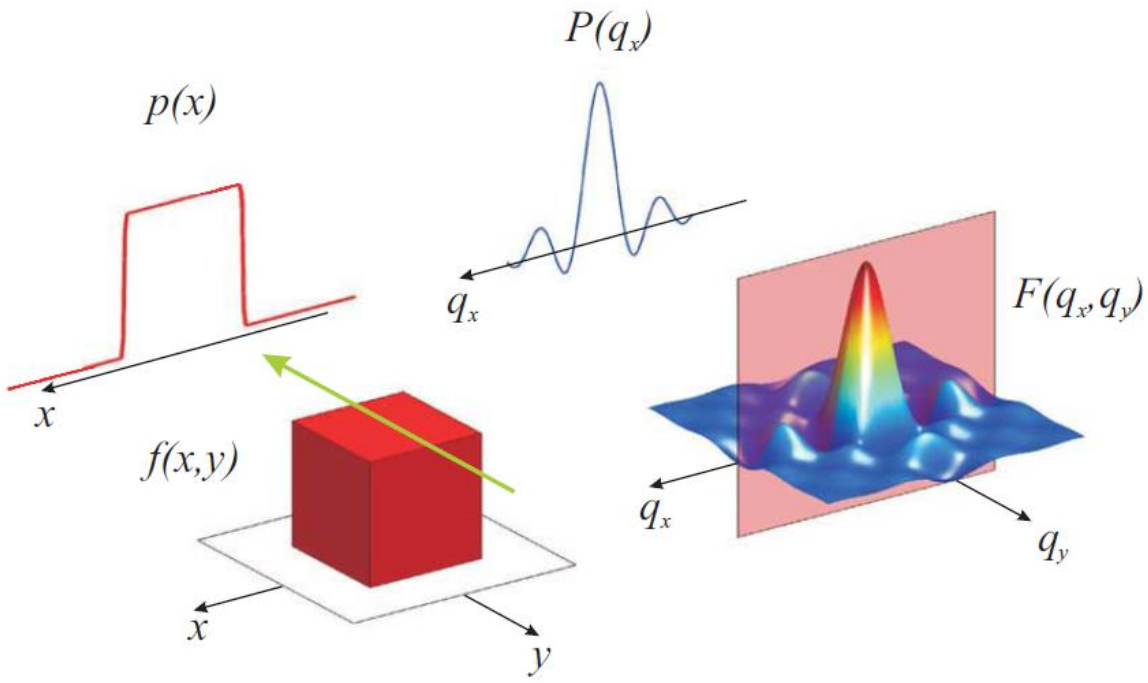

Fig. 3.3.5: illustration of the Fourier slice theorem. $f(x,y)$ is a 2D top-hat function with Fourier transform $F(q_x, q_y) = \left[\frac{sin\,(q_x x)}{q_x x}\right]\left[\frac{sin\,(q_y y)}{q_y y}\right]$. When $f(x,y)$ is integrated (projected) along the y axis, it generates a 1D top hat function $p(x)$ depending on x only. The Fourier transform of $p(x)$ is $P(q_x) = \left[\frac{sin\,(q_x x)}{q_x x}\right]$ which is the $q_y = 0$ slice through $F(q_x, q_y)$. Figure adapted from the literature.[130]

In other words, the Fourier transform of the projection along a particular line of a two-dimensional function $f(x,y)$ is equal to a slice through the Fourier transform of $f(x,y)$ along a line passing through the origin orthogonal to the propagation direction. In a more general way it is possible to consider the n-dimensional Radon transform $Rf(s)$ of a certain function $f(\boldsymbol{r})$ with $\boldsymbol{r} \in \mathbb{R}^n$ and $\boldsymbol{n}$ the normal vector of a hyperplane in $\mathbb{R}^{n-1}$ the Fourier transform with respect to s is:

$$F_s(Rf)(\rho) = \int \left(e^{-2\pi i \rho s} \int f(r)\delta(\boldsymbol{r}\cdot\boldsymbol{n} - s)d\boldsymbol{r} \right) ds = \int e^{-2\pi i \rho \boldsymbol{r}\cdot\boldsymbol{n}} f(\boldsymbol{r})d\boldsymbol{r} \qquad (3.3.23)$$

where the order of integration has been changed and then integral over s have been carried out, eliminating the delta distribution. It is possible to define a coordinate vector in the reciprocal space as $\boldsymbol{v} = \rho\boldsymbol{n}$, so that the equation above is proportional to the Fourier transform of the function $f(\boldsymbol{r})$ with respect to the coordinate vector $\boldsymbol{r}$:

$$F_s(Rf)(\rho) = \int e^{-2\pi i \boldsymbol{v}\cdot\boldsymbol{r}} f(\boldsymbol{r})d\boldsymbol{r} = F_r(f)(\boldsymbol{v}) \qquad (3.3.24)$$

This last equation is the Fourier slice theorem, which states that the one-dimensional Fourier transform of a projection $Rf(s)$ gives the values of the n-dimensional Fourier transform of the starting function $f(r)$ along a slice through the origin defined by $\boldsymbol{v} = \rho\boldsymbol{n}$ and it proves that the Radon transform is invertible for every dimension n. For the 2D case (*Fig. 3.3.5*), the theorem will be:

$$\tilde{p}_\alpha(\rho) = F_s(R_\alpha f)(\rho) = \tilde{f}(\rho cos\alpha, r sin\alpha) \qquad (3.3.25)$$

using the typical notation for the Fourier transformed quantities. Accordingly, the Fourier space can be totally determined and the initial function $f(\boldsymbol{r})$ can be reconstructed by mean of the inverse Fourier transform. This method is known as direct Fourier method. Looking at *Fig. 3.3.6* it is noticeable that, when the function $f(\boldsymbol{r})$ is given at discrete points and a limited number of angles is used, the Fourier space won't be uniformly sampled, as the points won't be on a rectangular grid. For this reason, it is necessary to apply interpolation methods with the consequent introduction of artefacts in the real space. In addition, there is a difference in point density between low (near the origin) and large spatial frequencies, therefore the high resolution information in the image can be under-represented.

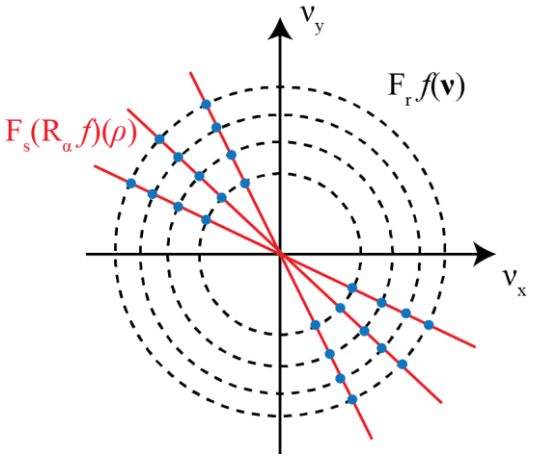

*Fig. 3.3.6: illustration of the Fourier slice theorem for 2D case (n = 2): the 1D Fourier transform of a projection $R_\alpha f(s)$ corresponds to a slice (red line) in the 2D Fourier transform of the object f(**ν**). If the function is given at discrete points, indicated by the black dots, a sophisticated interpolation in Fourier space is necessary to avoid artefacts in the tomographic reconstruction. Both Radon transform and Fourier slice theorem are fundamental in the acquisition of the CT data.*

3.3.6 Artifacts

The inhomogeneity of the CT images is one of the most common problems occurring during a CT acquisition. This non-uniform aspect, appearing both in the sample region and outside of it, can have different origins, such as:

- the temporal instability of the incident X-ray beam due to a possible decrement of the machine current, to the vibrations or thermal drift of the optics;
- the spatial inhomogeneity of the incident beam due to some intrinsic inhomogeneity of the X-ray beam from the wiggler, defects or deformations of optical elements or similar;
- the detector response inhomogeneity caused by electron noise or pixels having different gain.

Darknoise (taken without the beam illuminating the detector) and whitefield (taken with beam impinging on the detector without the sample in the X-ray path) images are used for a proper renormalization of the signal, usually implemented before the slices' reconstruction, according to:

$$normalized\ image = \frac{raw\ image - darknoise}{whitefield - darknoise} \tag{3.3.26}$$

performed on a pixel by pixel basis. One recurrent artifact that typically degrades image quality in X-ray tomography is the presence of concentric rings arising from the non-uniform pixel response of detector elements. The ring shape results from the back projection of stripe artifacts in the sinogram image, which is interpreted as a feature that always has the same distance to the center of rotation and is present at each angle. A usual way to overcome this problem is to de-stripe the sinogram or, in order to avoid the need of additional imaging data (i.e. whitefield and sinogram images), it is possible to apply a de-striping filter directly to the transformed reconstructed slices. To do that, images are transformed into polar coordinates (where the value of the x axis represents the distance from the rotation axis and the values on the y axis the rotation angle) so that the rings show up as lines parallel to the y direction. Many approaches

are exploited to remove these defects. The Rivers method at first computes the average row of the sinogram. The average should have the minimal high-frequency content, as real objects usually move in the sinogram, and are blurred out when computing the average row. In fact, anomalous detector elements will show up as spikes or dips. The magnitude of these detector anomalies is then computed by subtracting a smoothed version of the average row from the unmodified average. Subtracting the result of this last step from each row in the sinogram, will result in a final sinogram with much less vertical striping.[131]

3.3.7 Experimental setup

The measurements presented in this thesis were performed at the SYRMEP beamline, of Elettra Sincrotrone Trieste (schematic representation of the beamline reported in *Fig. 3.3.7*). SYRMEP is the acronym of Synchrotron Radiation for Medical Physics. Thanks to the horizontal acceptance covered by the front-end port (7 mrad), the beamline is able to provide, at a distance of about 30 m from the source, a mono-energetic laminar beam with an area of (210 x 5) mm^2 at 20 keV. Depending on the experimental setup, a white beam can be exploited, or a monochromator may be placed between the source and the sample.[132]

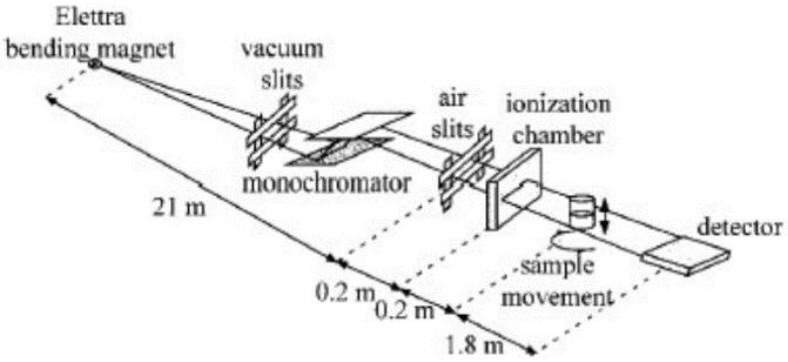

Fig. 3.3.7: schematic sketch of the SYRMEP beamline.

In the applications of this thesis only the "white-beam" configuration mode was exploited, illuminating the sample with the polychromatic X-ray beam filtered for low energy components by means of 0.5 mm of silicon, resulting in a mean energy of about 17 keV. The imaging system consists in a Hamamatsu sCMOS detector, optically coupled to a 17 µm thick GGG (Gd3Ga5O12:Eu) scintillator utilizing a set of optical lenses, which enable different magnifications. The sCMOS sensor comprises 2048 × 2048 pixels (with a size of 6.5 µm × 6.5 µm) and features a dynamic range of 37,000:1. The optical magnification has been set to 3.25, which translates into a field of view of 4.1 mm × 4.1 mm and a pixel size of 2 µm × 2 µm. Given the sample diameter (less than 1 mm) and the parallel beam geometry, tomographic images were reconstructed from 1800 evenly spaced projections spanning over 180 degrees and collected in continuous rotation mode. The exposure time was set to 200 ms/projection, thus resulting in an overall sample exposure of approximately 6 minutes. The exposure time was chosen to have a good SNR and, at the same time, not damaging the plastic samples. The source-to-sample distance was approximately 22 m, while the sample-to-detector distance was 100 mm. Thanks to this geometry and to the spatial coherence of the source, projection images were

obtained in the propagation-based phase-contrast regime, in the near field diffraction regime. Prior to reconstruction (performed with a GPU-based filtered back-projection algorithm), projections were further processed by means of a TIE-Hom based phase-retrieval filter, obtaining a higher signal-to-noise ratio at cost of a loss of edge-enhancement signal.[133] After processing, the final CT reconstruction yields a 3D map, in which the contrast is substantially proportional to the linear attenuation coefficient of the sample.[134,135]

3.4 Bidirectional Reflectance Distribution Function

The Bidirectional Reflectance Distribution Function (BRDF) characterizes the visual appearance of materials i.e. it describes objects' colors and aesthetic appearance characteristics. The BRDF is extremely important to digitally reproduce an object through computer graphics, perform rendering but also describe materials properties that must be implemented in ray-tracing software. Evidently, the BRDF holds a great amount of information; therefore having the ability to both quantize and decode a BRDF is an important tool for the research.[136,137] Formally, BRDF quantifies the geometric radiance distribution which results from visible incident light. The term bidirectional is used as it is a function of both incident and reflected light directions. It is also a distribution function in the classical sense, as integration over the hemisphere above the sample surface results in the reflectance, ρ, which ranges from 0 to 1. Before introducing the BRDF, it is important to focus the attention on the radiance, irradiance and reflectance quantities.

3.4.1 Radiance, irradiance, reflectance

At first, consider an ideal point-like light source, i.e. an infinitesimal point radiating electromagnetic energy uniformly in all space directions. The radiant flux, or radiant power, Φ measured in Watts [W], is the radiant energy per unit time emitted, transmitted or received by an object. Light sources are not ideal point sources, in the sense that they have a finite size and they radiate differently in the space; thus, it is important to describe the amount of energy being emitted in a certain direction. In this case, we define the radiant intensity (I) as the amount of radiant flux Φ per solid angle $d\omega$ (graphic representation reported in *Fig. 3.4.1*):

$$I = \frac{d\Phi}{d\omega} \left[\frac{W}{sr}\right] \tag{3.4.1}$$

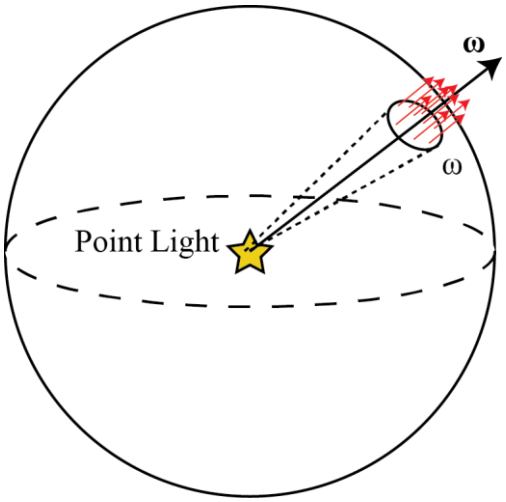

Fig. 3.4.1: ideal point-like light source radiating uniformly in the space. The radiant intensity is the radiant flux traveling through and infinitesimal solid angle ω in the $\boldsymbol{\omega}$ *direction.*

Rather than following the light along a specific ray (and so, a specific direction), it is of particular interest to observe a finite surface illuminated by a point light source. The term

characterizing the incident flux on the surface is called irradiance, E, and is defined as flux per unit sample area:

$$E = \frac{d\Phi}{dA} \quad \left[\frac{W}{m^2}\right]. \tag{3.4.2}$$

Consider now an isotropic point light source. The flux emitted by the source at a distance r will be distributed across a spherical surface with the same radius. Moreover, irradiance is in general proportional to the cosine between surface normal and the direction of the light. This fact is highlighted in the *Fig. 3.4.2*, where two equal surfaces receive different amount of radiation due to their diverse orientation and distance from the source.

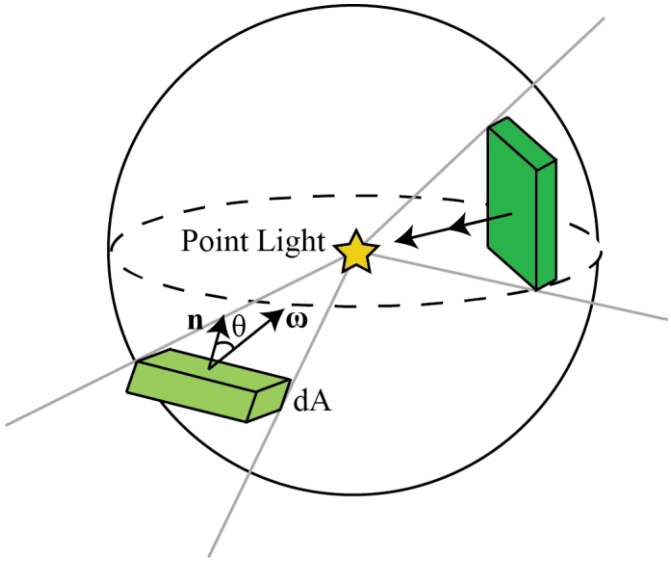

Fig. 3.4.2: examples of orientation and distance from the source dependence on the irradiance.

In the same way, it is possible to consider the light leaving a surface. This phenomenon is described by the physical quantity denominated radiance, usually labelled with L. Radiance is defined as the flux received per solid angle, per projected surface area:

$$L = \frac{d^2\Phi}{dA_{proj}d\omega} = \frac{d^2\Phi}{dAcos\theta d\omega} \quad \left[\frac{W}{m^2 \cdot sr}\right] \tag{3.4.3}$$

where θ is the angle between the surface normal and the direction to the observer/detector. Clearly, the received power should be normalized to the solid angle covered by the aperture of the camera (ω) and to the projected surface area.

It is worthwhile at this point to explore the definition of reflectance. The reflectance is defined as the ratio of the power leaving a surface to the power incident upon a surface.

$$\rho = \frac{\Phi_r}{\Phi_i} \tag{3.4.4}$$

Reflectance is therefore unitless and, if the material taken in consideration is passive (i.e. no energy being added by the material itself), has a range limited to $0 \leq \rho \leq 1$. There exist multiple variants of reflectance, like the hemispherical reflectance, describing the total amount of power leaving a surface relative to the total amount entering, or the spectral reflectance focusing on the ratio at a specified wavelength (Nicodemus notes nine different reflectance

specifications),[138] but the quantity of interest is the Total Integrated Scatter (*TIS*) also known as Directional Hemispherical Reflectance (*DHR*). The *TIS*, $\rho_{TIS}(\theta_i, \Phi_i, \lambda)$, is the ratio of the total energy reflected from a surface into the subtending hemisphere to that incident on the surface from a precise given direction.

$$\rho_{TIS}(\theta_i, \varphi_i, \lambda) = \frac{\Phi_r}{\Phi_i} \tag{3.4.5}$$

The difference with respect to the general reflectance is that now the direction of the incident light is considered. As with other radiometric terms, the *TIS* is also a function of wavelength. The diffuse reflectance, $\rho_d(\theta_i, \varphi_i, \lambda)$, can be described as the *TIS* minus the specular reflectance coming from a small solid angle encompassing the specular lobe. The ratio of diffuse reflectance to *TIS* may then be used as a measure of a diffuseness, *d*, of a material

$$d = \frac{\rho_d(\theta_i, \varphi_i, \lambda)}{\rho_{TIS}(\theta_i, \varphi_i, \lambda)} \tag{3.4.6}$$

A perfect diffusive (also known as Lambertian) surface would therefore have d = 1, while a perfect mirror would have d = 0.

3.4.2 Bidirectional Reflectance Distribution Function definition and properties

Using the quantities introduced above, the precise definition of the BRDF will be given, which was originally coined by Nicodemus.[139] Theoretically, the BRDF describes the spectral reflectance properties of a material as a function of both illumination or viewing angles. BRDF is defined in radiometric terms as the surface radiance of a sample due to the scattering of the incident radiation from a defined direction of irradiation:

$$f_r(\omega_i, \omega_r, \lambda) = f_r(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) = \frac{dL_r(\boldsymbol{\omega}_r, \lambda)}{dE_i(\boldsymbol{\omega}_i, \lambda)} = \frac{dL_r(\boldsymbol{\omega}_r, \lambda)}{L_i(\boldsymbol{\omega}_i, \lambda) cos\theta_i d\omega_i} \quad \left[\frac{1}{sr}\right] \tag{3.4.7}$$

and describes the ratio between the radiance reflected off a surface in a specific direction and the irradiance hitting the surface from another specific direction. The BRDF depends on the incidence and reflection directions $(\theta_i, \varphi_i, \theta_r, \varphi_r)$ in the polar and azimuthal coordinates formalism, and on the wavelength (λ) of the incident light. Hence, the BRDF is a 5-dimensional quantity.[140] BRDF is theoretically specified for a point source and detector, as well as an infinitesimal surface area, but practical measurement considerations results in some averaging over the source and detector solid angles (ω_i, ω_r) and surface area *A*. Going back to some concepts introduced before, it is possible to review them in terms of the BRDF. *TIS* is therefore given by

$$\rho_{TIS}(\theta_i, \varphi_i, \lambda) = \int f_r(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) d\omega_r \tag{3.4.8}$$

being

$$d\omega_r = cos\theta_r sin\theta_r d\theta_r d\varphi_r \tag{3.4.9}$$

while incident flux, and reflected flux are defined as

$$\Phi_i = \int E_i(\theta_i, \varphi_i, \lambda) dA_i \tag{3.4.10}$$

and

$$\Phi_r = \iint f_r(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) E_i(\theta_i, \varphi_i, \lambda) dA_i d\omega_r. \tag{3.4.11}$$

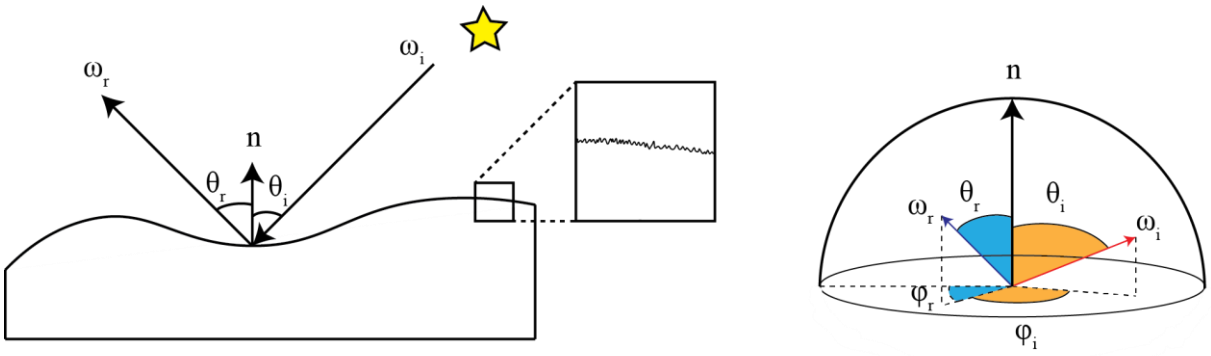

Fig. 3.4.3: the BRDF (f_r) relates the amount of radiance leaving a surface in a specific direction, ω_r, relative to the amount of irradiance hitting the surface from a direction ω_i.

The observed reflected radiance L_0 at a direction ω_r can be calculated by inserting the surface BRDF into the so-called rendering equation:[141]

$$L_0(\omega_r, \lambda) = L_e(\omega_r, \lambda) + L_r(\omega_r, \lambda) = L_e\omega_r + \int f_r(\omega_i, \omega_r, \lambda) L_i(\omega_i, \lambda) d\omega_i \tag{3.4.12}$$

where the integration is performed over the whole solid angle. L_e is the radiance contributed by the surface, and for passive materials the term can be neglected. In this case, the above equation is referred to as the "reflected radiance equation", and $L_0(\omega_r, \lambda) = L_r(\omega_r, \lambda)$. To keep a consistent notation with r-subscripts the term L_r will be used in spite of L_0 in the whole thesis. In contrast to the reflectance, ρ, the BRDF may locally take values greater than 1 sr^{-1}. Instead, energy conservation states that the integrated BRDF (over all possible directions of observation) should never exceed 1 sr^{-1}:

$$\forall\, \omega_i : 0 \leq \int f_r(\omega_i, \omega_r, \lambda) d\omega_r \leq 1. \tag{3.4.13}$$

The BRDF of a perfect mirror is a Dirac delta function, having maximum intensity when the observation direction coincides with the reflected direction of the light, and zero everywhere else. On the other hand, a perfect diffuse surface with no absorption will have, instead, a constant BRDF of $f_d = \frac{1}{\pi}$ (since the integral of the cosine over the hemisphere equals π). If the reflector is Lambertian, which means the surface is ideal matte and the brightness is the same regardless of the observer angle, therefore

$$f(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) = f_d(\theta_i, \varphi_i, \lambda) \tag{3.4.14}$$

which is constant over the reflected hemisphere, and that the irradiance is uniform. The resulting derivation is

$$\rho(\theta_i, \varphi_i, \lambda) = \frac{\iint f(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) E_i(\theta_i, \varphi_i, \lambda) dA_i d\omega_r}{\int E_i(\theta_i, \varphi_i, \lambda) dA_i} \tag{3.4.15a}$$

$$= \int_0^{2\pi} \int_0^{\pi/2} f(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) cos\theta_r sin\theta_r d\theta_r d\varphi_r \tag{3.4.15b}$$

$$= 2\pi f_d(\theta_i, \varphi_i, \lambda) \int_0^{\pi/2} cos\theta_r sin\theta_r d\theta_r = \pi f_d(\theta_i, \varphi_i, \lambda) \tag{3.4.15c}$$

hence

$$f_d(\theta_i, \varphi_i, \lambda) = \frac{\rho(\theta_i, \varphi_i, \lambda)}{\pi} \tag{3.4.16}$$

and remembering that the reflector is ideal ($\rho(\theta_i, \varphi_i, \lambda) = 1$) the constant BRDF value of $\frac{1}{\pi}$ is obtained. Unfortunately, none of the two models presented before (perfect specular and perfectly diffusing) is valid to describe a real surface, due to the presence of surface roughness and inelastic scattering effects. The simplest way to handle this is to create a model where a perfectly diffusive, or Lambertian, component is simply added to a specular component. This is about as far as the discussion can be taken without looking at a specific model. A three dimensional depiction of the diffuse component, specular component, and the diffuse-specular combination BRDF models are shown in *Fig. 3.4.4*.

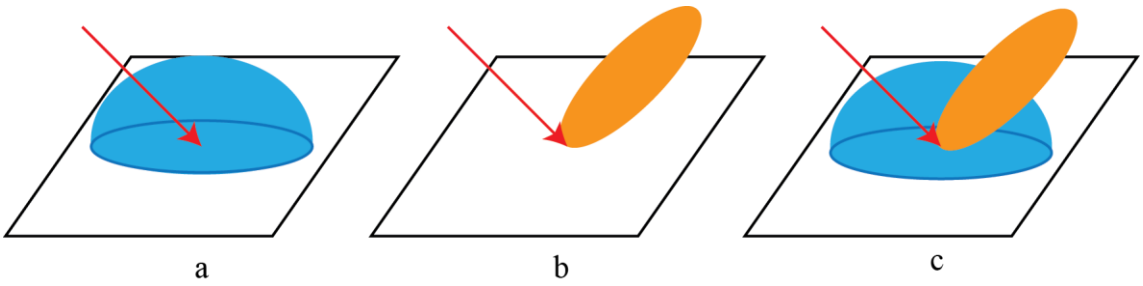

Fig. 3.4.4: (a) Lambertian reflection model, (b) specular reflection model, (c) sum of the Lambertian and specular models.

Another important property of the BRDF is that it obeys Helmholtz reciprocity:[142]

$$f_r(\omega_i, \omega_r) = f_r(\omega_r, \omega_i) \tag{3.4.17}$$

stating that a ray of light's path is reversible. In computer graphics, this property is utilized for example in ray-tracing,[143] and back-tracing.[144] These two techniques have been developed to consider the kinematics of electromagnetic waves along with their associated propagation and reflection properties to perform optical simulations. The forward ray-tracing technique consists of projecting rays from a source and collecting the rays at an observer. Realistic scenarios directly follow this approach. The backward ray-tracing technique (also known as back-tracing) consists of projecting rays from an observer and collecting them at a source. The forward technique is valid for small sources when compared to the observer and the backward technique

is valid for the opposite.[145] The Helmholtz reciprocity is also very convenient in a measurement scheme since

$$f_r(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) = f_r(\theta_i, \varphi_i + \pi, \theta_r, \varphi_r + \pi, \lambda) \quad (3.4.18)$$

implying that only half of the hemisphere needs to be covered: $\theta \in \left[0, \frac{\pi}{2}\right]$ and $\varphi \in [0, \pi]$. For a large group of materials, rotating the surface around its normal will not affect the amount of reflected light in any given light/view configuration. Materials holding this property are called isotropic and materials not holding it are called anisotropic. Isotropy implies that the BRDF is invariant with respect to rotations about the surface normal, i.e.

$$f_r(\theta_i, \varphi_i, \theta_r, \varphi_r, \lambda) = f_r(\theta_i, \varphi_i + \alpha, \theta_r, \varphi_r + \alpha, \lambda) \; \forall \, \alpha \in \mathbb{R} \quad (3.4.19)$$

3.4.3 Experimental setup

The BRDF measurements were performed at the Light&Lighting laboratory of the KU Leuven University. A home-built full three-dimensional BRDF instrument (see *Fig. 3.4.5*) was used. The light source represented in the illumination section (*Fig. 3.4.5a*) consists of a xenon arc lamp mounted in a lamp housing, in front of which a diaphragm is positioned; an image of the diaphragm aperture is formed at the detector plane by use of a collimating mirror. The detection section (*Fig. 3.4.5b*) includes a collector lens, imaging the sample area onto an aperture of an integrating cavity, which in turn is coupled to a spectrometer with interchangeable gratings, by use of a high-grade fused silica fiber bundle. The detector head is mounted on a bench, which can be rotated by aid of two motorized rotation stages, enabling alignment at any viewing angle with respect to the sample. The sample holder allows manual positioning and alignment of the specimen by adjusting two rotation stages and one translation stage. For a more detailed description of the instrument, see Leloup et al.[146]

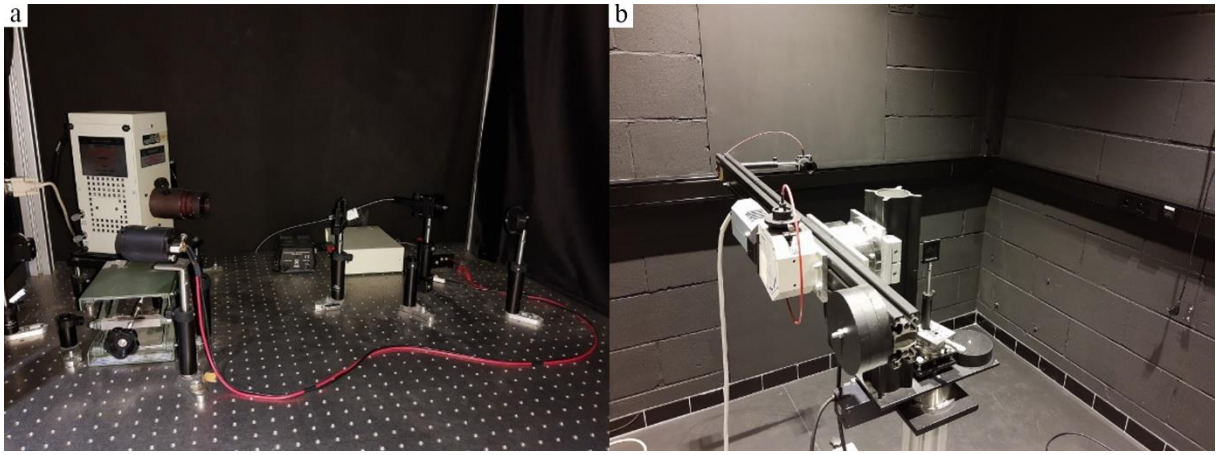

Fig. 3.4.5: Light&Lighting laboratory BRDF measurement setup showing (a) the illumination and (b) the detection sections.

BRDF measurements were performed at an incident angle of 45°. In particular, to better emphasize the wavelength dependence the 45°,0°: 45°,180° geometry was used, while to stress the angular dependence the 45°,0°: x,180° geometry, with x in the interval [20°,70°], was exploited.[147] The integration time was optimized for each viewing angle. All measurements

were dark current corrected. The angular distribution of the incident beam is reported in *Fig. 3.4.6* and it affects the angular dependence of the BRDF. This limits the angular resolution of the BRDF spectra.[148]

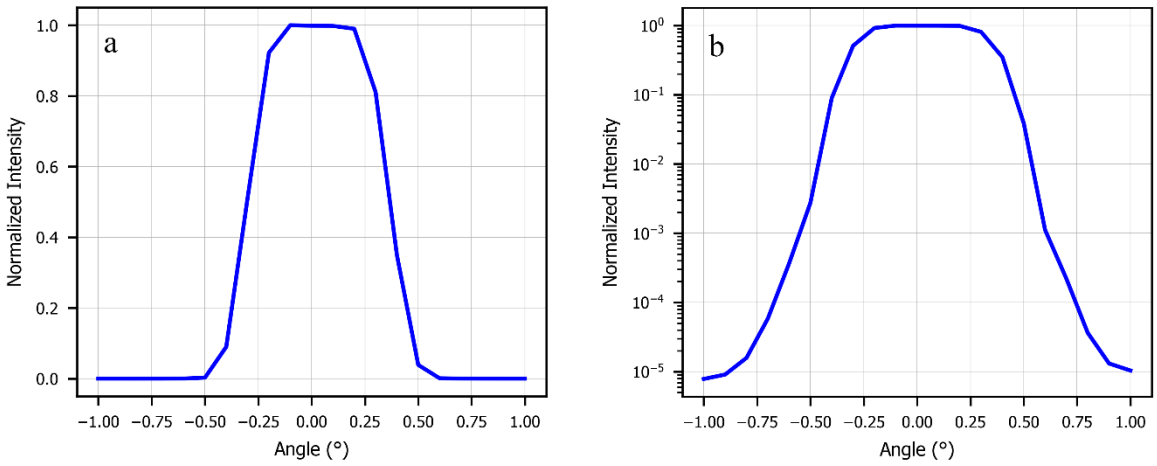

Fig. 3.4.6: angular distribution of the incident beam represented using linear scale (a) and semi-log scale (b).

4. Computer-based inspection and characterization approaches

4.1 Introduction to machine vision

In this chapter, a general overview regarding some basic concepts of computer vision algorithms will be presented. In particular, the fundamental notions useful to understand the experimental section involving the filtering and segmentation processes exploited to tune the machine for the automatic detection of defects (*Chapter 5*) will be introduced. It is firstly convenient to give a definition of a digital greyscale image. An greyscale image may be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial (planar) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y, and the intensity values of f are all finite, discrete quantities, the image is called digital image. Pixel is the term used most widely to denote the elements of a digital image.

4.1.1 Intensity transformation and spatial filtering

The processing techniques exploited for the realization of an automatic visual inspection system have been implemented to operate in the spatial domain, which is the plane containing the pixels. These techniques work directly on the pixels of an image as opposed to the frequency domain, in which operations are performed on the Fourier transform of the image itself. Usually, spatial domain techniques are computationally more efficient and require fewer processing resources. They can be denoted by

$$g(x, y) = T[f(x, y)] \tag{4.1.1}$$

where $f(x, y)$ is the input image, $g(x, y)$ is the output image, and T is an operator applied on f. For any location (x, y), the output image value g at those coordinates is obtained applying T to the neighborhood with origin at (x, y) in f. Typically, the process starts at the top left of the input image and proceeds pixel by pixel in a horizontal scan, one row at a time. When the origin of the neighborhood is at the border or in the corner of the image, part of the neighborhood will reside outside the image. The procedure is either to ignore the outside neighbors in the computations specified by T or to pad the image with a border of 0s or some other specified intensity values. The procedure just defined is denominated spatial filtering, while the neighborhood is called a spatial filter (also referred to as a spatial mask, kernel or template). The smallest possible neighborhood is of size 1×1. In this case, g depends only on the value of f at a unique point (x, y) and T becomes an intensity (also called gray-level) transformation function of the form

$$s = T(r) \tag{4.1.2}$$

where, for simplicity in notation, s and r are variables denoting the intensity of g and f at any point (x, y), respectively. Some processing approaches can be formulated with intensity transformation functions. These transformations can be used for image enhancement and image

segmentation, as it will be presented in detail in the next sections. Approaches whose results depend only on the intensity at a point sometimes are called point processing techniques, as opposed to the neighborhood processing techniques discussed earlier in this section. Intensity transformations are among the simplest of all image processing techniques. As denoted previously, these values are related by an expression of the form $s = T(r)$. One example of such intensity transformations is represented by the thresholding.[149]

Thresholding

Suppose that the intensity histogram of an image $f(x, y)$, composed of light objects on a dark background, is similar to the one represented in *Fig. 4.1.1a* with object and background pixels that have intensity values peaked into two dominant regions.

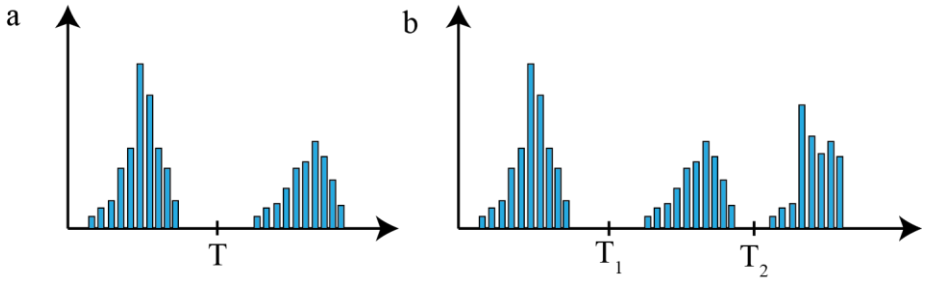

Fig. 4.1.1: representation of two typologies of intensity histograms: in panel (a) two intensity peaks are evidenced, leading to a single threshold, while in panel (b) two peaks are visible and the thresholding process requires a dual threshold.

One possibility to extract the objects from the background is to fix a threshold T that separates the two main modes. Then, any point (x, y) in the image at which $f(x, y) > T$ is called an object point; otherwise, the point is called a background point. The segmented image $g(x, y)$, is given by

$$g(x,y) = \begin{cases} 1 & if\ f(x,y) > T \\ 0 & if\ f(x,y) \leq T. \end{cases} \tag{4.1.3}$$

If the constant T is the same over the entire image, the process described above is referred to as "global thresholding", while if the value of T changes over an image, the procedure is called "variable thresholding". *Fig. 4.1.1b* shows a more complex thresholding problem involving a histogram with three dominant modes. Here, multiple thresholding classifies a point (x, y) as belonging to the background if $f(x, y) \leq T_1$, to one object if $T_1 < f(x, y) \leq T_2$, and to the other object if $f(x, y) > T_2$. That is, the segmented image is given by

$$g(x,y) = \begin{cases} a & \text{if } f(x,y) > T_2 \\ b & \text{if } T_1 < f(x,y) \leq T_2 \\ c & \text{if } f(x,y) \leq T_1 \end{cases} \tag{4.1.4}$$

where a, b and c are three distinct constant values. To obtain a good thresholding, some aspects are important: (1) the separation between peaks; (2) the noise content in the image; (3) the relative sizes of objects and background; (4) the uniformity of the illumination source; and (5) the uniformity of the reflectance properties of the image. Illumination inhomogeneity may bring to the erroneous classification of pixels, while the relative size of objects with respect to

background is important to perform a distinction between the objects. If one of the features above described is not well tuned, sometimes it is not possible to perform a good thresholding: in order to exploit this process, the experimental setup must be adequately prepared.[150]

Spatial correlation and convolution

Correlation and convolution belong to the class of linear spatial filtering features, modifying an image f by replacing the value at each pixel with some linear function of the values of nearby pixels. Correlation is the process of moving a filter mask over the image and computing the sum of products at each location.[151] The convolution is the same procedure, except for the fact that the filter is first rotated by 180°. The differences will be now highlighted using a 1D example.

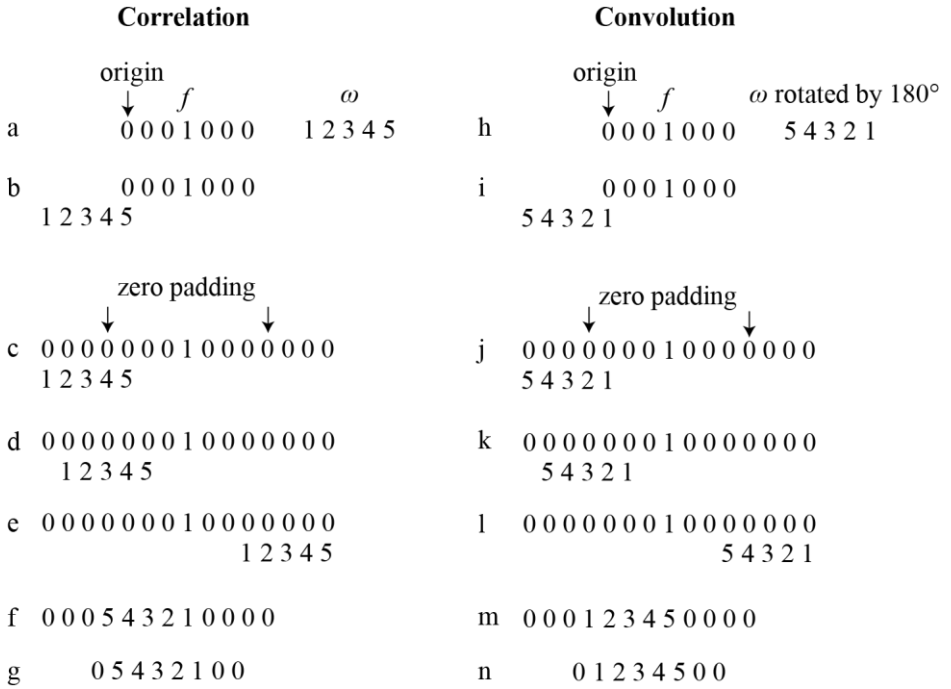

Fig. 4.1.2: 1D correlation (left column, panels a-g) and convolution (right column, panels h-n) of a filter ω with a discrete unit impulse. Taking into account the correlation the processes are: introduction of function and filter (panel a), alignment of the two elements in the starting position (b), zero padding (c), overall position after one filter shift (d), final position of the filter (e), correlation result (f) and cropped correlation result (g). The same steps are applied for the convolution (panels h-n).

Fig. 4.1.2a shows a 1D function, f, and a filter, ω, while *Fig. 4.1.2b* shows the starting position to perform correlation. If there are parts of the functions (f and ω) that do not overlap, it is possible to pad f with enough 0s on either side to allow each pixel in ω to visit every pixel in f (*Fig. 4.1.2c*). If the filter is of size m, $m-1$ 0s on either side of f need to be added. The first value of correlation is the sum of products of f and ω for the initial position shown in *Fig. 4.1.2c* (the sum of products is 0 in the example). This corresponds to no displacement of the filter: indicating with x the filter displacement, this case is typified by $x = 0$. To obtain the second value of correlation, ω is shifted one pixel location to the right (displacement of $x = 1$, panel *d*) and the sum of products is computed. Proceeding in this manner, the full correlation

result is obtained in *Fig. 4.1.2f*. Note that it took 12 values of x (i.e., 0,1,2,…,11) to fully slide ω past f so that each pixel in ω visited every pixel in f. The correlation is, therefore, a function of displacement of the filter, which means that the first value corresponds to no displacement of the filter, the second to one unit displacement and so on. The correlation of a filter, with a function containing all the values equal to 0 and only one value equals to 1 (a discrete unit impulse), will give a copy of the filter rotated by 180°.[152] On the contrary, as learned above, the convolution applied to a discrete unit impulse does not rotate the filter, highlighting the complementarity of these two filtering features (panels *h-n*). The procedure illustrated can be easily extended to images, as shown in *Fig. 4.1.3*. For a filter of size $m \times n$ is required padding the image with a minimum of $m - 1$ rows of 0s at the top and bottom and $n - 1$ columns of 0s on the left and right. In this case, m and n are equal to 3, then it is important to pad f with 2 rows of 0s above and below and 2 columns of 0s to the right and left, as *Fig. 4.1.3b* shows.

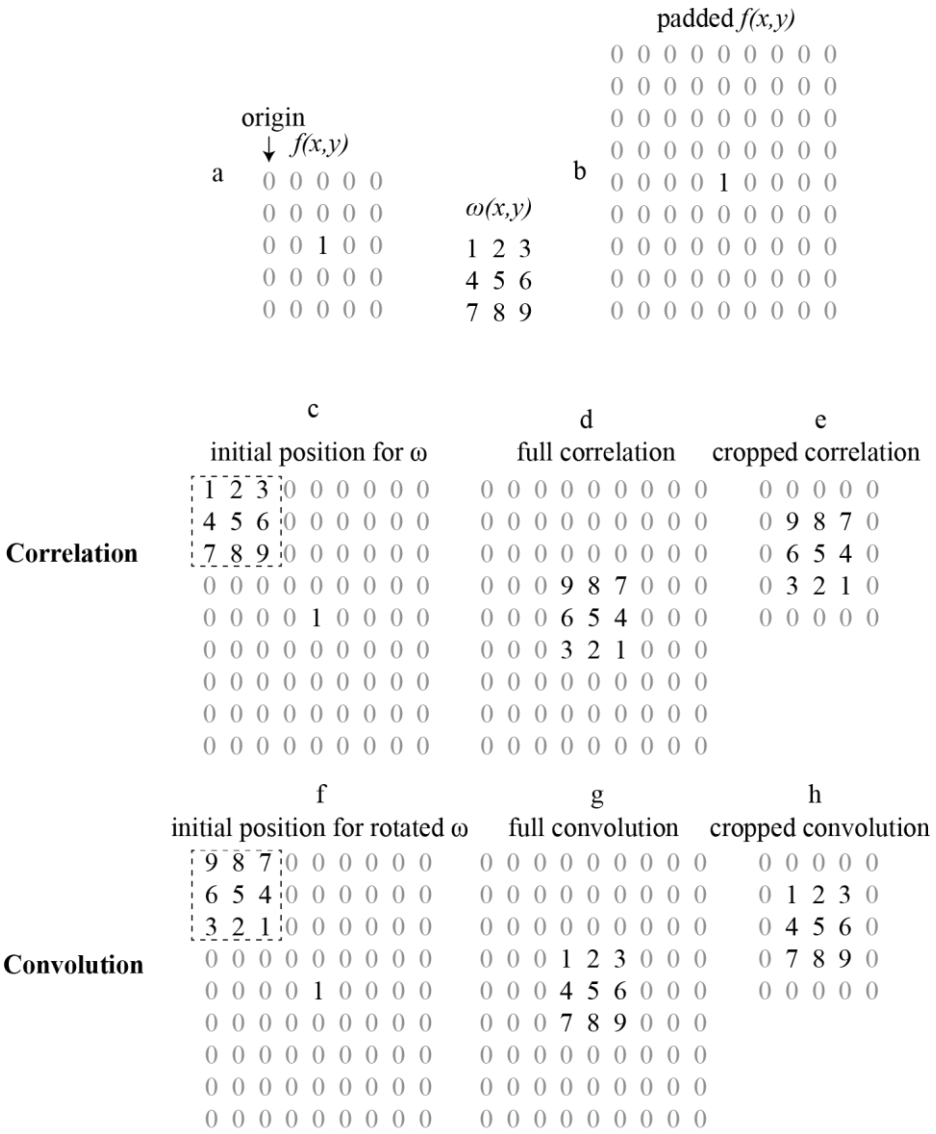

Fig. 4.1.3: correlation (middle row) and convolution (final row) of a 2D filter ω with a 2D discrete unit impulse f (original function represented in panel a, filter in panel b). In panel c the starting position of the filter on the padded image is represented, then the full correlation result is showed (panel d) together with the cropped result (e). Same for the convolution (panels f-g).

Fig. 4.1.3c shows the initial position of the filter mask for performing correlation, and *Fig. 4.1.3d* shows the full correlation result (in *Fig. 4.1.3e* the result is cropped). Note again that the result is rotated by 180°. For convolution process it is possible to obtain the same result if a pre-rotation of the mask is performed and the sliding sum of products is repeated. *Figs. 4.1.3f-h* show the convolution result. Clearly, if the filter mask is symmetric, correlation and convolution yield the same result. Correlation and convolution can be therefore used also to find matches between images, and this is the way it has been exploited for the realization of the machine. Formally, the correlation of a filter $\omega(x,y)$ of size $m \times n$ with an image $f(x,y)$, denoted as

$$\omega(x,y) \blacksquare f(x,y) \tag{4.1.5}$$

is given by the equation:

$$\omega(x,y) \blacksquare f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} \omega(s,t) f(x+s, y+t)\,. \tag{4.1.6}$$

This equation is evaluated for all values of the displacement variables x and y to permit all elements of ω to visit every pixel in f, assuming f has been padded appropriately; $a = \frac{(m-1)}{2}$ and $b = \frac{(n-1)}{2}$, and for notational convenience m and n are odd integers. In a similar manner, the convolution of $\omega(x,y)$ and $f(x,y)$, denoted by $\omega(x,y) \dagger f(x,y)$ is given by the expression

$$\omega(x,y) \dagger f(x,y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} \omega(s,t) f(x-s, y-t) \tag{4.1.7}$$

where the minus signs on the right flip f (i.e., rotate it by 180°). Flipping and shifting f instead of ω is done for notational simplicity and also to follow convention. This equation is evaluated, as for the correlation, for all values of the displacement variables x and y so that every element of ω visits every pixel in f, which we assume has been padded appropriately. To sum up, if correlation has to be performed, it must be inserted ω as an input into the algorithm; for convolution, the input ω has to be rotated by 180°.

4.1.2 Sharpening spatial filters

The principal objective of sharpening is to highlight intensity transitions, and this can be accomplished by spatial differentiation. In fact, the intensity of the response of a derivative operator is proportional to the discontinuity of the image where the operator is applied. Thus, image differentiation enhances edges and discontinuities while it deemphasizes areas with slowly or not varying intensities.

The Laplacian

The implementation of a 2D, second order derivatives algorithm is useful for image segmentation. The approach consists on defining a discrete formulation of the second-order derivative and then build up a filter mask based on that formulation. The main interest relies on

isotropic filters, whose responses are independent of the direction of the discontinuities: they are, in fact, rotationally invariant, meaning that rotating the image and then applying the filter gives the same result as applying the filter to the image first and then rotating the result. The simplest isotropic derivative operator is the Laplacian, which, for a function (image) $f(x, y)$ of two variables, is defined as

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \tag{4.1.8}$$

Because derivatives of any order are linear operations, the Laplacian is a linear operator. To express this equation in discrete form. In the x-direction, we have

$$\frac{\partial^2 f}{\partial x^2} = f(x+1, y) + f(x-1, y) - 2f(x, y) \tag{4.1.9}$$

and similarly, in the y-direction

$$\frac{\partial^2 f}{\partial y^2} = f(x, y+1) + f(x, y-1) - 2f(x, y) \tag{4.1.10}$$

Therefore, it follows from the preceding three equations that the discrete Laplacian of two variables is

$$\nabla^2 f = f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y). \tag{4.1.11}$$

Eq. 4.1.11 can be implemented using the filter mask in *Fig. 4.1.4a*, which gives an isotropic result for rotations in increments of 90°.

a

0	1	0
1	-4	1
0	1	0

b

1	1	1
1	-8	1
1	1	1

c

0	-1	0
-1	4	-1
0	-1	0

d

-1	-1	-1
-1	8	-1
-1	-1	-1

Fig. 4.1.4: panel (a), filter mask representing the Laplacian equation. Panel (b), extension of filter described in panel (a) including the diagonal terms. In panels (c, d) two other representations of the Laplacian filter are shown.

The mechanics of implementation of the filters are the same as the ones described before for the convolution and correlation. It is possible to obtain a mask that yields isotropic results in increments of 45° by incorporating the diagonal directions by adding two more terms to the equation of the Laplacian to one for each of the two diagonal directions. Each diagonal term also contains a $-2f(x,y)$ term, the total subtracted from the difference terms now would be $-8f(x,y)$ because four elements have to be added (*Fig 4.1.4b*). Other two filters are represented in *Fig. 4.1.4c* and *Fig. 4.1.4d*: they are obtained from definitions of the second derivatives that are the negatives of the ones used in the equation above, and as such, they yield equivalent results, but the difference in sign must be kept in mind when combining (adding or subtracting) a Laplacian-filtered image with another image. Because the Laplacian is a derivative operator, it will produce images that have grayish edge lines and other discontinuities, all superimposed on a dark, featureless background. Background features can be "recovered" while preserving the sharpening simply by adding the Laplacian image to the original. Thus, the basic way in which we use the Laplacian for image sharpening is

$$g(x,y) = f(x,y) + c[\nabla^2 f(x,y)] \tag{4.1.12}$$

where $f(x,y)$ and $g(x,y)$ are the input and sharpened images, respectively. The constant is $c = -1$ if the Laplacian filters in *Fig. 4.1.4a* or *Fig. 4.1.4b* are used, and $c = 1$ if either of the other two filters are used.

The Gradient

First derivatives in image processing are implemented using the magnitude of the gradient. For a function $f(x,y)$, the gradient of f at coordinates (x,y) is defined as the two-dimensional column vector

$$\nabla f = grad(f) = \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}. \tag{4.1.13}$$

This vector has the important geometrical property that it points in the direction of the greatest rate of change of f at location (x,y). The magnitude (length) of vector ∇f denoted as $M(x,y)$, where

$$M(x,y) = \sqrt{g_x^2 + g_y^2} \tag{4.1.14}$$

is the value at (x,y) of the rate of change in the direction of the gradient vector. Note that $M(x,y)$ is an image of the same size as the original, created when x and y vary over all pixel locations in f. This is the gradient image (or simply the gradient, when the meaning is clear). The partial derivatives exploited from the gradient are not rotationally invariant (isotropic), but the magnitude of the gradient vector is. In some implementations, it is more computationally suitable to approximate the squares and square root operations by absolute values:

$$M(x,y) \approx |g_x| + |g_y| \tag{4.1.15}$$

This expression still preserves the relative changes in intensity, but the isotropic property is lost in general. The most popular masks used to approximate the gradient are

isotropic at multiples of 90°. These results are independent of whether we use the magnitude as sum of squares or as sum of absolute values, therefore nothing of significance is lost in using the latter equation if we choose to do so. Discrete approximations, as for the Laplacian case, can be formulated to build the appropriate filter masks. In *Fig 4.1.5* some examples of 2×2 and 3×3 gradient filters are illustrated.

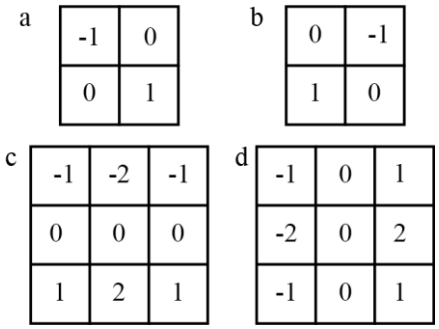

Fig. 4.1.5: panels (a, b) representation of 2×2 *gradient filters. In panels (c, d),* 3×3 *gradient filters are shown.*

4.1.3 Morphological image processing

Erosion and dilatation

Erosion and dilation are two fundamental operations to morphological processing. With A and B as sets in $\mathbb{Z}^2$, the erosion of A by B, denoted $A \ominus B$ is defined as

$$A \ominus B = \{z|(B)_z \subseteq A\}. \quad (4.1.16)$$

Eq. 4.1.16 indicates that the erosion of A by B is the set of all points such that B, translated by z, is contained in A. In the following, set B is assumed to be a structuring element, that is a small set or used to probe an image under study for properties of interest. The exact same definition of erosion can be achieved remembering that the statement "B has to be contained in A" is equivalent to "B not sharing any common elements with the background", therefore:

$$A \ominus B = \{z|(B)_z \cap A^c = \emptyset\} \quad (4.1.17)$$

where, A^c is the complement of A and $\emptyset$ is the empty set. *Fig. 4.1.6* shows an example of erosion.

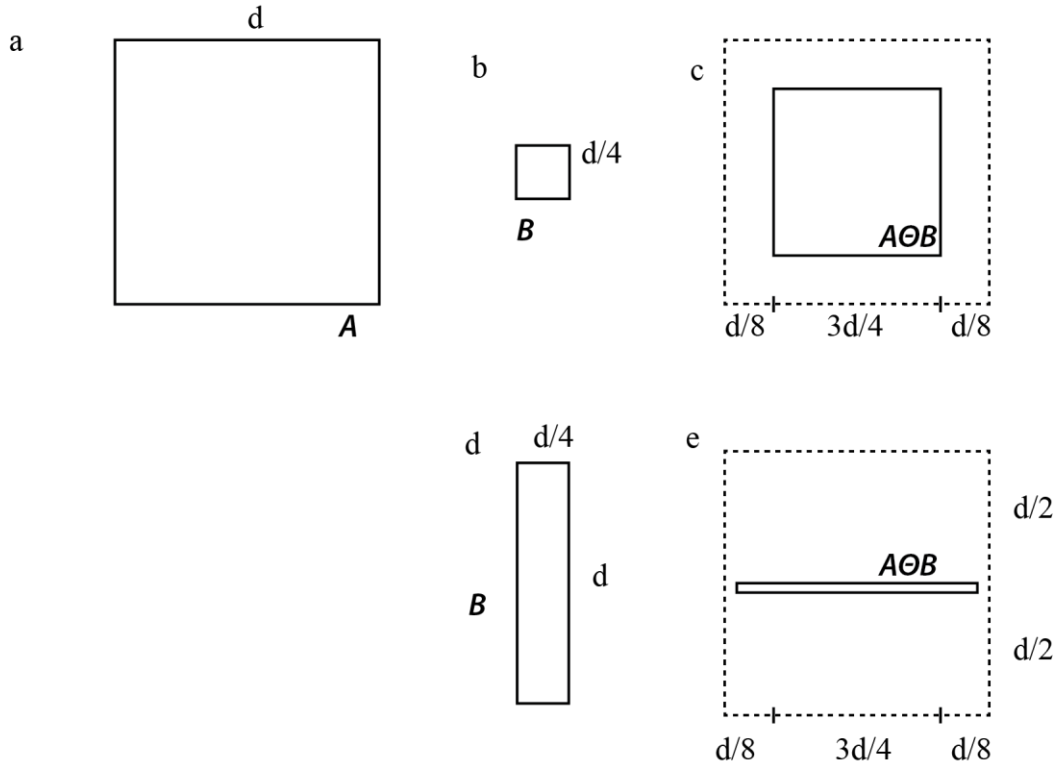

Fig. 4.1.6: representation of a set A (panel a), a square structuring element (b) and the erosion of A by B (c). Another structuring element (d), with shape different from the square one can be used to perform the erosion of A by B (e).

The solid boundary in *Fig. 4.1.6c* is the limit beyond which further displacements of the origin of B would cause the structuring element to cease being completely contained in A. Thus, the locus of points (locations of the origin of B) within (and including) this boundary constitutes the erosion of A by B. *Fig. 4.1.6d* shows structuring element with shape different to the square, and panel *e* shows the erosion of *A* by this element.

With A and B as sets in $\mathbb{Z}^2$, the dilation of *A* by *B* denoted $A \oplus B$ is defined as

$$A \oplus B = \left\{ z \middle| \left[\left(\hat{B} \right)_z \cap A \right] \neq \emptyset \right\}. \tag{4.1.18}$$

Eq. 4.1.18 is based on reflecting B about its origin and shifting this reflection by z. The dilation of *A* by *B* then is the set of all displacements z, such that $\hat{B}$ and *A* overlap by at least one element. Based on this interpretation, *Eq. 4.1.18* can be written equivalently as

$$A \oplus B = \left\{ z \middle| \left[\left(\hat{B} \right)_z \cap A \right] \subseteq A \right\} \tag{4.1.19}$$

As before, referring to *Fig. 4.1.7*, *B* is a structuring element and *A* is the set (image objects) to be dilated. The process of rotating *B* about its origin and then successively displacing it so that it slides over set (image) *A* is analogous to spatial convolution introduced before. Keep in mind, however, that dilation is based on set operations and therefore is a nonlinear operation, whereas convolution is a linear operation. Unlike erosion, which is a shrinking or thinning operation, dilation "grows" or "thickens" objects in a binary image. The specific manner and extent of this thickening is controlled by the shape of the structuring element used.

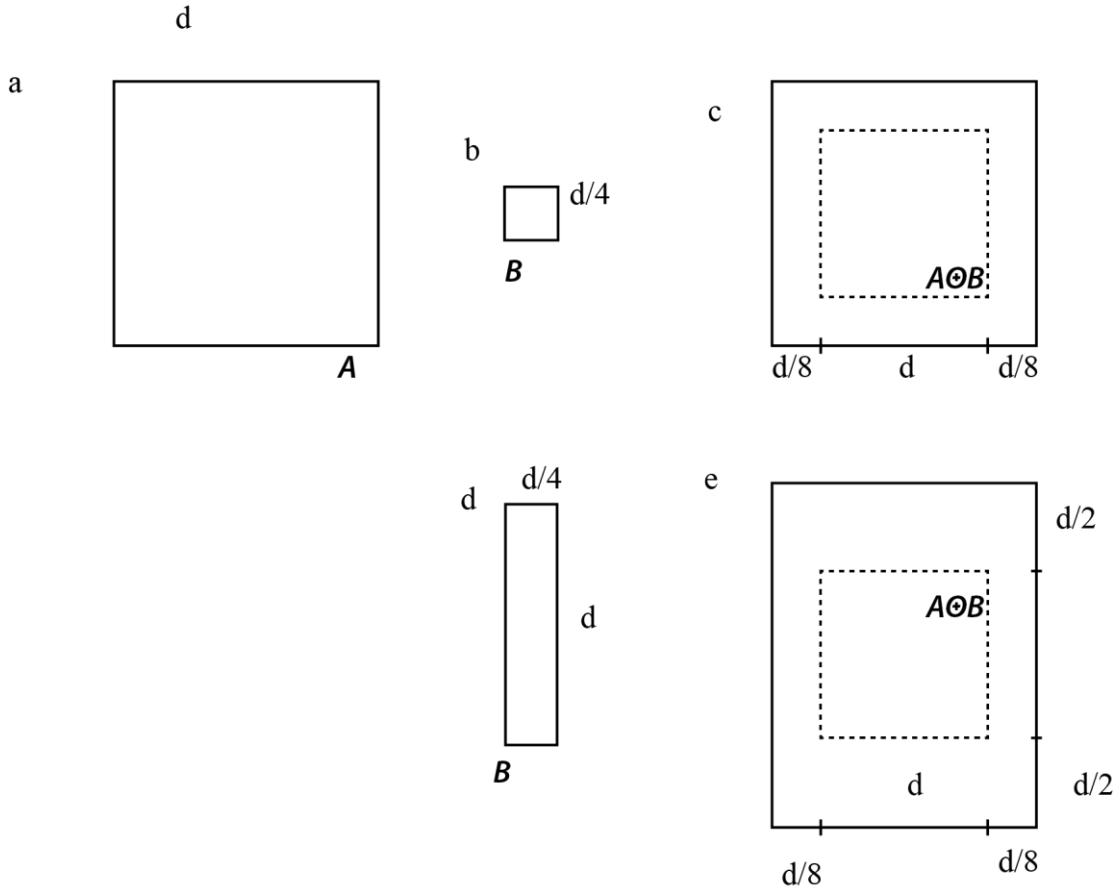

Fig. 4.1.7: representation of a set A (panel a), a square structuring element (b) and the dilatation of A by B (c). Another structuring element (d), with shape different from the square one can be used to perform the dilatation of A by B (e).

Fig. 4.1.7a shows the same set used in *Fig. 4.1.6a*, and *Fig. 4.1.7b* shows the structuring element (in this case $\hat{B} = B$ because it is symmetric about its origin) taken into account. The dashed line in *Fig. 4.1.7c* shows the original set for reference, and the solid line shows the limit beyond which any further displacements of the origin of $\hat{B}$ by z would cause the intersection of $\hat{B}$ and A to be empty. Therefore, all points on and inside this boundary constitute the dilation of A by B. *Fig. 4.1.7d* shows a different structuring element, and *Fig. 4.1.7e* shows the dilation achieved with this element.

Opening and closing

Dilation expands the components of an image and erosion shrinks them. Opening and closing are two important morphological operations which can be obtained starting from dilatation and erosion. Opening generally smooths the contour of an object, breaks narrow edges, and eliminates thin protrusions. Closing also tends to smooth sections of contours but, as opposed to opening, it generally eliminates small holes, and fills gaps in the contour. The opening of set A by structuring element B, denoted $A \circ B$, is defined as

$$A \circ B = (A \ominus B) \oplus B \tag{4.1.20}$$

thus, the opening A by B is the erosion of A by B followed by a dilation of the result by B. Similarly, the closing of set A by structuring element B, denoted $A \cdot B$ is defined as

$$A \cdot B = (A \oplus B) \ominus B \tag{4.1.21}$$

which states that the closing of A by B is simply the dilation of A by B, followed by the erosion of the result by B. The opening operation has a simple geometric interpretation *(Fig. 4.1.8*).

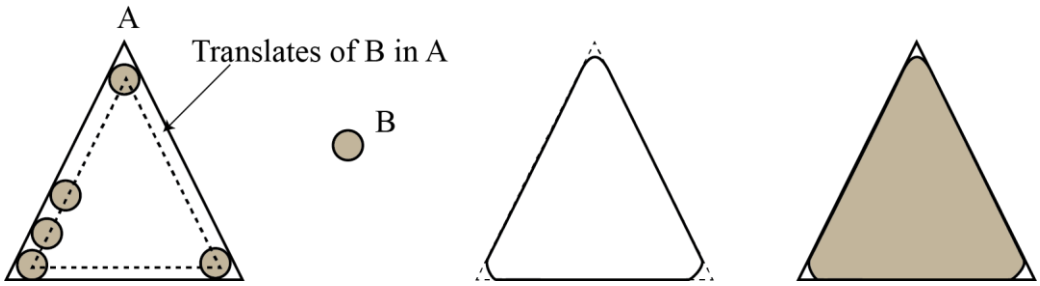

Fig. 4.1.8: the structuring element B is made rolling along the border of A (on the inside) obtaining the opening (right).

Suppose that the structuring element B is represented as a ball. The boundary of $A \circ B$ is then determined by the points in B that reach the farthest into the boundary of A as B is rolled around the inside of this boundary. This leads to a set-theoretic formulation, which states that the opening of A by B is obtained by taking the union of all translates of B that match into A. Therefore, opening can be expressed as a fitting process such that

$$A \circ B = \bigcup \{(B)_z | (B)_z \subseteq A\} \tag{4.1.22}$$

where the big $\cup\{\cdot\}$ denotes the union of all the sets inside the braces. Closing has a similar geometric interpretation, except that now B is rolled on the outside of the boundary (Fig. 4.1.9).

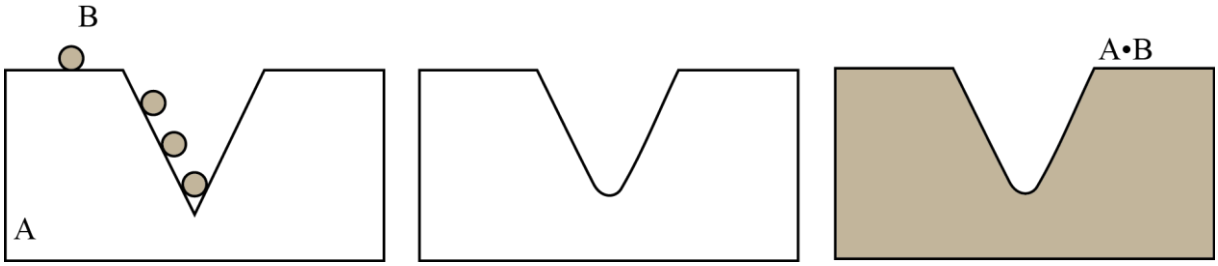

Fig. 4.1.9: the structuring element B is made rolling along the border of A (on the outside) obtaining the closing (right).

Opening and closing are duals of each other. Geometrically, a point ω is an element of $A \cdot B$ if and only if $\{(B)_z \cap A \neq \emptyset\}$ for any translate $(B)_z$ of that contains ω. *Fig. 4.1.9* illustrates the basic geometrical properties of closing. As in the case with dilation and erosion, opening and closing are duals of each other with respect to set complementation and reflection, which means

$$(A \cdot B)^c = (A^c \circ \hat{B}) \tag{4.1.23}$$

and

$$(A \circ B)^c = \left(A^c \cdot \hat{B}\right). \tag{4.1.24}$$

Of importance is the boundary extraction of a feature. The boundary of a set A, denoted by $\beta(A)$, can be obtained by first eroding A by a suitable structuring element B and then performing the set difference between A and its erosion, as reported in *Eq. 4.1.25*

$$\beta(A) = A - (A \ominus B). \tag{4.1.25}$$

4.1.4 Image segmentation

Segmentation algorithms are based on discontinuity and similarity. In the first category, the partition of an image is based on abrupt changes in intensity, such as edges (edge-based segmentation).[153,154] The principal approaches exploited in the second category are based on partitioning an image into regions that are similar according to a set of predefined criteria (region-based segmentation).[155,156]

Edge-based segmentation

For line detection, second derivatives result in a stronger response and produce thinner lines than first derivatives. Thus, a Laplacian mask can be exploited for the purpose. Often, interest lies in detecting lines in specific directions. Consider the masks in *Fig. 4.1.10a*. Suppose that an image with a constant background and containing various lines (oriented at 0°, 45° and 90°) is filtered with the first mask. The maximum responses would occur at image locations in which a horizontal line passed through the middle row of the mask. A similar experiment would reveal that the mask depicted in *Fig. 4.1.10b* responds best to lines oriented at +45°, the one represented in *Fig. 4.1.10c* mask to vertical lines, and the one reported in *Fig. 4.1.10d* to lines in the -45° direction. The preferred direction of each mask is weighted with a larger coefficient (i.e., 2) than other possible directions. The coefficients in each mask sum to zero, indicating a zero response in areas of constant intensity. If the interest lies in detecting all the lines in an image in the direction defined by a given mask, it is sufficient to run the mask through the image and threshold the absolute value of the result. The points that are left are the strongest responses which correspond closest to the direction defined by the mask.

a

-1	-1	-1
2	2	2
-1	-1	-1

b

2	-1	-1
-1	2	-1
-1	-1	2

c

-1	2	-1
-1	2	-1
-1	2	-1

d

-1	-1	2
-1	2	-1
2	-1	-1

Fig. 4.1.10: filters exploited for the line detection: in particular, they focus on horizontal (a), +45° (b), vertical (c) and -45° (d) lines.

Edge detection is the approach used most frequently for segmenting images based on abrupt local changes in intensity. Edge models are classified according to their intensity profiles. A step edge involves a sudden transition between two intensity levels, ideally over the distance of 1 pixel (*Fig. 4.1.11a*).

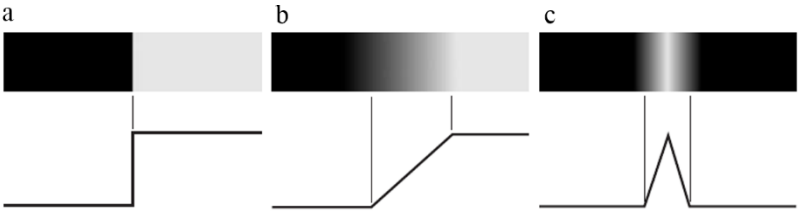

Fig. 4.1.11: representation of step (a), ramp (b) and roof edges (c) with their corresponding intensities. Adapted from the literature.[157]

These ideal edges can occur over the distance of 1 pixel, provided that no additional processing (such as smoothing) is used to make them look "real". In practice, digital images have edges that are blurred and noisy; in such situations, edges are more closely modeled as having an intensity ramp profile, such as the one represented in *Fig. 4.1.11b*. The edge point now is any point contained in the ramp, and an edge segment would then be a set of such points that are connected. A third model of an edge is the so-called roof edge, having the characteristics illustrated in *Fig. 4.1.11c*. Roof edges are models of lines through a region, with the base (width) of a roof edge being determined by the thickness and sharpness of the line. In the limit, when its base is 1 pixel wide, a roof edge is really nothing more than a 1-pixel-thick line running through a region in an image. It is not unusual to find images that contain all three types of edges. Although blurring and noise result in deviations from the ideal shapes, edges in images that are reasonably sharp and have a moderate amount of noise do resemble the characteristics of the edge models in *Fig. 4.1.11*. The models in *Fig. 4.1.11* allow us is to write mathematical expressions for edges in the development of image processing algorithms. The performance of these algorithms will depend on the differences between actual edges and the models used in developing the algorithms.

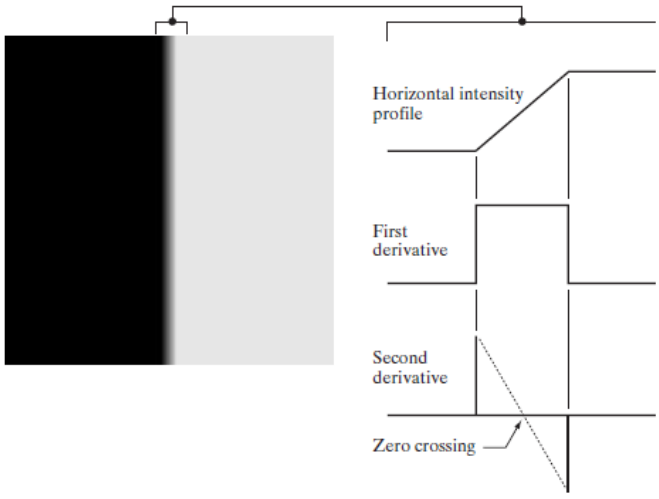

Fig. 4.1.12: on the left, enlargement of the ramp depicted in figure 4.1.11b. On the right, description of the profile together with its first and second derivatives. Adapted from the literature.[157]

The left part of *Fig. 4.1.12* shows an enlargement of the step represented in *Fig. 4.1.11b*, while on the right a horizontal intensity profile is depicted. This figure shows also the first and second derivatives of the intensity profile. Moving from left to right, the first derivative takes a positive value at the onset of the ramp and at points on the ramp, while it is equal to zero in

areas of constant intensity. Obviously, the second derivative is positive at the beginning of the ramp, negative at the end of the ramp, zero at points on the ramp, and zero at points of constant intensity. The intersection between the zero intensity axis and a line extending between the extrema of the second derivative marks a point called the zero crossing of the second derivative. The magnitude of the first derivative can be used to detect the presence of an edge at a point in an image. Similarly, the sign of the second derivative can be used to determine whether an edge pixel lies on the dark or light side of an edge. However, the conclusions reached using those models are the same as with an ideal ramp and working with the latter simplifies theoretical formulations. Finally, although attention thus far has been limited to a 1D horizontal profile, a similar argument applies to an edge of any orientation in an image. We simply define a profile perpendicular to the edge direction at any desired point and interpret the results in the same manner as for the vertical edge just discussed.[158]

Region based segmentation - Maximally stable extremal region

Extraction of invariant regions has been the focus of intense study supporting a wide variety of applications such as recognition, image retrieval, 3D reconstruction, tracking, robot navigation and more.[159,160] Maximally Stable Extremal Regions (MSER) described in the literature (Matas et al.)[161] have become one of the commonly used region detector types, due to their high repeatability and because of their complementarity to many other commonly used detectors. They have commonly been used in recognition as well as tracking and these two fields are the one exploited in this thesis. The standard algorithms for computing MSER are similar to a flooding simulation algorithm for computing a watershed segmentation.[162,163] Maximally stable extremal region (MSER) is a technique to detect particular features in a digital image. The formal definition of the MSER concept and the fundamental auxiliary definitions as introduced by Matas et al. are given in the definitions below:

Image I is mapping $I: D \subset \mathbb{Z}^2 \rightarrow S$. Extremal regions are well defined on images if:

- S is totally ordered, i.e. reflexive, antisymmetric and transitive binary relation exists. In this thesis only $S = \{0,1, \dots ,255\}$ is considered.
- A neighborhood relation $A \subset D \times D$ is defined.

Region: Q is a contiguous subset of D, i.e. for each $p, q \in Q$ there is a sequence $p, a_1, a_2, \dots, a_n, q$ and $pAa_1, a_iAa_{i+1}, \dots, a_nAq$.
Outer Region Boundary: $\partial Q = \{q \in D \backslash Q: \exists p \in Q: qAp\}$ i.e. the boundary ∂Q of Q is a set of pixels being adjacent to at least one pixel of Q but not belonging to Q.
Extremal Region: $Q \subset D$ is a region such that for all $p \in Q, q \in \partial Q: I(p) > I(q)$ (maximum intensity region) or $I(p) < I(q)$ (minimum intensity region).
Maximally Stable Extremal Region (MSER): let $Q_1, \dots, Q_{i-1}, Q_i, \dots$ be a sequence of nested extremal region, i.e. $Q_i \subset Q_{i+1}$. Extremal region Q_{i*} is maximally stable if $q(i) = |Q_{i+\Delta} \backslash Q_{i-\Delta}|/|Q_i|$ has a local minimum at i^*. (where $|\cdot|$ denotes cardinality and $\Delta \in S$ is a parameter).

The above definitions can be interpreted as follow. The first step to detect MSERs is to compute all the binary regions by thresholding the image at all possible gray levels. Suppose that all the pixels with intensity below a given threshold are white and all those above or equal are black. At the beginning, the image will be totally black, but performing the MSER procedure there will be white spots which will merge and at the end the image would be completely white.

While changing the threshold, the area of each region is monitored; regions whose rate of change of area with respect to the threshold is minimal are defined as "maximally stable". Such regions are therefore invariant to both affine geometric and photometric (linear bias-gain) transformations. The word extremal refers to the property that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary. Over a large range of thresholds, the local binarization is stable in certain regions, and have the following properties: 1) they are invariant to affine transformation of image intensities 2) only the regions whose support is nearly the same over a range of thresholds are selected 3) can be applied to both large and fine structure.[164]

Region based segmentation - Marching squares

One approach to isosurface construction will be now described, the widely used Marching Cubes algorithm by Lorensen and Cline.[165] The Marching Cubes algorithm is based on two ideas. First, the isosurface can be constructed piecewise within each cube of the grid without reference to other grid cubes. Second, the combinatorial structure of each isosurface patch in a grid cube can be retrieved from a lookup table. Since the main operation is retrieving this structure from the lookup table, the algorithm runs in time proportional to the number of grid cubes. In particular, in this thesis the 2D version of the algorithm has been exploited, the so-called Marching Squares, by means of which it is possible to construct two-dimensional isocontours. Given a regular scalar grid and an isovalue σ, it is convenient to assign "+" and "–" labels to each grid vertex based on the relationship between every single scalar value and the selected isovalue σ. A grid vertex is positive, "+", if its scalar value is greater than or equal to σ (strictly positive if its scalar value is not equal to σ) and it is negative ("–") if its scalar value is less than σ. Since the scalar value of a negative vertex never equals the isovalue, there is no point in defining a similar "strictly negative" term. Grid edges can be characterized by the labels at their endpoints and they can be positive, if both endpoints are positive, negative, if both are negative and bipolar if one endpoint is positive and the other one is negative. Note that a grid vertex or edge is only positive or negative in relationship to a given isovalue. The above definitions apply also to curvilinear grids. They also apply to the vertices and edges of polyhedral meshes such as tetrahedral and simplicial meshes. The inputs to the Marching Squares algorithm are so an isovalue and a set of scalar values at the vertices of a two-dimensional regular grid. The algorithm has three steps:

1) Read in the isocontour lookup table.
2) For each square, find the corresponding set of isocontour edges (from the table), correctly describing the combinatorial structure of the isocontour in analysis, remembering that the endpoints of these edges form the isocontour vertices.
3) Assign geometric locations to the isocontour vertices based on the scalar values at the square edge endpoints.

A square has four vertices, therefore there are $2^4 = 16$ different configurations of square vertex labels. These configurations are listed in *Fig. 4.1.13*.

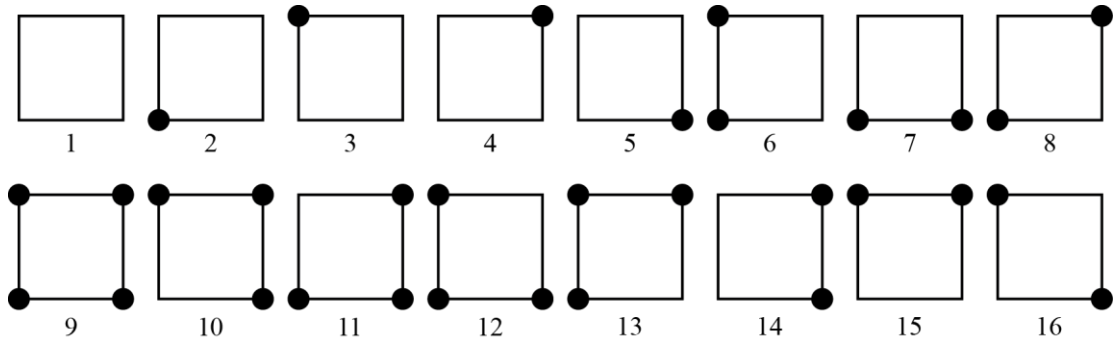

Fig. 4.1.13: representation of the 16 square configurations. The black vertices have values greater than the isovalue.

The structure of the isocontour within each square is ruled by the configuration of the square's vertex labels. The isocontour must intersect any square edge that has one positive and one negative endpoint. For each square configuration k, let $E_k^{+/-}$ be the set of bipolar edges. Note that the size of $E_k^{+/-}$ is either zero, two, or four, because the isocontour can touch the edges zero, two or four times. Pair the edges of $E_k^{+/-}$. Each such pair represents an isocontour edge with endpoints on the two elements of the pair. *Fig. 4.1.14* contains the sixteen square configurations and their isocontours. There, the isocontour edges are drawn connecting the midpoints of each square edge. This is for illustration purposes only. The geometric locations of the isocontour vertices are not defined by the lookup table.

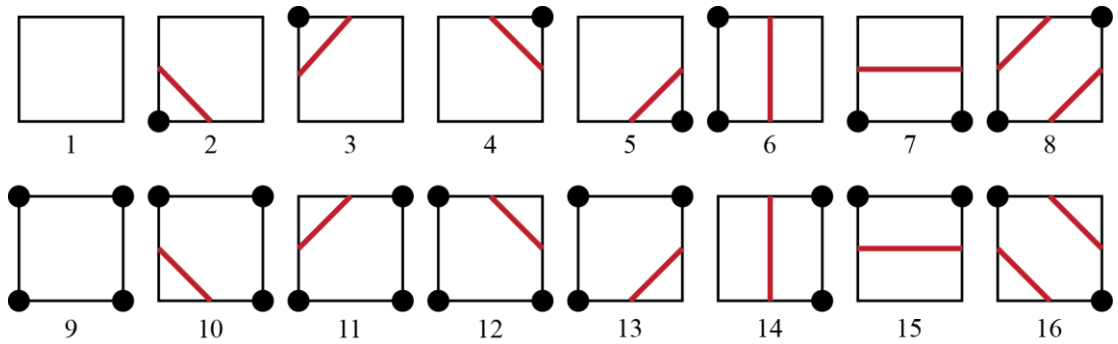

Fig. 4.1.14: all the possible square isocontours obtainable from the marching square algorithm.

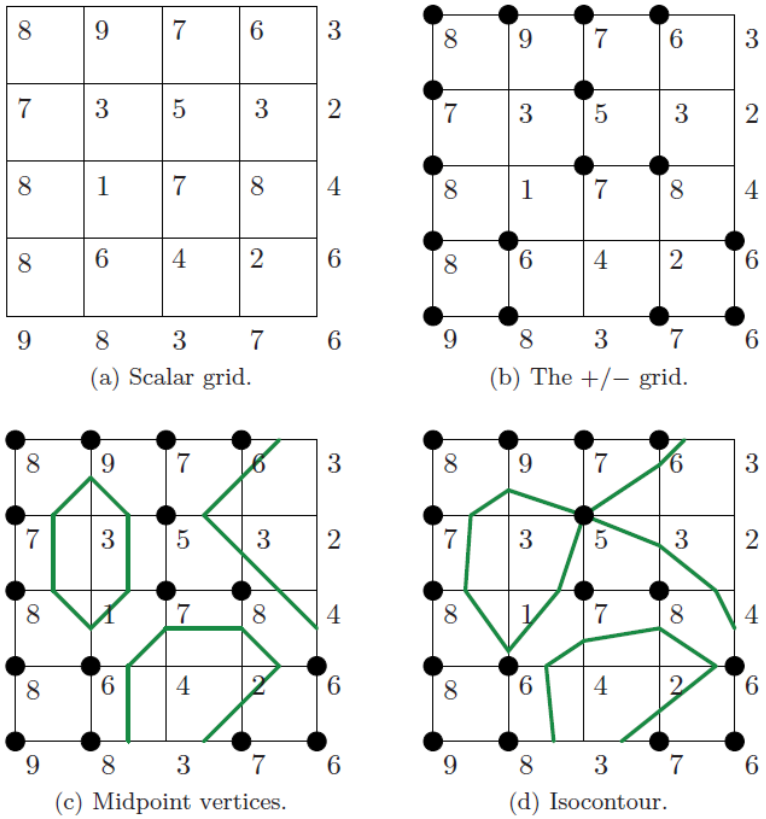

Fig. 4.1.15: example of isocontour with isovalue equals to 5. Panel a represents the grid, while in panel b the black spot is assigned to values greater than the isovalue. In panel c the vertices assignation is performed while in panel d the isocontour are drawn. Figure adapted from the literature.[166]

By means of the Marching Squares algorithm and selecting a proper isovalue it is possible to extract contours which are fundamental for the detection of the defects as will be expressed in the next chapter, as it is schematized in *Fig. 4.1.15*.[167,168]

4.2 Introduction to Support Vector Machine

Statistical learning theory is a framework built for machine learning purposes associated with statistics and functional analysis. The aim of statistical learning is to find a function f, which will be used for prediction and inference, based on the machine response to some predictor elements. To estimate this unknown function, a supervised approach is exploited, which suggests the usage of a training method to teach the machine how to estimate f. In fact, a given set of labeled data consisting of input and outputs are used to train the specific desired model. Every point in the training data is an input-output pair where both the input and output are well classified by the scientists. The statistical leaning problem consists of inferring the function that maps the input and the output, such that the learned function can be used to predict output from future input. A classification algorithm attempts to regulate the experimental data into different and distinct categories to locate new data in future easily. It consists of two main parts: model training and prediction. Support Vector Machines (SVMs) are sets of supervised learning methods with associated leaning algorithms exploited to analyze data. They have first been introduced as methods to solve classification problems.[169] However, due to many attractive features, they have recently been extended to the area of regression analysis. One of the main and hot topics in which SVMs are exploited is text categorization.[170,171] It consists in assigning labels or categories to a text according to its context (web news organization, academic paper classification, spam filtering, …) by means of linear SVM. In this thesis, the classification of defects arising on reflectors and the characterization of their optical properties have been implemented by means of SVM-based algorithms.

4.2.1 How does SVM work?

SVM algorithms generate learning models for automatic data classification based on a set of given data. Suppose that, in a 2D space, two different types of data are present, represented in *Fig. 4.2.1* as circles and squares. The aim is to find the best way to separate the two families. It is possible to exploit different straight lines, denominated H_1 and H_2 and represented in *Fig. 4.2.1a*. In the present case, the best separators are represented by straight lines, but in a more general case (3D or greater dimensions) they are addressed as hyperplanes. If one hyperplane is selected for the classifier model, then when a new data is received, the decision on its shape depends on its position relative to the hyperplane. Another possibility is to use nonlinear separator when data cannot be separated linearly. An example of a two-dimensional nonlinear separator has been shown in *Fig. 4.2.1b*. Decision between linear or nonlinear classification depends on distribution of samples in the input data.[172]

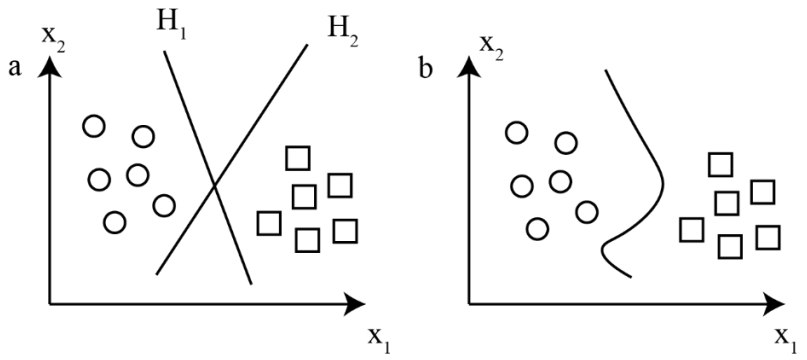

Fig. 4.2.1: examples of linear (a) and curvilinear (b) separators.

4.2.2 The Linear SVM

In SVM, the goal is to find the best separator line for a given set of data. Consider the following set containing N samples:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \tag{4.2.1}$$

x_i is the input vector in $\mathbb{R}^q$ and y_i is the corresponding output. In this general formulation of q-dimensional classification problem, the hyperplanes are subspaces of dimension $q - 1$ like straight lines in two-dimensional space.[173] To make things clearer, consider the problem discussed before which involves inputs in $\mathbb{R}^2$. Data are linearly separable and different hyperplanes (in this specific case, planes) can perform the separation, defined by the equation $w_1 x_1 + w_2 x_2 + b = 0$. For the classification of linearly separable data, this idea is to find among all the hyperplanes that minimize the error, the one with the largest margin.

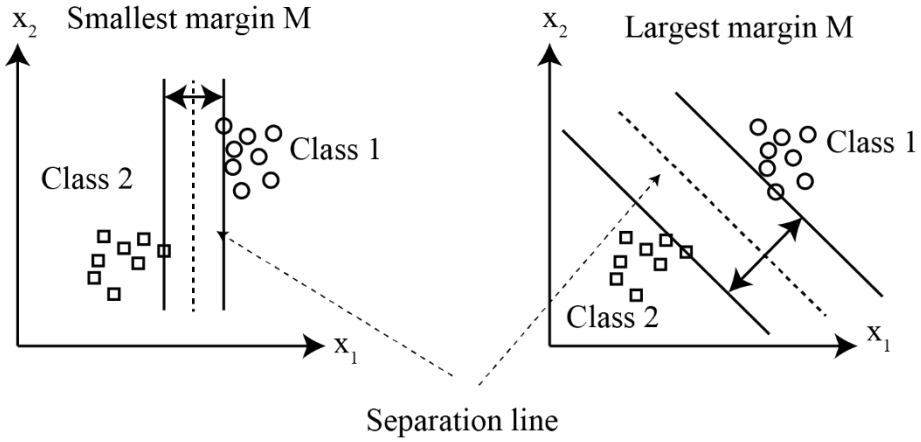

Fig. 4.2.2: a vertical separation (left) may not be the most correct choice. In fact, as evidenced by the graph on the right, an oblique separation line has a bigger margin for the case presented.

By using given training examples, during the learning stage, the machine finds parameters $\boldsymbol{w} = (w_1, w_2, \dots, w_n)$ and b of a decision function $d(\boldsymbol{x}, \boldsymbol{w}, b)$ given as

$$d(\boldsymbol{x}, \boldsymbol{w}, b) = \boldsymbol{w} \cdot \boldsymbol{x} + b = \sum_{i=1}^{n} w_i x_i + b. \tag{4.2.2}$$

After the successful training stage, the learning machine, by using the weights obtained and given a previously unseen pattern $\boldsymbol{x}_p$, produces output 0 (zero) according to an indicator function given as

$$i_F = 0 = sign(d(\boldsymbol{x}_p, \boldsymbol{w}, b)) \tag{4.2.3}$$

where 0 is the standard notation for the output from the learning machine. In other words, the decision rule is:

- if $d(\boldsymbol{x}_p, \boldsymbol{w}, b) > 0$ the pattern $\boldsymbol{x}_p$ belongs to class 1 ($y_1 = +1$)
- if $d(\boldsymbol{x}_p, \boldsymbol{w}, b) < 0$ the pattern $\boldsymbol{x}_p$ belongs to class 2 ($y_2 = -1$)

It is important to introduce the expression for the calculation of a distance (the margin, indicated with the letter M) between the closest members from two diverse classes: $M = \frac{2}{\|\boldsymbol{w}\|}$. From the formula, it is clear that the minimization of a norm of a hyperplane normal weight vector $\|\boldsymbol{w}\| = \sqrt{(\boldsymbol{w}^T\boldsymbol{w})}$ leads to a maximization of a margin M. The square root is a monotonic function, therefore the minimization of $\|\boldsymbol{w}\|$ equals to the minimization of $\boldsymbol{w}^T\boldsymbol{w} = \sum_{i=1}^{n} w_i^2$. Hence, the learning problem is

$$minimize\ \frac{1}{2}(\boldsymbol{w}^T\boldsymbol{w}) \tag{4.2.4}$$

(multiplication by 0.5 is for numerical convenience only). The separating hyperplane with the largest margin defined by $M = \frac{2}{\|\boldsymbol{w}\|}$, is denominated Optimal Canonical Hyperplane (OCH) and specifies the support vectors, i.e. training data points which satisfy the equation

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) = 1 \tag{4.2.5a}$$

for all $i = 1,2, \dots, N_{SV}$. For all the other (non-SVs) data points the OCH satisfies inequalities

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) > 1. \tag{4.2.5b}$$

In general, OCH should satisfy the following constraints

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1, \qquad i = 1,2, \dots, n \tag{4.2.5c}$$

where n denotes a number of training data points, and N_{SV} stands for a number of SVs. This is a quadratic optimization problem with inequality constraints, which can be solved by the saddle point of the Lagrange functional (Lagrangian)

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1] \tag{4.2.6}$$

where α_i are Lagrange multipliers. The search for an optimal saddle point $(\boldsymbol{w}_0, b_0, \alpha_0)$ is necessary because the Lagrangian must be minimized with respect to w and b, and maximized with respect to nonnegative α_i ($\alpha_i \geq 0$ should be found). This problem can be solved either in a primal space (which is the space of parameters w and b) or in a dual space (which is the space

of Lagrange multipliers α_i). The second approach gives insightful results. In order to do that, Karush-Kuhn-Tucker (KKT) conditions must be used for the optimum of a constrained function. These conditions imply that at the saddle point ($\boldsymbol{w}_0, b_0, \alpha_0$), the derivatives of the Lagrangian L with respect to primal variables should vanish

$$\frac{\partial L}{\partial \boldsymbol{w}_0} = 0 \tag{4.2.7a}$$

$$\frac{\partial L}{\partial b_0} = 0 \tag{4.2.7b}$$

therefore,

$$\boldsymbol{w}_0 = \sum_{i=1}^{n} \alpha_i \, y_i \boldsymbol{x}_i \tag{4.2.8a}$$

$$\sum_{i=1}^{n} \alpha_i \, y_i = 0 \tag{4.2.8b}$$

and the KKT complementarity conditions below (products between dual variables and constraints at the solution point must be equals zero) must also be satisfied,

$$\alpha_i [y_i (\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1] = 0. \tag{4.2.9}$$

and performing the substitution, it is possible to obtain

$$L_d(\alpha) = -\frac{1}{2} \sum_{i,j=1}^{n} \alpha_i y_i \alpha_j y_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) + \sum_{i=1}^{n} \alpha_i. \tag{4.2.10}$$

To find the correct hyperplane, the dual Lagrangian $L_d(\alpha)$ has to be maximized with respect to nonnegative α_i and with respect to the equality constraint as follows

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \qquad i = 1,2, \dots, n \tag{4.2.11a}$$

$$\alpha_i \geq 0, \qquad i = 1,2, \dots, n \tag{4.2.11b}$$

Note that the dual Lagrangian $L_d(\alpha)$ is expressed in terms of training data and depends only on the scalar products of input patterns ($\boldsymbol{x}_i \cdot \boldsymbol{x}_j$). Note also that the number of unknown variables equals the number of training data *n*. After the learning process, the number of free parameters is equal to the number of SVs, which is not depend on the dimensionality of input space. Such a standard quadratic optimization problem can be expressed in a matrix notation (out of the scope of this thesis). Solutions α_{0i} of the dual optimization problem above determine the parameters $\boldsymbol{w}_0$ and b_0 of the optimal hyperplane

$$\boldsymbol{w}_0 = \sum_{i=1}^{n} \alpha_{0i} \, y_i \boldsymbol{x}_i \tag{4.2.12a}$$

$$b_0 = \frac{1}{N_{SV}}\left(\sum_{s=1}^{N_{SV}}\left(\frac{1}{y_s} - \boldsymbol{x}_s \cdot \boldsymbol{w}_0\right)\right). \tag{4.2.12b}$$

The fact that the summation in *Eq. 4.2.12a* goes over all training data patterns (i.e., from 1 to n) is irrelevant because the Lagrange multipliers for all non-support vectors (NSV) equal zero (α_{0i} = 0, for i = NSV + 1,…, n). Finally, having calculated $\boldsymbol{w}_0$ and b_0 a decision hyperplane $d(\boldsymbol{x})$ and an indicator function i_F are obtained

$$d(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_{0i} x_i + b_0 = \sum_{i=1}^{n} \alpha_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b_0 \tag{4.2.13a}$$

$$i_F = 0 = sign\big(d(\boldsymbol{x})\big). \tag{4.2.13b}$$

Training data patterns having non-zero Lagrange multipliers are called support vectors. For linearly separable training data, all support vectors lie on the margin and they are generally just a small portion of all training data (typically, $NSV << n$).

Linear soft margin classifier for overlapping classes

The learning procedure presented before is valid for linearly separable data which means, from the experimental point of view, training data sets with no overlapping data. Quadratic programming solutions as explained above cannot be used in the case of overlapping because the constraints $y_i[\boldsymbol{w} \cdot \boldsymbol{x}_i + b] \geq 1$, for $i = 1, \ldots, n$ cannot be satisfied. The overlapped data points cannot be correctly classified, and this means that for any misclassified training data point $\boldsymbol{x}_i$, the corresponding α_i will tend to infinity. In such a situation, the algorithm chooses almost all training data points as support vectors. To find a classifier with a maximal margin, the algorithm must be changed allowing some data to be unclassified or leaving some of them as incorrectly classified. In practice, a soft margin is allowed and all data inside this margin (whether on the correct or wrong side of the separating line) are neglected. The width of a soft margin can be controlled by a corresponding penalty parameter C. Similarly to the case presented before, the goal is to minimize the equation

$$minimize: \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C(number\ of\ misclassified\ data) \tag{4.2.14}$$

where C is a penalty parameter which takes into account the margin size (defined by $\boldsymbol{w}$) and the number of misclassified points. Large C leads to small number of misclassifications, and consequently to a smaller margin and vice versa. Obviously taking $C = \infty$ requires that the number of misclassified data is zero and, in the case of an overlapping, this is not possible. Hence, the problem may be feasible only for some value $C < \infty$.

To be more precise, it is convenient to measure the distances ξ_i of the points in the region enclosed by the two margins with respect to the corresponding and trade their sum for the margin size as given by *Eq. 4.2.15*

$$minimize: \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C(sum\ of\ distances\ of\ the\ wrong\ side\ points). \tag{4.2.15}$$

This is exactly how the problem of the data overlapping was solved in the literature. The separating hyperplane must now satisfy

$$minimize: \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}\xi_i \tag{4.2.16}$$

subject to

$$y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geq 1 - \xi_i, \qquad for\ i = 1, \dots, n, \qquad \xi_i \geq 0 \tag{4.2.17}$$

which means

$$(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geq 1 - \xi_i, \qquad for\ y_i = +1, \qquad \xi_i \geq 0 \tag{4.2.18a}$$

$$(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \leq -1 + \xi_i, \qquad for\ y_i = -1, \qquad \xi_i \geq 0. \tag{4.2.18b}$$

Hence, for such a generalized optimal separating hyperplane, the functional to be minimized comprises an extra term. The function above can be even more general

$$minimize: \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}{\xi_i}^k \tag{4.2.19}$$

subject to the same constraints. This is a convex programming problem that is usually solved only for k = 1 or k = 2, and such soft margin SVMs are dubbed L1 and L2 SVMs respectively. By choosing exponent k = 1, neither ξ_i nor their Lagrange multipliers β_i appear in a dual Lagrangian L_d. Same as for a linearly separable problem presented previously, for L1 SVMs (k = 1), the solution to a quadratic programming problem, is given by the saddle point of the primal Lagrangian $L_P(\boldsymbol{w}, b, \xi, \alpha, \beta)$ shown below

$$\begin{aligned} L_P(\boldsymbol{w}, b, \xi, \alpha, \beta) = {} & \frac{1}{2}\|\boldsymbol{w}\|^2 + C\left(\sum_{i=1}^{n}\xi_i\right) \\ & - \sum_{i=1}^{n}\alpha_i[y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) - 1 + \xi_i] \\ & - \sum_{i=1}^{n}\beta_i\xi_i, \qquad for\ L1\ SVM \end{aligned} \tag{4.2.20}$$

where α_i and β_i are the Lagrange multipliers. The optimal saddle point $(\boldsymbol{w}_0, b_0, \xi_0, \alpha_0, \beta_0)$ must be found in order for the Lagrangian L_P to be minimized with respect to w, b, ξ and maximized with respect to nonnegative α_i and β_i. As before, the solution in a dual space is considered

$$\frac{\partial L}{\partial \boldsymbol{w}_0} = 0 \ \ i.e.\ \ \boldsymbol{w}_0 = \sum_{i=1}^{n}\alpha_i\, y_i\boldsymbol{x}_i \tag{4.2.21a}$$

$$\frac{\partial L}{\partial b_0} = 0 \ \ i.e.\ \ \sum_{i=1}^{n}\alpha_i\, y_i = 0 \tag{4.2.21b}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \;\; i.e. \;\; \alpha_i + \beta_i = C \tag{4.2.21c}$$

and the KKT complementarity conditions below

$$\alpha_i[y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1 + \xi_i] = 0 \quad for\ i = 1, \dots, n \tag{4.2.22a}$$

$$\beta_i \xi_i = (C - \alpha_i)\xi_i = 0 \quad for\ i = 1, \dots, n. \tag{4.2.22b}$$

At the optimal solution, due to the KKT conditions, the last two terms in the primal Lagrangian L_P vanish and the dual variables Lagrangian $L_d(\alpha)$, for L1 SVM, is not a function of β_i

$$L_d(\alpha) = -\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i y_i \alpha_j y_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) + \sum_{i=1}^{n} \alpha_i. \tag{4.2.23}$$

To find the optimal hyperplane, the dual Lagrangian $L_d(\alpha)$ has to be maximized with respect to nonnegative and (unlike before) smaller than or equal to C, α_i

$$C \geq \alpha_i \geq 0, \qquad i = 1, \dots, n \tag{4.2.24}$$

and under the constraint

$$\sum_{i=1}^{n} \alpha_i y_i = 0. \tag{4.2.25}$$

The final quadratic optimization problem is therefore the same as that for the separable case, with the only difference laying in the modified bounds of the Lagrange multipliers α_i. The penalty parameter C, which is now the upper bound on α_i, is determined by the user and the selection of a "proper" parameter is always done experimentally by using some cross-validation technique. The learning problem is expressed only in terms of unknown Lagrange multipliers α_i, and known inputs and outputs. Furthermore, optimization does not solely depend upon input, but it depends upon a scalar product of input vectors $\boldsymbol{x}_i$. Finally, expressions for both a decision function $d(\boldsymbol{x})$ and an indicator function $i_F = sign(d(\boldsymbol{x}))$ for a soft margin classifier are same as for linearly separable classes.

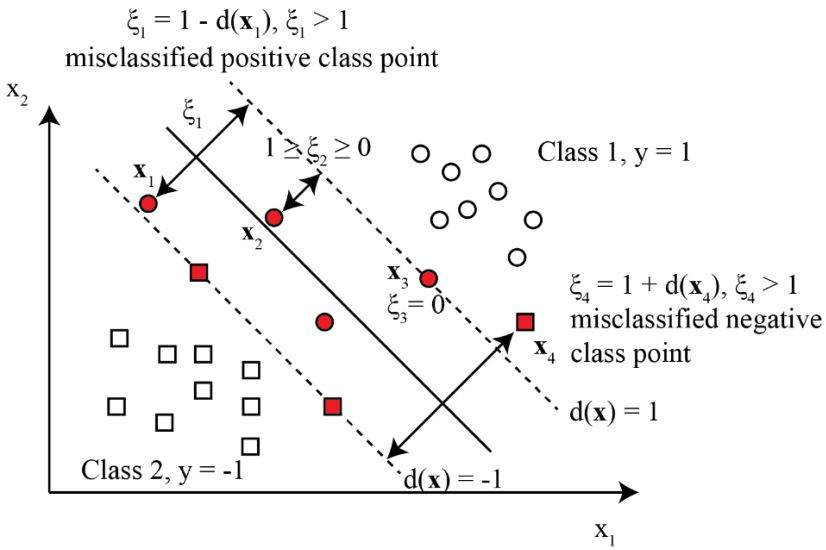

Fig. 4.2.3: visual representation of classification problem. The red objects are the support vectors. Two misclassifications are also highlighted.

Only three possible solutions for α_i (see *Fig. 4.2.3*):

- $\alpha_i = 0$, $\xi_i = 0$, data point $\boldsymbol{x}_i$ is correctly classified,
- $C > \alpha_i > 0$, → then, the two complementarity conditions must result in $y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1 + \xi_i = 0$, and $\xi_i = 0$. Thus, $y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) = 1$ and $\boldsymbol{x}_i$ is a support vector. The support vectors with $C > \alpha_i > 0$ are called free support vectors because they lie on the two margins,
- $\alpha_i = C$, → then, $y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1 + \xi_i = 0$, and $\xi_i \geq 0$, and $\boldsymbol{x}_i$ is a support vector. The support vectors with $\alpha_i = C$ are called bounded support vectors. They lie on the "wrong" side of the margin. For $1 > \xi_i \geq 0$, $\boldsymbol{x}_i$ is still correctly classified, and if $\xi_i \geq 1$, $\boldsymbol{x}_i$ is misclassified.

For L2 SVM the second term in the cost function (*Eq. 4.2.19*) is quadratic, i.e.,

$$C \sum_{i=1}^{n} {\xi_i}^2 \tag{4.2.26}$$

and this leads to changes in a dual optimization problem,

$$L_d(\alpha) = -\frac{1}{2} \sum_{i,j=1}^{n} \alpha_i y_i \alpha_j y_j \left(\boldsymbol{x}_i \cdot \boldsymbol{x}_j + \frac{\delta_{ij}}{C} \right) + \sum_{i=1}^{n} \alpha_i \tag{4.2.27}$$

subject to

$$\alpha_i \geq 0, i = 1, \ldots, n \tag{4.2.28}$$

and

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{4.2.29}$$

where, δ_{ij} = 1 for i = j, and it is zero otherwise.

4.2.3 The nonlinear SVM

The linear classifiers presented above are limited. In most cases, the categories are not only overlapped, but the separation functions are nonlinear hypersurfaces. The approach presented above can be extended to create nonlinear decision boundaries. The motivation for such an extension is that an SV machine that can create a nonlinear decision hypersurface will be able to classify nonlinearly separable data (example in *Fig. 4.2.4*).[174] This will be achieved by considering a linear classifier in the so-called feature space.[175]

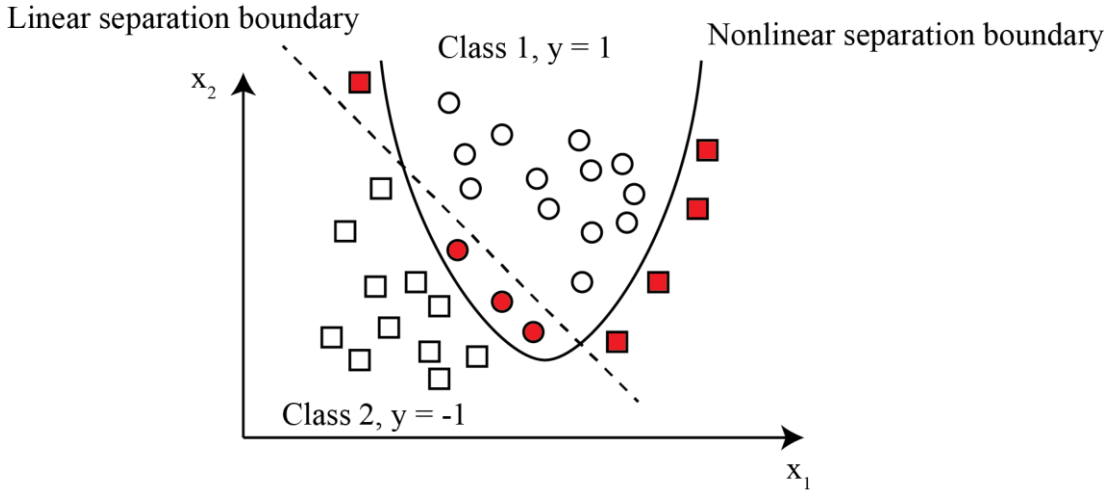

Fig. 4.2.4: visual representation of the nonlinear classification problem. A straight line is unable to correctly classify the squares and the circles, leading to several misclassifications (red objects). Instead, a parabolic line is able to distinguish correctly the two classes.

Usually, for m-dimensional input sets, an SVM will create a nonlinear separating hypersurface. The basic idea in designing nonlinear SVM is to map input vectors $\boldsymbol{x} \in \mathbb{R}^m$ into vectors $\boldsymbol{\Phi}(\boldsymbol{x})$ of a higher dimensional feature space F (where $\boldsymbol{\Phi}$ represents mapping: $\mathbb{R}^m \rightarrow \mathbb{R}^f$), and to solve a linear classification problem in this feature space

$$\boldsymbol{x} \in \mathbb{R}^m \rightarrow \boldsymbol{\Phi}(\boldsymbol{x}) = [\Phi_1(\boldsymbol{x}), \Phi_2(\boldsymbol{x}), \dots, \Phi_n(\boldsymbol{x})] \in \mathbb{R}^f \tag{4.2.30}$$

A mapping $\boldsymbol{\Phi}(\boldsymbol{x})$ is chosen in advance. i.e., it is a fixed function. An input space (in the present case, the $\boldsymbol{x}$-space) is spanned by components x_i of an input vector $\boldsymbol{x}$ and a feature space F (in the $\boldsymbol{\Phi}$-space) is spanned by components $\Phi_i(\boldsymbol{x})$ of a vector $\boldsymbol{\Phi}(\boldsymbol{x})$. The aforementioned mapping is useful for the algorithm to find, in a $\boldsymbol{\Phi}$-space, a linear separation images of $\boldsymbol{x}$ by applying the linear SVM formulation presented above. The solution for an indicator function

$$i_F(x) = sign(\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{x}) + b) = sign(\sum_{i=1}^{n} y_i \alpha_i \boldsymbol{\Phi}^T(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}) + b) \tag{4.2.31}$$

which is a linear classifier in a feature space, will create a nonlinear separating hypersurface in the original input space. $i_F(x)$ can therefore be rewritten as

$$
\begin{aligned}
i_F(x) &= sign\left(\sum_{i=1}^{n} y_i \alpha_i \boldsymbol{\Phi}^T(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}) + b\right) \\
&= sign\left(\sum_{i=1}^{n} \alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b\right) \\
&= sign\left(\sum_{i=1}^{n} v_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b\right)
\end{aligned}
\tag{4.2.32}
$$

where v_i correspond to the output layer weights of the "SVM's network" and $k(\boldsymbol{x}_i, \boldsymbol{x})$ denotes the value of the kernel function.[176] Now v_i are scalar values composing the weight vector $\boldsymbol{v}$ which dimension is equal to the number of training data points, in contrast to the previous case, where $\boldsymbol{w}$ is the vector of the same dimension as the feature space vector $\boldsymbol{\Phi}(\boldsymbol{x})$. As stated before, the $(n - N_{SVs})$ components of $\boldsymbol{v}$ are null, and only N_{SVs} entries of $\boldsymbol{v}$ are nonzero elements.

Two problems can arise when mapping an input in the x-space to the higher dimensionality F-space:

- the choice of the vector $\boldsymbol{\Phi}(\boldsymbol{x})$ should be correct, resulting in a "rich" class of hypersurfaces,
- the scalar product $\boldsymbol{\Phi}^T(\boldsymbol{x})\boldsymbol{\Phi}(\boldsymbol{x})$ can be computationally expensive due to the extremely large number of features f (which represents the dimensionality of the feature space).

The second obstacle is related to a phenomenon called "curse of dimensionality". For example, to construct a decision surface corresponding to a polynomial of degree two in an n-D input space, a dimensionality of a feature space $f = \frac{n(n+3)}{2}$ is required. Performing a scalar product operation with vectors of such, or higher, dimensions, is not a cheap computational task. This explosion in dimensionality can be avoided by noticing that in the quadratic optimization problem, as well as in the final expression for a classifier, training data only appear in the form of scalar products $\boldsymbol{x}_i^T \cdot \boldsymbol{x}_j$. These products will be replaced by scalar products

$$
\boldsymbol{\Phi}^T(\boldsymbol{x})\boldsymbol{\Phi}(\boldsymbol{x}_i) = [\Phi_1(\boldsymbol{x}), \Phi_2(\boldsymbol{x}), \dots, \Phi_n(\boldsymbol{x})]^T [\Phi_1(\boldsymbol{x}_i), \Phi_2(\boldsymbol{x}_i), \dots, \Phi_n(\boldsymbol{x}_i)] \tag{4.2.33}
$$

in a feature space F, and the latter can be and will be expressed by using the kernel function

$$
K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\Phi}^T(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}_j) \tag{4.2.34}
$$

which is a function in the input space.[177] Thus, the basic advantage in using kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is in avoiding performing a mapping $\boldsymbol{\Phi}(\boldsymbol{x})$. Instead, the required scalar products in a feature space $\boldsymbol{\Phi}^T(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}_j)$ are calculated directly by computing kernels $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$ for given training data vectors in an input space. In this way, the extremely high dimensionality of a feature space F is avoided. In addition, as will be shown below, by applying kernels we do not even have to know what the actual mapping $\boldsymbol{\Phi}(\boldsymbol{x})$ is. Further details and a complete list of kernels are findable in the literature.[178]

4.2.4 Regression by support vector machine

In the regression procedure, the functional dependence of the output variable $y \in \mathbb{R}$ is estimated based on the m-dimensional input variable $\boldsymbol{x}$. Thus, the object of study are real valued functions with the mapping going from an input $\mathbb{R}^m$ space to an output space of dimensionality $\mathbb{R}^1$. Same as in the case of classification, this will be achieved by training the SVM model on a training data set. Initially developed for solving classification problems, SV techniques can be successfully applied in regression, i.e., for a functional approximation problem. The general regression learning problem is set as follows: the learning machine is given n training data from which it attempts to learn the input-output relationship (dependence, mapping or function) $f(\boldsymbol{x})$. A training data set $D = \{[\boldsymbol{x}(i), y(i)] \in \mathbb{R}^n \times \mathbb{R}$, being $i = 1, \ldots, n\}$ consists of n pairs $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)$, where the inputs $\boldsymbol{x}$ are n-dimensional vectors $\boldsymbol{x} \in \mathbb{R}^n$ and system responses $y \in \mathbb{R}$, are continuous values. As for the previous case, a regression hyperplane $f(\boldsymbol{x}, \boldsymbol{w})$ can be expressed as

$$f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w} \cdot \boldsymbol{x} + b \,. \tag{4.2.35}$$

In the case of SVM's regression, the error of approximation is exploited in spite of the margin. The Vapnik's linear loss (error) function with ε-insensitivity may be exploited, which is defined as

$$E(\boldsymbol{x}, y, f) = |y - f(\boldsymbol{x}, \boldsymbol{w})|_\varepsilon = \begin{cases} 0, & \text{if } |y - f(\boldsymbol{x}, \boldsymbol{w})| \le \varepsilon \\ |y - f(\boldsymbol{x}, \boldsymbol{w})| - \varepsilon, & \text{otherwise} \end{cases} \tag{4.2.36a}$$

or as

$$e(\boldsymbol{x}, y, f) = \max(0, |y - f(\boldsymbol{x}, \boldsymbol{w})| - \varepsilon) \tag{4.2.36b}$$

Thus, loss = 0 is obtained if the difference between the predicted $f(\boldsymbol{x}_i, \boldsymbol{w})$ and the measured value y_i is less than ε. Vapnik's ε-insensitivity loss function defines an ε tube, represented in *Fig. 4.2.5*. If the predicted value is within the tube, the loss is zero. For all other points outside the tube, the loss equals the magnitude of the difference between the predicted value and the radius ε of the tube.

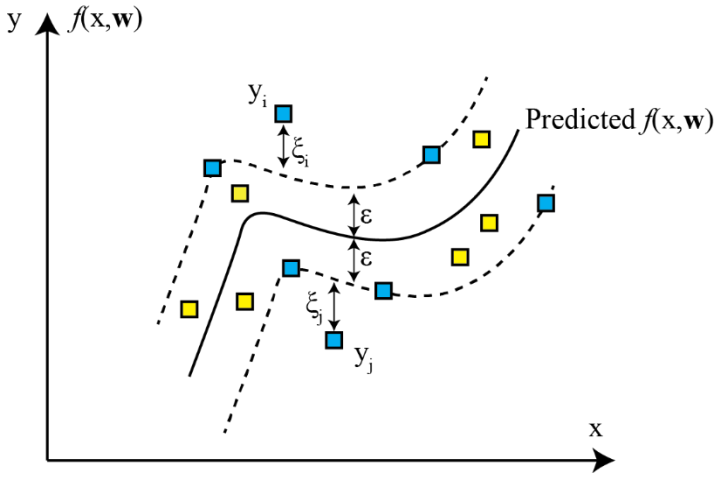

Fig. 4.2.5: visual representation of the ε tube. Cyan squares on the boundary of the tube are support vectors, yellow ones are not.

The two classic error functions are: a square error, i.e., L_2 norm $(y-f)^2$, as well as an absolute error, i.e., L_1 norm, least modulus $|y-f|$. Note that for ε = 0, Vapnik's loss function equals a least modulus function. Typical graph of a (nonlinear) regression problem as well as all relevant mathematical objects required in learning unknown coefficients w_i are shown in *Fig. 4.2.6*. As for the classification problem, SVM regression's algorithm will be formulated for the linear case first and then, applying mapping to a feature space, and using the kernel trick, they will be extended to nonlinear regression hypersurface. Here, for the regression, the empirical error term R_{emp} is measured using by Vapnik's ε-insensitivity loss function given by *Eq. 4.2.36* and shown in *Fig. 4.2.6c*.

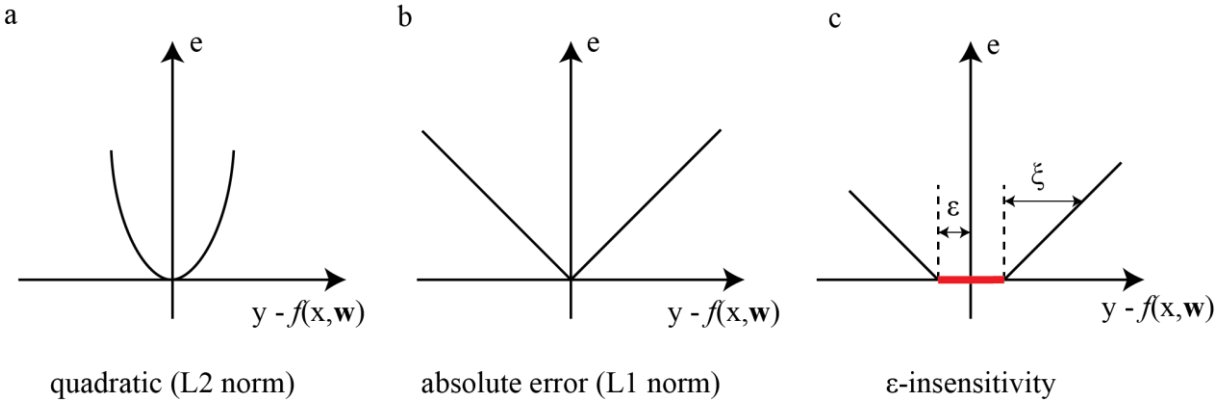

Fig. 4.2.6: different error typologies.

The empirical risk is given as

$$R_{emp}^{\varepsilon}(\boldsymbol{w},b)=\frac{1}{n}\sum_{i=1}^{n}|y_i-\boldsymbol{w}\cdot\boldsymbol{x}_{\mathrm{i}}-b|_{\varepsilon} \tag{4.2.37}$$

Fig. 4.2.7 shows two linear approximating functions as dashed lines inside an ε-tube having the same empirical risk R_{emp}^{ε} as the regression function $f(\boldsymbol{x},\boldsymbol{w})$ on the training data.

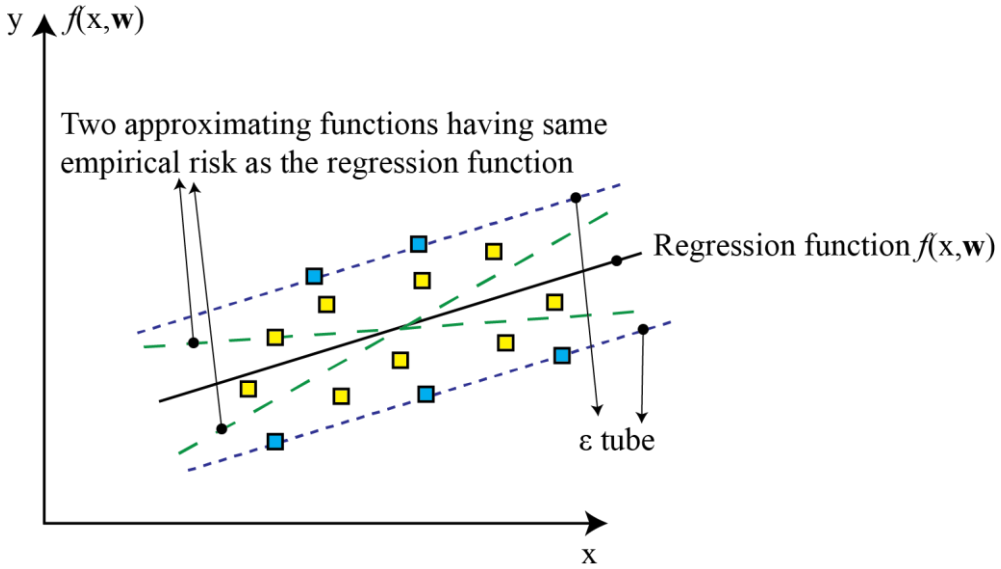

Fig. 4.2.7. two linear approximations inside an ε tube (green dashed lines) have the same empirical risk R_{emp}^{ε} on the training data as the regression function.

The goal is to simultaneously minimize both the empirical risk R_{emp}^{ε} and $\|\boldsymbol{w}\|^2$. Thus, as in the classification case, a linear regression hyperplane $f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{x} + b$ is built by minimizing

$$R = \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=l}^{n} |y_i - f(\boldsymbol{x}_i, \boldsymbol{w})|_{\varepsilon} \ . \tag{4.2.38}$$

Note that Vapnik's ε-insensitivity loss function instead of a squared error is used now. It is possible to distinguish the training data as

$$|y - f(\boldsymbol{x}, \boldsymbol{w})| - \varepsilon = \xi \tag{4.2.39}$$

for data "above" an ε-tube or

$$|y - f(\boldsymbol{x}, \boldsymbol{w})| - \varepsilon = \xi^* \tag{4.2.40}$$

for data "below" an ε-tube. Thus, minimizing the risk R is the same process as the minimization of

$$R_{\mathrm{w},\xi,\xi^*} = \left[\frac{1}{2} \|\boldsymbol{w}\|^2 + C \left(\sum_{i=l}^{l} \xi_i + \sum_{i=1}^{l} \xi_i^* \right) \right], \tag{4.2.41}$$

under constraints

$$y_i - \boldsymbol{w} \cdot \boldsymbol{x}_i - b \leq \varepsilon + \xi_i \,, \quad i = 1, \ldots, n \tag{4.2.42a}$$

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b - y_i \leq \varepsilon + \xi_i^* \,, \quad i = 1, .., n \tag{4.2.42b}$$

$$\xi_i \geq 0, \quad \xi_i^* \geq 0, \quad i = 1, \ldots, n \tag{4.2.42c}$$

where ξ_i and ξ_i^* are positive slack variables; α_i and α_i^*, will be nonzero values for training points "above" and "below" an ε-tube respectively. Since no training data can be on both sides of the tube, either α_i or α_i^* will be nonzero. For data points inside the tube, both multipliers will be equal to zero. Thus $\alpha_i \alpha_i^* = 0$.

Note also that the constant C is a parameter that is chosen by the user. An increase in C penalizes larger errors because it forces ξ_i and ξ_i^* to be small. This leads to an approximation error decrease which is achieved only by increasing the weight vector norm $\|\boldsymbol{w}\|$. Another design parameter is represented by the ε value that defines the size of an ε-tube. The choice of ε value is easier than the choice of C and it is given as either maximally allowed or some given or desired percentage of the output values y_i (say, ε = 0.1 of the mean value of $\boldsymbol{y}$).

The constrained optimization problem is solved by forming a primal variables Lagrangian as follows,

$$
\begin{aligned}
L_p(\boldsymbol{w}, b, \xi_i, \xi_i^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*) \\
&= \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\beta_i^*\xi_i^* + \beta_i\xi_i) \\
&- \sum_{i=1}^{n}\alpha_i[\boldsymbol{w}\cdot\boldsymbol{x}_i + b - y_i + \xi + \xi_i^*] \\
&- \sum_{i=1}^{n}\alpha_i^*\,[y_i - \boldsymbol{w}\cdot\boldsymbol{x}_i, -b + \xi + \xi_i^*]
\end{aligned}
\tag{4.2.43}
$$

The primal variables Lagrangian $L_p(\boldsymbol{w}, b, \xi_i, \xi_i^*, \alpha_i\alpha_i^*, \beta_i, \beta_i^*)$ expressed above must be minimized with respect to primal variables $\boldsymbol{w}, b, \xi i$ and ξ_i^* and maximized with respect to Lagrange multipliers α, α_i^*, β and β_i^* having the saddle point at the solution $(\boldsymbol{w}_0, b_0,\ \xi_{i0}, \xi_{i0}^*)$ to the original problem. At the optimal solution the partial derivatives of L_p in respect to primal variables vanish (because of the nature of saddle point):

$$
\frac{\partial L_p(\boldsymbol{w}_0, b_0, \xi_{i0}, \xi_{i0}^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*)}{\partial \boldsymbol{\omega}} = \boldsymbol{\omega}_0 - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\boldsymbol{x}_i = 0 \tag{4.2.44a}
$$

$$
\frac{\partial L_p(\boldsymbol{w}_0, b_0, \xi_{i0}, \xi_{i0}^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*)}{\partial b} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \tag{4.2.44b}
$$

$$
\frac{\partial L_p(\boldsymbol{w}_0, b_0, \xi_{i0}, \xi_{i0}^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*)}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \tag{4.2.44c}
$$

$$
\frac{\partial L_p(\boldsymbol{w}_0, b_0, \xi_{i0}, \xi_{i0}^*, \alpha_i, \alpha_i^*, \beta_i, \beta_i^*)}{\partial \xi_i^*} = C - \alpha_i^* - \beta_i^* = 0. \tag{4.2.44d}
$$

Substituting the KKT into the L_p, the problem shifts to the maximization of a dual variables Lagrangian $L_d(\alpha_i, \alpha_i^*)$

$$
\begin{aligned}
L_d(\alpha_i, \alpha_i^*) &= -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\big(\boldsymbol{x}_i \cdot \boldsymbol{x}_j\big) \\
&\quad - \epsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)y_i \\
&= -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\big(\boldsymbol{x}_i \cdot \boldsymbol{x}_j\big) - \sum_{i=1}^{n}(\varepsilon - y_{i_i})\alpha_i + \sum_{i=1}^{n}(\varepsilon + y_{i_i})\alpha_i{}^*
\end{aligned}
\tag{4.2.45}
$$

subjects to constraints

$$
\sum_{i=1}^{n}\alpha_i^* = \sum_{i=1}^{n}\alpha_i \tag{4.2.46a}
$$

$$
0 \le \alpha_i \le C, \qquad i = 1, \dots, n \tag{4.2.46b}
$$

$$
0 \le \alpha_i{}^* \le C, \qquad i = 1, \dots, n \tag{4.2.46c}
$$

Note that $L_d(\alpha_i, \alpha_i^*)$ is expressed in terms of Lagrange multipliers α_i and α_i^* only. After the learning stage, n Lagrange multiplier pairs (α_i, α_i^*) are obtained and the number of nonzero parameters α_i or α_i^* is equal to the number of SVs (independent on the dimensionality of the input space). Given that at least one element of each pair (α_i, α_i^*), i = 1, … n, is zero, the product of α_i and α_i^* is always zero, i.e., $\alpha_i \alpha_i^* = 0$. At the optimal solution the following KKT complementarity conditions must be fulfilled

$$\alpha_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b - y_i + \epsilon + \xi_i) = 0 \tag{4.2.47a}$$

$$\alpha_i^*(-\boldsymbol{w} \cdot \boldsymbol{x}_i - b + y_i + \epsilon + \xi_i^*) = 0 \tag{4.2.47b}$$

$$\beta_i \xi_i = (C - \alpha_i)\xi_i = 0 \tag{4.2.47c}$$

$$\beta_i^* \xi_i^* = (C - \alpha_i^*)\xi_i^* = 0 \tag{4.2.47d}$$

From the above equations it is possible to conclude that for $0 < \alpha_i < C$, $\xi_i = 0$ holds. Similarly, it follows that for $0 < \alpha_i^* < C$, $\xi_i^* = 0$ and, for $0 < \alpha_i \alpha_i^* < C$, follows,

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b - y_i + \epsilon = 0 \tag{4.2.48a}$$

$$-\boldsymbol{w} \cdot \boldsymbol{x}_i - b + y_i + \epsilon = 0 \tag{4.2.48b}$$

Thus, for all the data points fulfilling $y - f(x) = \epsilon$, α_i must be in the interval $0 < \alpha_i < C$, and for the ones satisfying $y - f(x) = -\epsilon$, the dual variable α_i^* take on values $0 < \alpha_i^* < C$. These data points are called the free (or unbounded) support vectors. They role is to help computing the bias term b

$$b = y_i - \boldsymbol{w} \cdot \boldsymbol{x}_i - \varepsilon, for\ 0 < \alpha_i < C \tag{4.2.49a}$$

$$b = y_i - \boldsymbol{w} \cdot \boldsymbol{x}_i + \varepsilon, for\ 0 < \alpha_i{}^* < C \tag{4.2.49b}$$

The best way to compute the bias b is taking the average over all the free support vectors, because the calculation requires high precision. For all the data points outside the ε-tube, $\xi_i > 0$ and $\xi_i^* > 0$, $\alpha_i = C$ for the points above and $\alpha_i^* = C$ for the points below the tube. These data are the so-called bounded support vectors. It is important to note that all the training data points within the tube are neither support vectors nor they construct the decision function $f(x)$.

The optimal weight vector of the regression hyperplane can be written as

$$\boldsymbol{w}_0 = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\boldsymbol{x}_i \tag{4.2.50}$$

and the best regression hyperplane obtained is given by

$$f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{w}_0 \cdot \boldsymbol{x} + b = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\boldsymbol{x}_i \cdot \boldsymbol{x} + b \tag{4.2.51}$$

The nonlinear task is a more challenging problem. The generalization to nonlinear regression (NR) can be performed by carrying the mapping to the feature space, or by using kernel functions. Therefore, the NR function in an input space will be devised considering a

linear regression hyperplane in the feature space. Recalling the explanation given for the classification problem, the input vectors $\boldsymbol{x} \in \mathbb{R}^m$ are mapped into vectors $\boldsymbol{\Phi}(\boldsymbol{x})$ of a higher dimensional feature space F (where $\boldsymbol{\Phi}$ represents mapping: $\mathbb{R}^m \rightarrow \mathbb{R}^f$) then the linear regression problem in this feature space can be solved exploiting the algorithms expressed above. In fact, by performing the described mapping in a $\boldsymbol{\Phi}$-space, the hope is that the learning algorithm will be able to perform a linear regression hyperplane by applying the linear regression SVM formulation, expecting this approach to lead to solving a quadratic optimization problem with inequality constraints in the feature space. The linear (in space F) solution for the regression hyperplane $f = \boldsymbol{w} \cdot \boldsymbol{\Phi} + b$, will then generate a nonlinear regressing hypersurface in the original input x space. An optimal weighting vector of the kernel's expansion can be found as

$$\boldsymbol{v}_0 = \boldsymbol{\alpha} - \boldsymbol{\alpha}^*. \tag{4.2.52}$$

With respect to the previous (classification) case, in a non-linear SVMs' regression, the optimal weight vector $\boldsymbol{w}_0$ could often be of infinite dimension. This is the reason why the best NR function is created by using the weighting vector $\boldsymbol{v}_0$ and the kernel (Gramian matrix **G** where $G_{ij} = \boldsymbol{\Phi}^T(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}_j)$) as follows,

$$f(\boldsymbol{x}, \boldsymbol{w}) = \boldsymbol{G}\boldsymbol{v}_\mathbf{0} + b \tag{4.2.53}$$

therefore

$$\boldsymbol{G}_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{4.2.54}$$

In the same way it is possible to obtain:

$$w_0 = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\boldsymbol{\Phi}(\boldsymbol{x}_i) \tag{4.2.55}$$

and

$$f(x, w) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\boldsymbol{\Phi}^T(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}) + b = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)K(\boldsymbol{x}_i, \boldsymbol{x}) + b. \tag{4.2.56}$$

If the bias term b is explicitly used as in *Eq. 4.2.53* then, for an NL SVMs' regression, it can be calculated from the upper SVs as,

$$b = y_i - \sum_{j=1}^{N\,freeupper\,SV} \left(\alpha_j - \alpha_j^*\right)K(\boldsymbol{x}_i, \boldsymbol{x}_j) - \epsilon \tag{4.2.57b}$$

or from the lower ones as,

$$b = y_i - \sum_{j=1}^{N\,freelower\,SV} \left(\alpha_j - \alpha_j^*\right)K(\boldsymbol{x}_i, \boldsymbol{x}_j) + \epsilon \tag{4.2.57b}$$

Note that $\alpha_j^* = 0$ in *Eq. 4.2.57a* and so is $\alpha_j = 0$ in *Eq. 4.2.57b*. Similarly to the linear case, it is better to calculate the bias term b averaging over all the free support vector data points. All three parameters' sets involving the insensitivity zone ε, the penalty parameter C shape and the kernel function should be selected by the user. The parameter selection is usually

a though procedure. The most popular method to facilitate the task is the cross-validation, which exploits at the beginning a trial and error approach to find at least one of the three parameters described.

4.2.5 Case study

The SVM algorithms have been exploited for the commissioning of the prototype machine to detect defects and classify reflectance properties of reflectors. For this case, non-linear classifiers have been used due to the nature of the experimental data. In fact, the input data are not classifiable with a linear dependence and non-linear techniques must be used. In particular, the Gaussian Radial Basis function (RBF) has been chosen as kernel which can be expressed as

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \tag{4.2.58}$$

or

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = e^{-\gamma \|x_i - x_j\|^2}. \tag{4.2.58}$$

This kernel is of the general type where $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ depends on $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ only through the norm

$$K(\boldsymbol{x_i}, \boldsymbol{x_j}) = k\left(\left\|\boldsymbol{x_i} - \boldsymbol{x_j}\right\|^2\right). \tag{4.2.59}$$

It nonlinearly maps samples into a higher dimensional space with the possibility to handle the case where relation between class labels and attributes is nonlinear.[179] It is based on a supervised learning methodology; in fact, it must be given to the algorithm some input-output pairs as examples in order to be trained. The RDF kernel is fast, does not require high computing performances and it supplies reliable results.[180,181] The advantages of using the RDF kernel, with respect to polynomial ones are multiple: it has less hyperparameters, which influence the complexity of the described model, and its kernel is limited to values in the range $0 < K(x_i, x_j) \leq 1$ in contrast to polynomial ones which may go to zero or to infinite. However, there are some situations where the RBF kernel is not suitable. Particularly, when the number of features is very large, one may just use the linear kernel and try to find the correct linear approximation. Two parameters are present in the RBF kernel: C and γ (γ explicitly expressed in the kernel formula while C is the well-known penalty parameter). The values of the two parameters are not known beforehand; consequently, some kind of model selection (parameter search) must be performed to correctly analyze the data. The goal is to identify good (C, γ) so that the classifier can accurately predict unknown data (the so-called testing data). To test the algorithms based on the kernel described above, two sets of samples have been prepared: the training set, used to teach the prototype machine how to recognize compliant and non-compliant components, and the validation set, exploited to check if the detection rate is sufficient for the detection.[182]

5. Automated characterization of outer lenses

5.1 Introduction

In the past, monitoring the quality of products was a human task. A dedicated operator had to take the decision whether a part was compliant or not, evaluating it in every single detail. Thus, several cofactors could have influenced the judgment: for example, poor ambient lighting and environmental conditions could have led to a quality check failure, with the presence of both false positive and false negative results. In addition, depending on the typology and dimensions of the component, the analysis could have taken a considerable amount of time, thus slowing or even limiting as a bottle neck the production chain.[183] An automated testing process for components is the key to improve the situation. Since computer vision systems can provide a fast and reliable analysis, in the present times automated inspection is exploited in multiple and various branches.[184,185] In the specific case of the automotive industry, in order to fulfill the market requests asking for progressively novel and more performing components, the need of an automated analysis system is strongly rising to maintain high productive standards and to reduce the inspection time.[186] The automotive lighting industry is not an exception in this constant upgrading progression: new technologies are exploited, and novel inspection methods are required to monitor different steps of the production chain. Spots of different color, cracks, incompleteness, excess and/or lack of material are just some examples of defects affecting the components of a rear lamp, which lead to discard the component or, even worse, the entire lamp. In this chapter, the attention is focused on the detection of structural and aesthetical defects that may arise during the production of the Mercedes E-class outer lens.

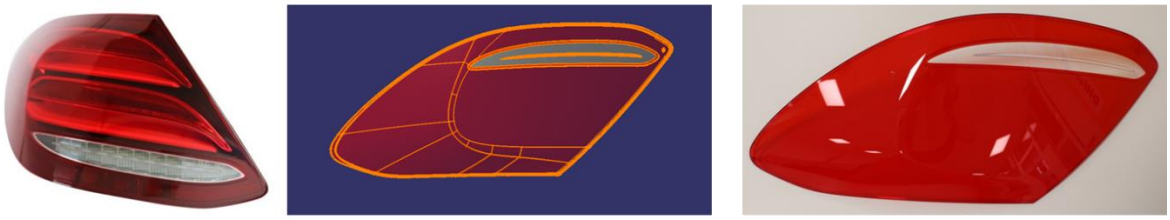

Fig. 5.1.1: left panel, example of rear lamp. Central panel, CAD model of the outer lens scanned with the prototype machine. Right panel, picture of the real outer lens.

The outer lens is the most external part of an automotive lamp and it must be qualitatively perfect, without flaws (examples representing the complete lamp, the CAD model and the outer lens are reported in *Fig. 5.1.1*). It is manufactured via injection molding of PMMA granules and the desired shape is obtained by means of steel molds in which the liquid PMMA is flown. The process is extremely accurate, but in some cases it is possible to have some defects in the shape or in the aesthetical aspects of the lens. In the specific case, selected examples are:

- Black spot: it is a small, dark, opaque area, which is observable due to the high contrast with the PMMA on which the defect is formed.
- Vinyl disk defect: in the zone affected by this flaw a wave-like behavior denoted by high surface roughness is visible, reminding the surface of a vinyl disk.

- Incompleteness: this macro-defect arises from an incomplete filling of the mold by the liquid PMMA and it is observable in a partial lack of a part (or more) of the lens.
- Scratch: visible as a scrape of various size and depth on the surface of the lens, the nature of this defect is various, but it is mainly caused by physical contact of the sample with other sharp objects.
- Air bubble: it is a region where the filling of the fluid PMMA is not complete cause of a small quantity of air inside the mold, leading to the formation of a zone with different transparency with respect to the rest of the lens.
- Crack: macro-flaw visible as rupture or rift in the lens, due to problems in the injection of the liquid PMMA.

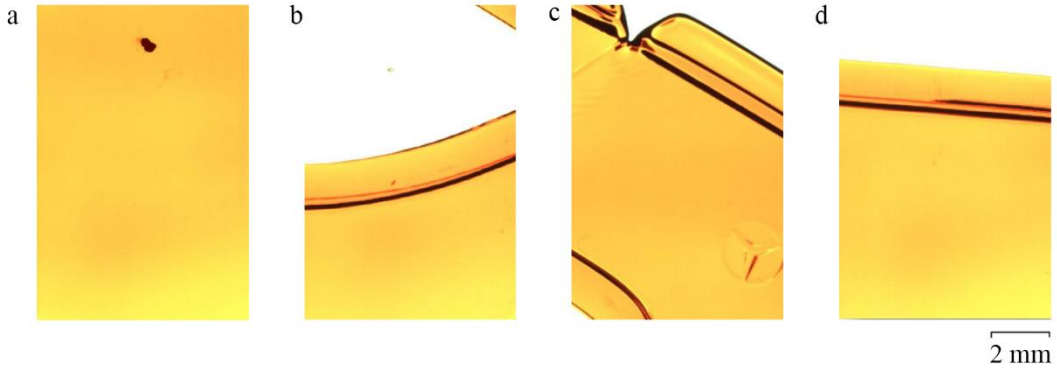

Fig. 5.1.2: images representing some of the defects described. Black spot (a), air bubble (b), incompleteness (c), and crack (d).

Fig. 5.1.2 reports some of the abovementioned defects. In the view of designing an automated inspection process, three steps are necessary to classify a defect univocally: data acquisition (scanning of the lens), image analysis (defects spotting), and defect evaluation (classification and attribution of a severity index). The full description of these three steps will be presented in the following. In addition, optical simulations performed to replicate all the possible defects' aspects and positions will be introduced. These simulations have been crucial to determine the feasibility of the different detection procedures and, therefore, fundamental for the entire prototype machine.

5.2 Layout and simulations

The pieces under examination are transparent and do not emit light by themselves therefore, in order to be analyzed, an external light source must be exploited. The lighting system is placed behind the lens, opposite to the camera and, taking advantage of the transparency of the outer lens, the signal can be acquired in transmission mode. The defects presented in the last section have different nature and they need different inspection methodologies to be analyzed. In order to be able to detect all the different types of flaws, a homemade illuminator has been designed. Optical simulations involving ray-tracing (both forward- and back-tracing) have been performed to determine the setup configuration yielding proper illumination conditions; moreover, the simulations were fundamental in order to understand how to detect all the possible defects arising on the lenses. These have been performed by means of a proprietary working environment allowing the ray-tracing calculation in CATIA, a multi-platform software suite for computer-aided design (CAD). To better detect the defects on the transparent component, the light source has been designed to achieve a near Lambertian distribution, so that its brightness is the same regardless the observer's angle of view. In *Fig. 5.2.1a,* the lateral view of the illuminator is schematized. The white foil contributes to the diffusion of light and to reduce the intensity peaks in the region close to each LED, while the aluminum foil on the back helps reflecting the light in the correct direction. A LED light strip composed of 40 LEDs is placed on the bottom of the illuminator and in *Fig. 5.2.1b,* the geometric distribution of the LEDs on the stripe is represented.

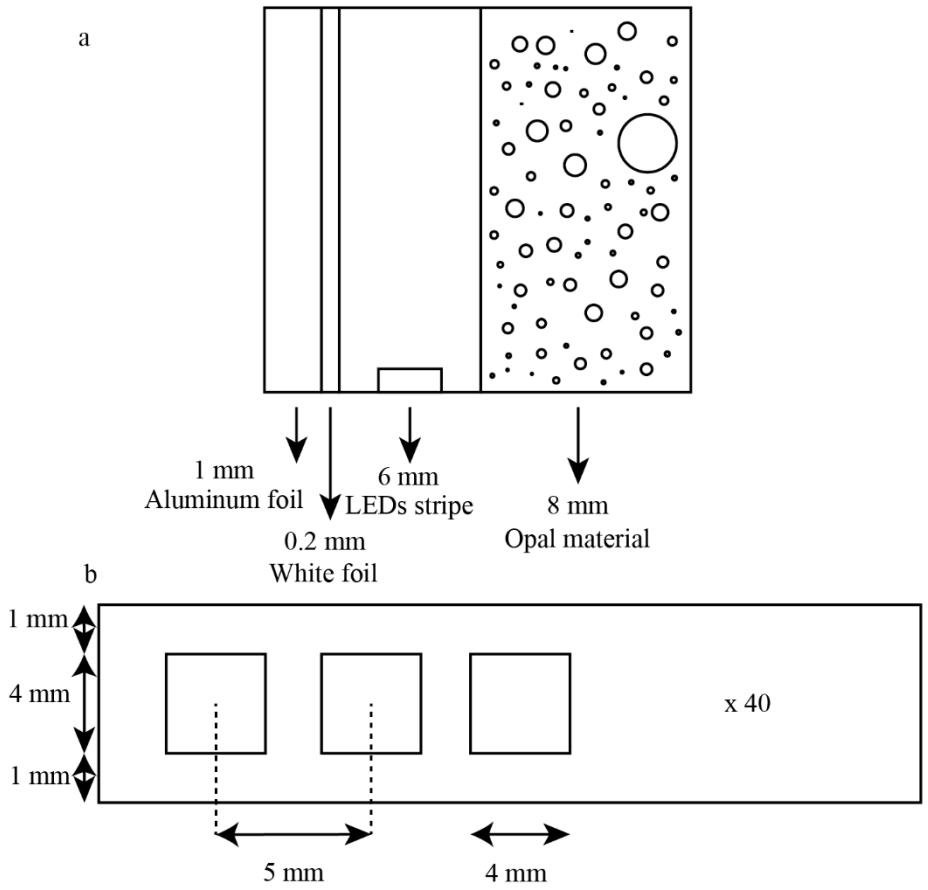

Fig. 5.2.1: (a) lateral view of the homemade illuminator (light source). (b), sizes of the LED stripe.

The emission spectrum of each LED is represented in *Fig. 5.2.2*. The white light is obtained combining the blue and yellow peaks evidenced in the emission spectrum.

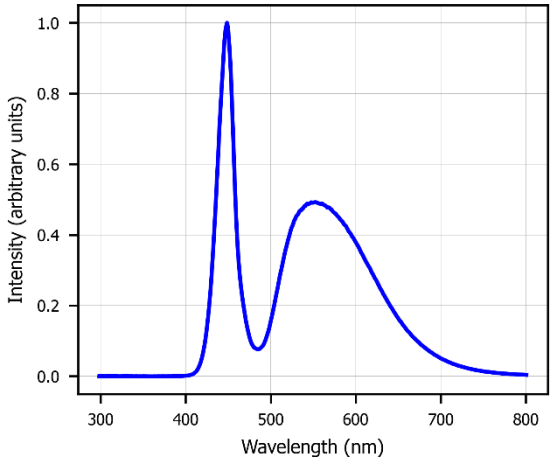

Fig. 5.2.2: emission intensity of the LEDs exploited for the illumination system. The white light is obtained by combining blue and yellow light, whose peaks are reported for a wavelength of 450 nm and 560 nm.

An opal diffuser, placed in the front side of the illuminator, helps spreading the light. An opal material is characterized by random micro-structures inside it having different refractive index with respect to the material in which they are immersed. These random "blisters" help scattering light randomly in all directions, as it is possible to see from the picture reported in *Fig.5.2.3*.

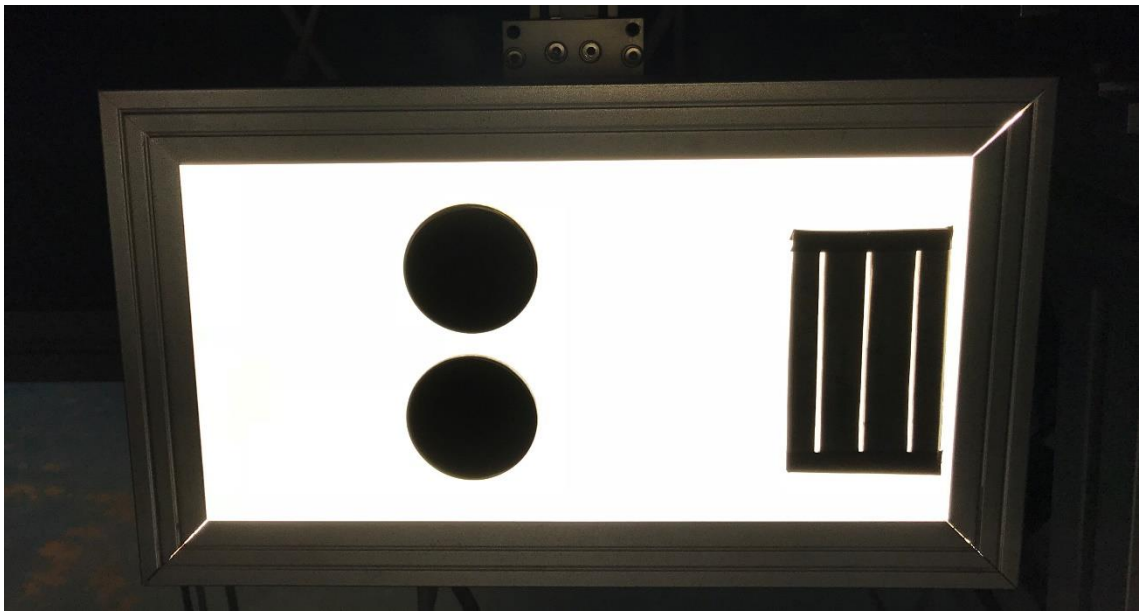

Figure 5.2.3: frontal picture of the illuminator.

The preliminary simulations performed to control if the illuminator can be compared to a Lambertian emitter are represented in *Fig. 5.2.4*.

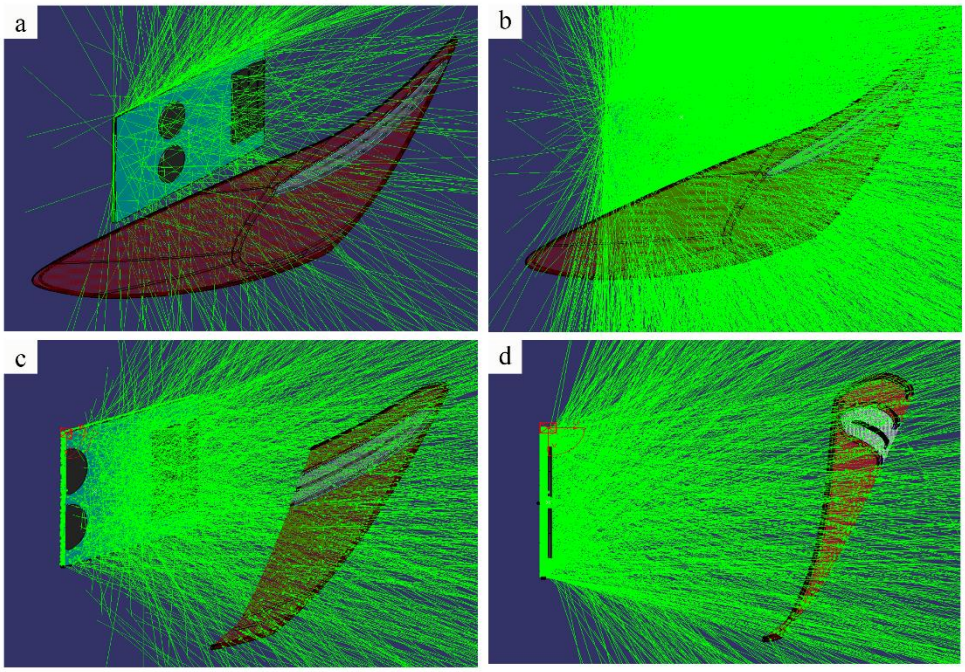

Fig. 5.2.4: ray-tracing simulation of the light source, with 2000 rays (a), 10000 rays (b) and 8000 rays from two different perspectives (c and d).

These simulations were performed exploiting the ray-tracing rendering technique selecting a different number of rays and a different perspective. The spatial uniformity of the emitted light testified the goodness of the setup. In the next pages, the simulations of defects appearance on the outer lens are reported. The simulations have been performed prior to the realization of the machine to test different geometrical configurations to better detect every kind of flaw. All defect-simulations have been performed in the back-tracing configuration, which is far more realistic. Moreover, even if back-tracing simulations take more time and require a more powerful GPU, it is the only method to simulate the passive functions of the lamp (i.e. components which do not emit light by themselves).

The actual inspection system is surrounded by dark panels, which are used to shield light coming from the surrounding environment. The prototype machine is illustrated in *Fig. 5.2.5*.

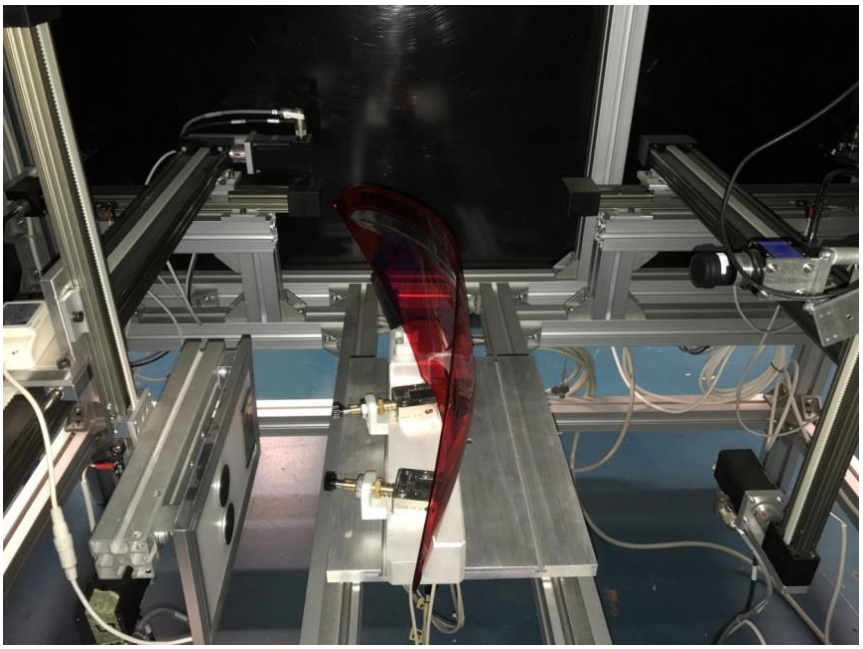

Fig. 5.2.5: representation of the prototype machine: the illuminator (on the left) lights up the outer lens (in the middle); the images are then acquired by a camera (on the right). It is possible to see the dark panels used to shield the system from the external light.

The detector is a GigE "mvBlueCOUGAR-X104C (2048x1088)" camera. The camera and the illuminator can move along three axes, while the lens can only rotate around the center of the footboard where it is placed, to maintain the orthogonality between its surface and the line connecting the illuminator and the camera. By moving the camera and the illuminator together, it is possible to collect a set of high-resolution images (which are always taken at some pre-determined fixed positions during the scans) and to analyze them without the severe distortion induced by the shape of the outer lens. The coordination of the source-camera movements is fundamental in pursuance of minimizing the distortion induced by the curved lens surface. A filter with three structures (two dark disks and one dark grating) is superimposed on the illuminator (scheme in *Fig. 5.2.6*). Thanks to this framework, it is possible to distinguish on the illuminator three different regions: the "transmission area" which has no structures on it, the "refraction area" with two dark disks on it, and the "pattern area" with the dark grating.

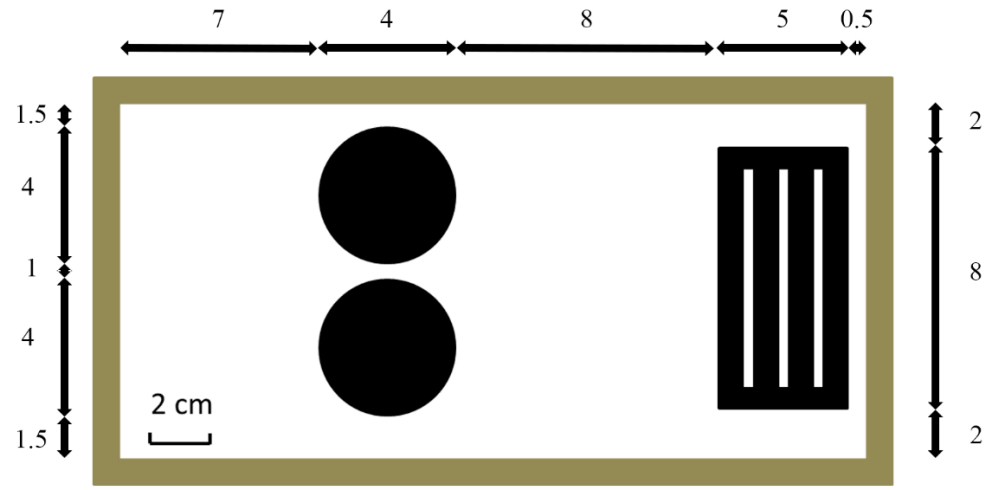

Fig. 5.2.6: schematic representation of the illuminator. Three zones are clearly visible: from left to right, the first is the region with the two black disks (refraction area), the second is the region without features on it (transmission area), and the last one is characterized by the presence of the pattern (pattern area). The numbers outside the golden frame represent the physical sizes of the illuminator and are expressed in cm.

These zones are specifically designed to detect different defects on the lens. The three areas are associated with different light scattering properties: in the “transmission area” the intensity of the light reaching every pixel of the camera is evaluated. This zone allows for a uniform illumination of the portion of the lens under examination in order to detect component’s regions with low transparency due to the presence of defects. The defects here analyzable could belong to one of two main families: presence of a dark spot that reflects or absorbs light and presence of discontinuity of the plastic material that inhibits light propagation in precise directions. Here in the specific black spots, air bubbles, cracks and incompleteness defects will be analyzed. In the “refraction zone”, due to the presence of the disks on the illuminator, the light comes with an incident angle that is not normal to the surface. Therefore, part of the refracted light can hit the camera if a defect on the surface of the lens modifies the path of the light beam. The blemish will be detected looking for high intensity signal in low intensity regions (dark background due to the presence of the disks in the background). In the “pattern zone” the interaction of light with the surface of the lens is similar to the one described for the “refraction zone”, but here the defects that are meant to be analyzed produce a quite regular and smooth distortion of the image, enhanced by the pattern superimposed to the illuminator to better recognize the defects in examination. Vinyl disk flaws are analyzed here.

5.3 Working principle

Different illumination conditions are used to better detect the defects under examination, therefore diverse algorithms are used for different blemishes. In fact, they are different in nature, physical aspect and properties. The implemented algorithms will be synthesized here, divided in categories.

5.3.1 Transmission region

In this region no filters on the backlight are present and multiple defects can be analyzed thanks to this setup: black spots, air bubbles, cracks, incompleteness.

Black spots

The presence of black spots (possible causes described in *Section 2.1.2*) inhibits the correct transmission of light in a region of the size of the defect itself. To perform the simulation of a black spot on the surface of an outer lens, the defect has been represented as a perfect circle made by a different material with respect to the whole compliant lens, showing dark appearance. In particular, the real part of the refractive index for the outer lens is 1.493, while the black spot has been treated like a perfect absorber, which means all the light rays impinging on it are absorbed. The output of the backtracing simulation is reported in *Fig. 5.3.1*.

Fig. 5.3.1: back-tracing simulation of the black spot detected in the transmission region. Rays used: 8.000.000.

The actual result obtained starting from the images, taken by the machine and reported in *Fig. 5.3.2*, is close to and in agreement with the simulations.

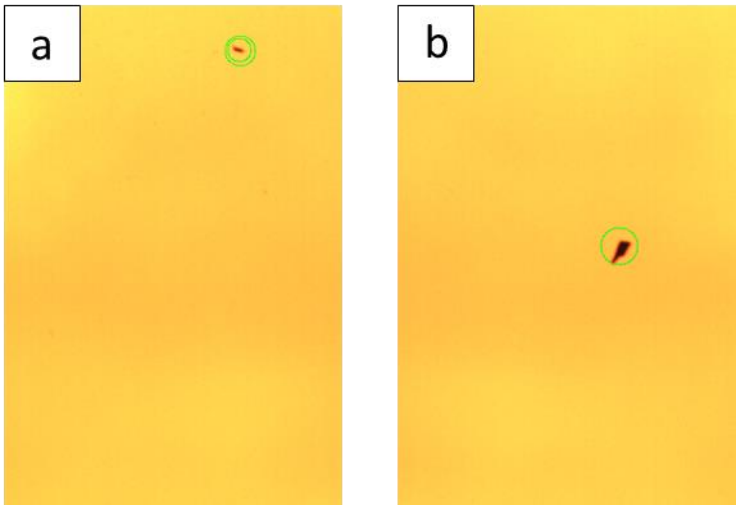

Fig. 5.3.2: two possible cases of black spot. In the detection phase the multi-scale kernels permit to identify approximate position and scale (highlighted by the green circles). In the first case (a), two scales have been associated to the same position, while in the second (b) just one scale is used.

Defining I as the image taken by the camera and s the scale of an image, the detection algorithm for the black spots processes the image by creating a set of masks $G(s, I)$ convolving the red channel (because the main color of the outer lens is red) of the image I with kernels $k(s)$ of different sizes s. An example of a kernel shape is represented in *Fig. 5.3.3*.

1	1	1	1	1
1	-1	-1	-1	1
1	-1	-1	-1	1
1	-1	-1	-1	1
1	1	1	1	1

Fig. 5.3.3: example of a 5x5 kernel $k(s)$ used to convolve the red channel of the image and obtain the desired masks $G(s, I)$.

Each cell of the kernel has a variable pixel size depending on the scale s, therefore $G(s, I)$ returns a list of possible black spots in the points (x, y), which correspond to local maxima in the masks $G(s, I)$. For each defect candidate the radius is estimated as proportional to the scale of the mask $G(s, I)$, with maximum intensity in the corresponding point (x, y). The output of this analysis is a list of candidates represented by point-coordinates (x, y) in which the results of the convolution process is maximum. Every black spot-nominate has an esteemed radius which is proportional to the scale s for which the response function has a maximum. The convolution is cost-demanding, therefore the integral image technique is used to speed up the computation.[187–190] In a second phase, important features are extracted from every candidate in order to describe its characteristics and to find out whether it is a defect or not. Remarkable quantities are, for example, the compactness and bipolarity indexes. The former indicates how much the defect approximates the shape of a circle, while the latter indicates if the defect is

composed by a single spot or if it is a sum of some minor flaws. Defining $P(P_0, P_1, P_2, \dots)$ the parameter-space generated by the features (for example, $P_0 =$ compactness index, etc.) only the candidates which exhibit a feature-vector which falls inside the subspace compatible with the characteristics expected for a black spot are indicated as effective defects. For this, specific threshold values have been calibrated.

Air bubble

As the name suggests, air bubble defects are caused by small volumes of air trapped inside the PMMA; due to the presence of multiple inner interfaces, light refraction occurs at the defects, thus affecting the transmitted beam path. Two different simulations have been performed to determine how to better detect the air bubble on the illuminator: one simulation was performed with the blemish placed in correspondence of black disk, the other with the flaw on the bare illuminator. The air bubble has been simulated as a hemispherical structure having refractive index equal to the air's one. The effect of the air bubble is visible both on the disk and on the bare illuminator, as represented in *Fig. 5.3.4* and enlarged in *Fig. 5.3.5*. Due to the double possibility, the detection algorithm has been implemented to work both in the transmission and in the refraction region, in order not to confuse this structure with a black spot and raise the reliability of the results. The effect of these scattering points is visible both in the transmission area - as the light normal to the surface is partially deviated away from the impinging direction and the image exhibits a darker circular region - and in in the refraction area, where bright circular areas are observed on a darker background.

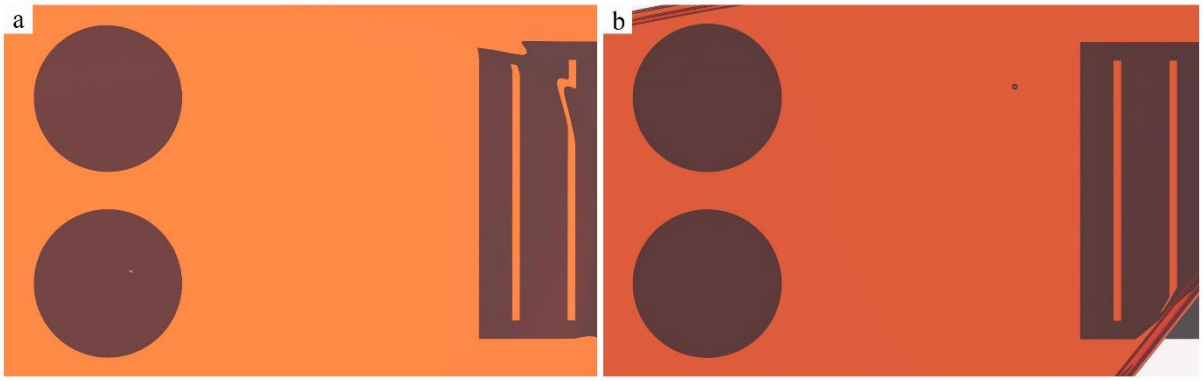

Fig. 5.3.4: full images of the back-tracing simulation of an air bubble detected on the black disk (a) and on the transmission region (b) (8.000.000 rays). It is of important to underline that the irregularities evidenced on the pattern structure in (a) are due to the natural curvature of the lens scanned and are not ascribable to a defect.

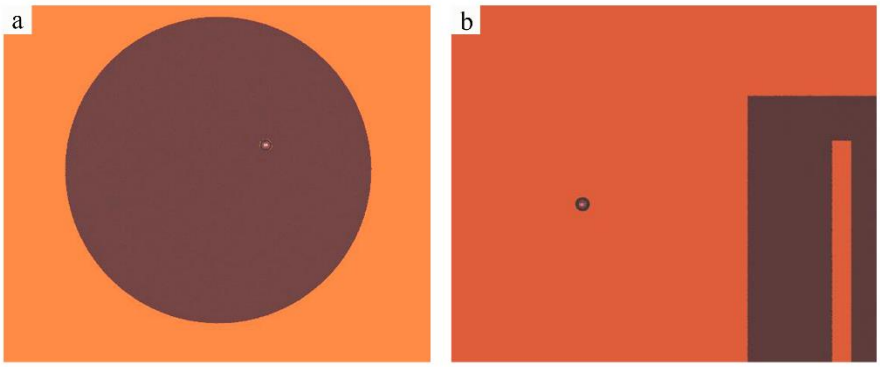

Fig. 5.3.5: enlargement of figure 5.3.4. Here it is better visible the influence of the air bubble on the dark disk background and on the bare illuminator.

A specific re-parameterization of the algorithm used to detect the black spots is exploited both in the transmission region, to detect small circular elements darker than the background, and in the refraction region, using the signal-inverted image instead of the original image to detect white circles on a dark background. Clearly, the same location on the lens surface can't be analyzed both in the reflection and in the transmission regions in the same frame. A precise mapping from the image coordinates (pixel) of a given frame to the lens surface physical coordinates (mm) can be used instead to detect correctly the air bubbles on the surface of the lens. This precise mapping is possible because the physical machine parameters are well known. The positions of air bubble candidates are then expressed in physical coordinates to locate them on the lens surface and only if a candidate detected in the refraction regions shares the same location with a candidate detected in the transmission region, it is confirmed as air bubble defect and validated (details in *Fig. 5.3.6*). The algorithm is powerful and very reliable due to the double check made in the two regions, thanks to which air bubbles are detected with approximately the 100% detection rate.

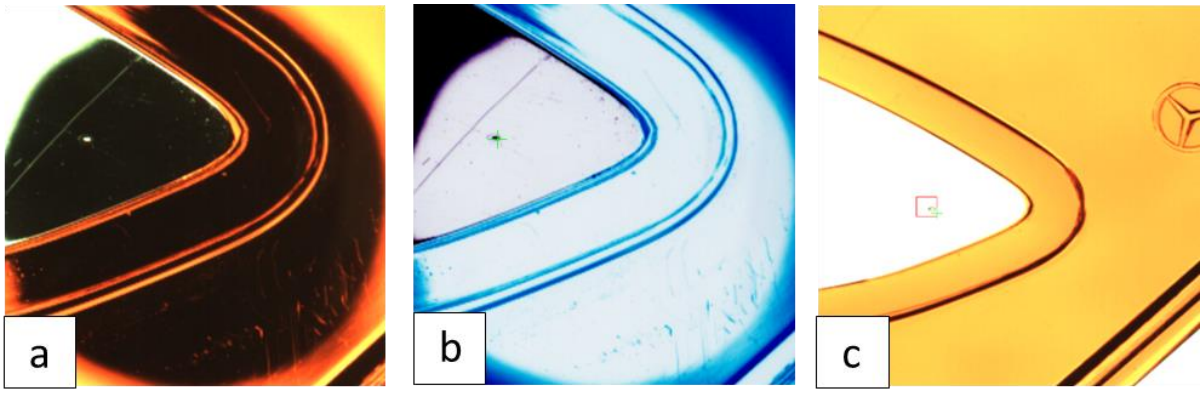

Fig. 5.3.6: (a), original frame from the refraction zone: the air bubble can be seen as a white spot on the dark background. In (b), inversion of the image has occurred, now the air bubble is a black spot on a white background and the algorithm is able to detect the position. In (c), the transformation from image coordinates to surface coordinates is performed and the proper frame is selected. Linking these results, it is possible to correctly spot the air bubble.

Cracks

It is possible to find cracks along the boundary between two different plastic castings or along the external profile of the lens. These flaws do not always appear at the center of the surface, though they are present at the borders of the lens. The detection algorithm looks therefore specifically for defects at the border of the surface or at the interfaces, where the cracks indeed take origin. The algorithm superimposes the i^{th} frame of a reference sample (i.e. without blemishes), $RF(i)$, to the i^{th} frame of the current acquisition, named $CF(i)$. Then it segments the connected areas that appear brighter than a threshold in the difference image $D(i) = CF(i) - RF(i)$, producing a series of contour candidates, as showed in the examples in *Fig. 5.3.7*.

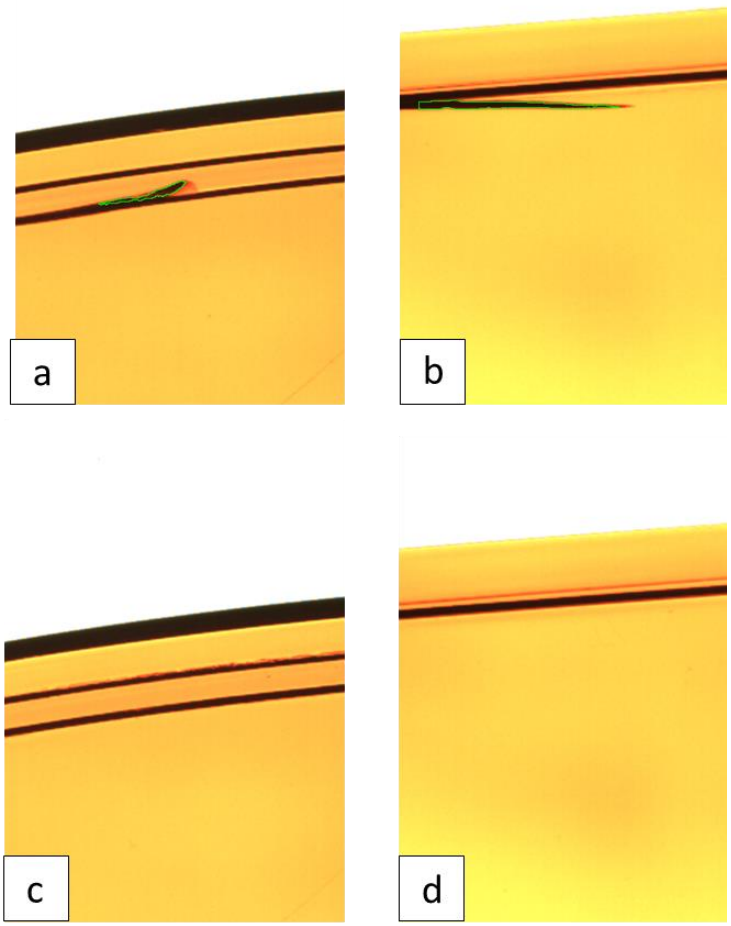

Fig. 5.3.7: the panels (a, b) on top show defects on a lens, while the panels on bottom (c, d) show the references, used as comparison. The mapping between image and reference must be very precise because usually the breaks are parallel to the boundaries and misclassifications can arise.

The superimposition and extraction process must take into account several aspects:

- Luminosity differences: the light of the source may vary in intensity or color due to source-related issues (different temperature of the source). These instabilities may induce different intensity levels in corresponding regions, yielding false negative

defects when identified by image subtraction algorithms. This issue is easily solved since cracks are generally strong absorbers, allowing for high threshold levels in the contour extraction process.

- Positioning zeroes: some mismatch in the positioning of the transparent may be possible. A feature of the transparent imaged in the frame $RF(i)$ can be seen in an offset position when imaged in the frame $CF(i)$. This latter issue can be corrected thanks to a rigid translation $Tr(i) = [dx(i); dy(i)]$ of the frame data, avoiding false positive detections due to the misalignment of the two frames. A registration algorithm maximizing the cross correlation between two frames is used to extrapolate automatically the translation $Tr(i)$ to correct for the rigid offset. The difference frame is then computed as $D(i) = CFT(i) - RF(i)$ where $CFT(i)$ is he frame $CF(i)$ translated by $Tr(i)$.
- Local deformation of the transparent: the extrapolated translation $Tr(i)$ between frame $RF(i)$ and $CF(i)$ is good enough to represent the sliding of the profile lines of the transparent, but is not always suitable for other features (internal junction lines, injection points, ...) far from them due to local deformations of the transparent caused by the printing process. As the breaks are searched only near the profile lines, this can't affect the candidate detection process, but can compromise the preliminary registration. Thus, a Gaussian mask is applied to the borders of the image as a local weighting in the cost function of the registry maximization problem. The farther the point from the border, the lower is its weight.

Candidates obtained in this segmentation phase are then sifted using a set of filters analogously to what described for the black spots. In this case, the features used to classify the images are mainly based on geometry: width, length, and elongation of the candidates.

The cracks haven't been simulated before the realization of the machine due to their extremely infrequent presence on the lenses.

Incompleteness

When the PMMA is injected into the mold it may happen that the material fails to reach the furthest zones with respect to the injection point. In our case study, the lack of material can arise only with red PMMA filling due to its wider area (with respect to the colorless one).

Fig. 5.3.8: back-tracing simulation of an incompleteness flaw on the outer lens (8.000.000 rays).

The simulation has been performed by simply removing part of the PMMA from the portion of the lens further from the injection point (used to fill the mold with liquid PMMA). *Fig. 5.3.8* demonstrates that the transmission region is the perfect zone to spot this flaw. Two types of incompleteness can be distinguished: color incompleteness, when the red casting cannot reach all the correct regions and cannot be correctly superimposed to the transparent casting, leaving uncolored regions, and figure incompleteness, when the material coming from both the two castings is missing, leading to an incorrect geometry of the lens. These two categories are clearly different, one concerning the shape and one concerning the perceived color of the lens, therefore two different algorithms have been implemented to detect both of them. For what concerns figure incompleteness defects, the algorithm is similar to the one used for the cracks, in fact it finds and extracts the contour that delimits the border of the lens in $CFT(i)$ and compares it with a reference. The maximum distance metric is used between the two borders (i.e. the one from the incomplete sample and the one from the reference sample) to expresses the presence of incompleteness or anomalies, as shown in *Fig. 5.3.9*. Regarding color incompleteness, the border is still present so can't be used as a discriminative feature as before, but it can be used to fine tune the reference-current image match. Once the two images are perfectly aligned, the areas covered by white pixels inside the border in the reference and current frame are expected to have similar extensions. Thus, the algorithm counts the number of pixels of the current frame in which the red channel intensity is similar to the intensities of the other channels (white area) giving back as result an area and compares this value with the expected extension precomputed from the reference frame. A color incompleteness is then reported if the calculated difference (expressed in pixel^2 or mm^2) is greater than a given absolute threshold. A threshold value of 9 mm^2 has been used in the final solution.

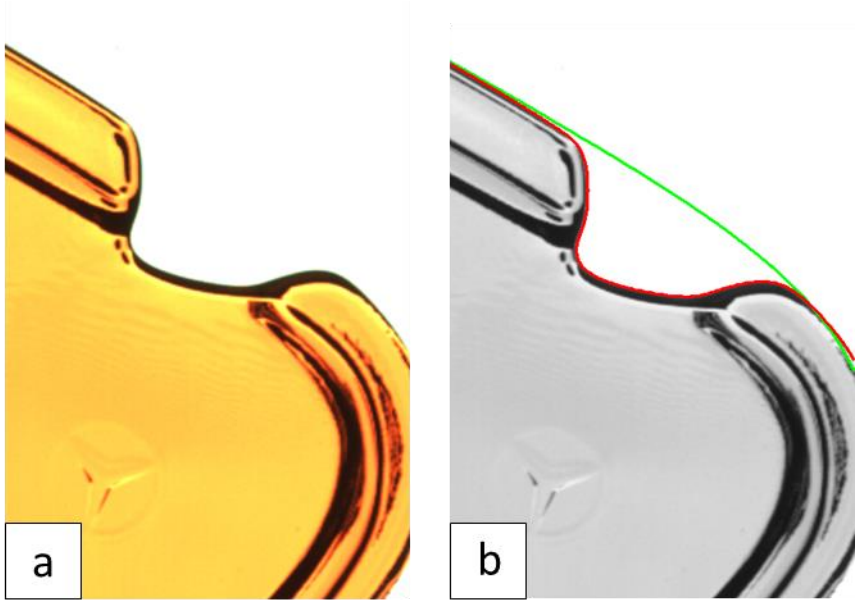

Fig. 5.3.9: in (a), an example of figure incompleteness is reported. Panel (b), highlights the differences between the border extracted from the reference (in green) and the border of the lens in analysis (in red). The maximum distance between the two borders is used as parameter to identify the defect.

5.3.2 Refraction region

This region is assigned to the identification and classification of scratches and other similar defects on the surface of the lens. In the refraction area, the black disks are exploited in order to screen the light propagating along the normal to the surface, thus yielding sensitivity to light beams diffracted and refracted at the defects. This is typical of the specific typology of defects that are intended to be analyzed here, scratches mainly.

Fig. 5.3.10: back-tracing simulation (8.000.000 rays) of a scratch on the outer lens. The scratch is places between the dark circle and the white background to evidence the importance of the dark structure in the scratch detection process.

The simulations performed to determine which configuration was the most correct to detect a scratch suggested to use a dark background. As it is possible to perceive from *Fig. 5.3.10*, the scratch is highly visible while it is superimposed to a dark circle, due to the high contrast. On the contrary, if the scratch is superimposed to the bare illuminator, it is extremely difficult to be spotted. In fact, the scratch modifies the smooth surface of the lens, changing the path of the light rays crossing the lens: the dark circles help increasing the contrast, making the defect more visible.

The algorithm exploited to detect the real defects will be now presented. Masks obtained from a reference without defects are used to compensate for the deformation of the black disks caused by the non-null curvatures of the lens. Masks delimit the regions on the images where the analysis will be performed. They also serve to cover the presence of structures inside the lens itself (e.g. injection point, junction lines, …) and the uncontrolled reflections of the light source on the different components of the machine, that can lead to erroneous analysis. Every frame has its own mask, created starting from the acquisition of a lens without defects. Applying the marching squares algorithm on the red channel of the image (since the lens is principally made of red PMMA, defects result enhanced in that spectral region), the global threshold level is chosen in order not to neglect faint flaws nor to classify dirt as a defect. The contours obtained are used to fill the masks, setting the external points to 0 and the internal ones to 1. When in a frame a junction line is present, caused by different fluid castings, the situation is more complex, due to the different position it can have depending on the physical properties of the injected material. In order to detect the junction line, the marching squares algorithm is applied also on the gradient of the image to enhance the sharp variations. Once the masks are built, each acquired frame is multiplied pixel by pixel with the corresponding mask. This process leads to a revised image that, in case of no defects, should appear dark (artificially set to zero the pixels

where the analysis is not of interest and with very low intensity values induced by the presence of the disks in the remaining pixels). To find the defects in the refraction region, in each image the marching squares algorithm is used a first time with a specified threshold and a constraint on the minimum area analyzable. Blemishes with intensity below that threshold or composed by multiple disconnected small zones are not detected at this stage. In order to locate also these defects, a second marching squares algorithm is run with a lower threshold, close to the background intensity of the image. Then, the pixel intensities near those contours are lifted proportionally to the proximity to the centroid and to the area delimited by them. Thanks to a third application of the marching square algorithm on this latter modified image, using the first threshold, it is possible now to identify defects composed by a multitude of small areas that without the enhancement process were impossible to be detected. The severity of a defect is proportional to light deviated by the blemish itself which is dependent on the perceived intensity of the flaw and on its extension. For what concerns the intensity, every pixel value in the range $[0, 255]$ is mapped with a piecewise quadratic polynomial function in the range $[0, 1]$; the mapped values represent weights that encode the severity associated to a defect of unitary area with all uniform intensity. All those values relative to pixels inside the contour that delimits the defect are summed up to obtain a value in the range $[0, +\infty[$ that brings with it also the information about the extension. Finally, the sum is transformed in the interval $[0, 1[$ with an inverse exponential function. The result represents an estimate of the severity of the blemish. An example of single and cluster defect is represented in *Fig. 5.3.11*. At the end of the analysis of the single frames, neighboring defects are clustered. For every cluster, a flaw is signaled in the centroid of its connected components. To estimate the clusters' severity, the assumption that the total light deviated is the sum of the light deviated by every single defect together with the choice of using an exponential function in the last step make the computation straightforward:

$$S(D_C) = 1 - \big(1 - S(D_1)\big)\cdot\big(1 - S(D_2)\big)\cdot \ldots \cdot (1 - S(D_n)) \tag{5.3.1}$$

where D_C is the clustered defect, $D_1, \ldots, D_n$ are the defects that constitute the cluster and $S(D)$ is the severity of a defect.

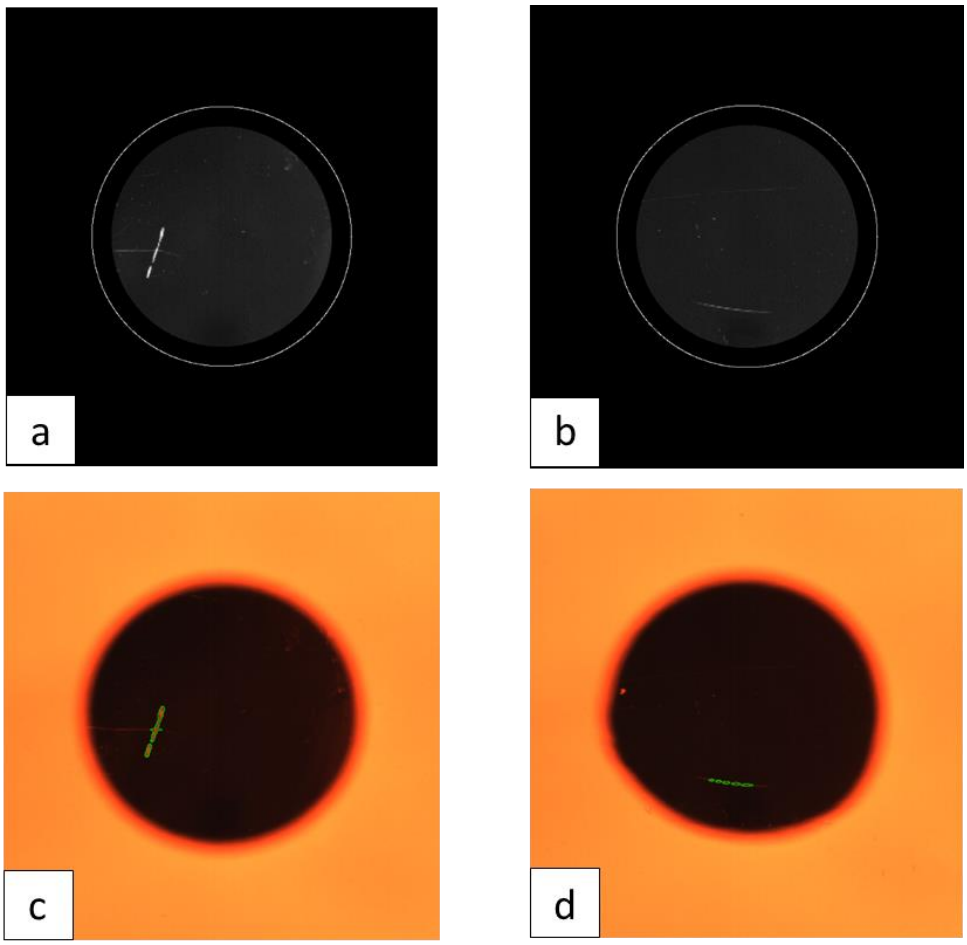

Fig. 5.3.11: on the top panels (a, b) it is possible to observe the mask performed on the black disk, on bottom panels (c, d), the original images with the highlighted defect are reported. For the first case (left side panels, a, c) the defect is sufficiently clear and it is recognized by the algorithm at first try, while in the second case (right side panels, b, d) only some spots of the defects have intensity higher than the threshold, so the clustering described in the last section is needed to classify correctly the defect.

5.3.3 Pattern region

The vinyl-disk defect is ascribed to a non-uniform filling process during injection. The flaws are mostly located in the regions that are far from the injection point. Therefore, the analysis was focused on specific frames where this type of defect is probably generated. This flaw does not alter in a significant way the absorption or the refraction of light, thus a new detection tool is required in order to recognize it. An opaque, black, rectangular pattern is exploited in this case, with three vertically-oriented holes in it. The mild light deflections introduced by the vinyl-disk blemish have the characteristics of being periodic and oriented, like ripples. Without the presence of flaws on the lens, the sides of the edges of the slots on the rectangular pattern may appear distorted due to the different curvature of the lens and blurred for the variety of incident angles of the light beams: in the described case the edge's deformation is very smooth. When a vinyl-disk defect affecting the lens is superimposed to the pattern, the pattern deformation becomes sharp and periodic. The algorithm first detects the areas of interest in the frame: the maximally stable extremal regions algorithm is employed to segment the pattern and isolate the slots profiles.[161,191] The area of the frame covered by each detected slot

position s is then processed applying the Laplace operator, to enhance the sudden variations of the slot profile due to the ripple-refraction of light (induced by the vinyl-disk defect) and producing the response $L(s)$. A Gaussian filter mask $G(s)$ is computed from each single slot region vertical borders to exclude out-of-slot occasional features, enhancing the ripples on the slot border and evaluating each slot singularly. The ratio

$$\frac{Avg[L(s) \cdot G(s)]}{Avg[G(s)]} \tag{5.3.2}$$

where $L(s) \cdot G(s)$ is the Laplacian response weighted pixel by pixel with the Gaussian mask. It represents the average intensity of ripples around the slot border and it is used as a $[0, 1]$ index to classify the *s*-slot area. Each frame can produce a total of three such indexes depending on the number of slots identified: if at least one of the indexes is higher than a threshold level, the flaw is recognized and characterized as vinyl-disk defect. An example is shown in *Fig. 5.3.12*.

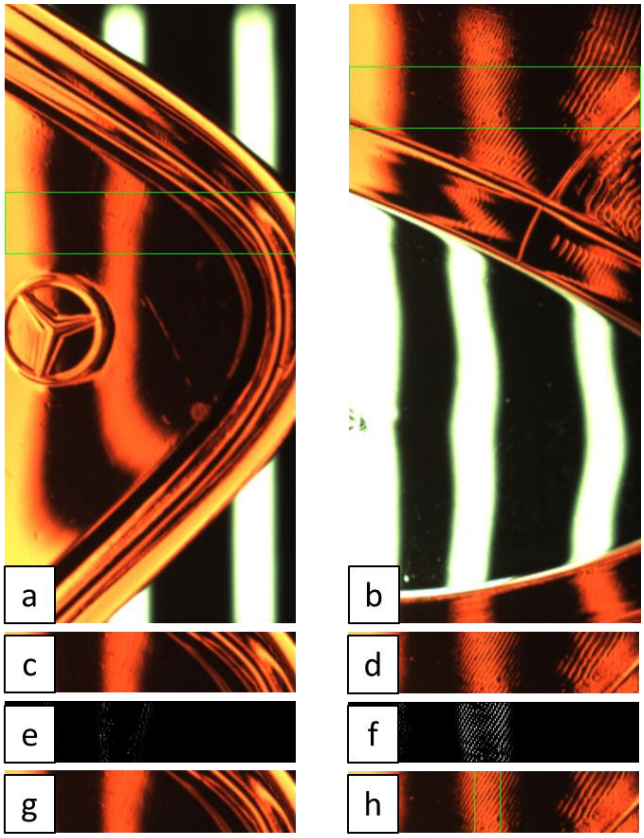

Fig. 5.3.12: examples of one compliant outer lens (left side, panels a, c, e, g) and one showing the vinyl-disk defect (right side, panels b, d, f, h). On top panels (a, b), automatic detection of the black-white-black pattern observed through the lens. The other panels represent, in order, the extracted area from the top panel (c, d), the result of the application of the Laplace operator to the extracted image (e, f) and the detection of the defect thanks to the application of the algorithm (g, h). No vertical green lines are visible in panel g because there is no vinyl-disk defect on that outer lens (in contrast to panel h).

5.4 Preliminary results

After a testing period when, in collaboration with quality operators, the parameters of the various algorithms have been calibrated to fulfill the requirements imposed by market requests, the prototype machine has undergone a validation phase, in which 204 lenses were scanned to retrieve some statistics about the reliability of the algorithms. The system has been configured to indicate three different values of severity for every lens scanned: “waste”, “warning”, or “fine”. In the two cases, “fine” and “waste”, there is no need of human operators, because the component would pass to the next stage of production or would be discarded, respectively. In the presence of the “warning” signal, the machine will report the typology of issue and the zone where the flaw has been detected to an operator, who will visually inspect and check the piece. The importance of the “warning” label is fundamental because it permits to save pieces without defects and to discard components that are affected by slight impurities, not recognized with certainty by the machine. The system has been tested with both compliant and non-compliant pieces, for a total number of 106 defects in 204 test parts. It is important to underline that, with respect to the standard production percentage, a higher number of non-compliant lenses have been selected to test the machine: it is not normal to have almost 50:50 ratio of compliant and non-compliant components. In most of the cases the machine was able to detect, without the need of an operator, if a component was faulty or not. In approximately 10% of the cases, the system was not able to sort the outer lens in examination, leading to a classification of the component as “warning”: in this case, a human operator is needed to state if the outer lens is compliant or not.

Less reliable is the case in which the machine classifies a piece as compliant even if it is, in practice, non-compliant and vice versa (i.e., false positives and false negatives). Speaking about the false positive, that is pieces that are compliant but are, indeed, discarded by the system without any possibility of a check by a human operator because classified as “waste”, from the validation period resulted that only the black spots show false positive. The amount of false positive ascribable to black spots is very small, the 0.61%. This means that the machine is able to detect approximately every compliant piece and a compliant component is incorrectly discarded only in very rare cases. For what concerns the false negatives, that is pieces that are classified by the prototype as compliant but show at least one severe defect and so they should be classified as non-compliant, the results are represented numerically in *Table 5.4.1*.

False negatives	
Cracks	7.46 %
Air bubbles	0.00 %
Black spots	0.00 %
Scratches	4.35 %
Incompleteness	0.00 %
Vinyl disk	0.00 %

Table 5.4.1: numeric results coming from the analysis of the false negative pieces. Only cracks and scratches show off false negative, in low percentage compared to the overall production. Speaking about air bubbles, black spots, lack of material and vinyl disk defects, no false negative has been detected.

The rates of false negative are higher than the false positives for the cracks and the scratches, while on the contrary the black spots do not show any false positive. The reason of these higher rates is due to the multiplicity of different cracks and scratches that may arise during the production of the lamp. Considering the scratches, they can occur with different shapes, depth and location, therefore it is not easy to set the algorithm in order to detect this high variety of possibilities. The percentage of the false negative are quite small, if compared to the entire production: only the 7.46% of breaks passing the automatic control were, effectively, non-compliant pieces and only the 4.35% of scratches. However, by implementing another camera (placed on top of the structure) it has been estimated that the detection accuracy related to the cracks will raise, reaching the 100%. Keeping in mind the beforementioned fact that the analyzed set has a non-standard 50:50 percentage of compliant and non-compliant pieces, the reliability of the prototype machine is extremely high. It never fails in detecting the air bubbles, vinyl disk and incompleteness defects, while the 0.61% of black sports have been detected as false positive and the 7.46% and 4.35% of cracks and scratches, respectively, as false negative.

5.5 Discussion

It has been possible to demonstrate that defects on the surface of the lamps can be automatically detected and identified, exploiting machine vision algorithms and a well synchronized axis motion and alignment system. The importance of having a specific machine is crucial inasmuch as the samples are almost transparent and with a non-constant surface curvature. A simple and straightforward CCD fixed camera geometry cannot be used to make all the work. The results, numerically speaking, are promising since human work can be reduced by about 90%, leading the operator to check only the components that the machine is not able to classify properly. The recognition efficiency of compliant pieces is extremely high, while the detection of non-compliant pieces is a little bit less efficient due to the high multiplicity of possible defects present on the surface of the lens. Moreover, the design of an easy and trivial visualization system to follow in real time the acquisition of the images (*Fig. 5.5.1*) makes it possible to easily implement the machine into the chain processes and to teach rapidly the human operators how to use it. This interface guarantees the visualization of the exact frame the camera is acquiring (top left image) and the last defect detected (bottom left). Furthermore, on a 2D sketch of the outer lens it is possible to visualize the location of the defects and their typologies, because at each different flaw category a diverse color is associated. The horizontal bar on the top (yellow in the image) defines if the component in analysis is compliant (green) not compliant (red) or warning (yellow). This interface helps and accelerates the work of the operator.

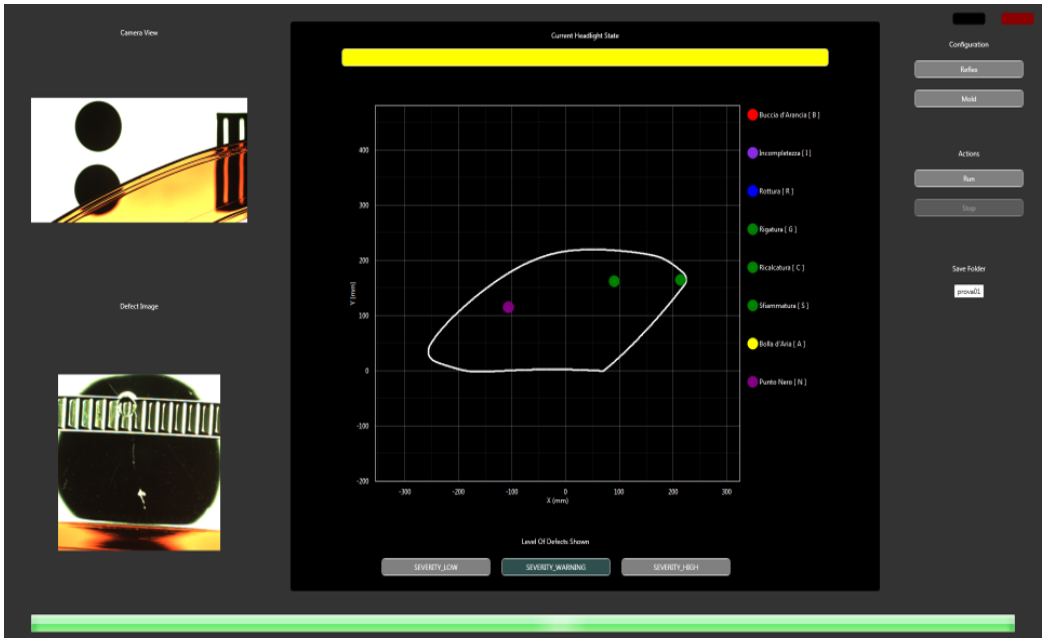

Fig. 5.5.1: user interface implemented on the automatic vision machine.

A further improvement of the algorithms and a minor modification of the inspection system are being developed and will further reduce the incidence of false positive results. To sum up, thanks to this machinery it is possible to scan the lens in a safe and light-controlled environment, implementing the automation in the automotive lighting industry sector. A faster, more precise and more objective quality control is thus attained with this system, implying a more correct identification of defects on the lens.

6. Characterization of reflectors

6.1 Surface defects

The target of the automotive lighting industry, as manufacturing industry, is to produce the highest number of compliant pieces reducing the production time, the costs, and the wastes, as recommended by the "lean manufacturing" method.[192–196] At the same time, stringent quality standards, together with compliance requirements have to be accomplished. Thus, this has to be performed striving for a constant improvement of both the quality of products and processes involved in the production.

In the specific case, particular attention has to be devoted to the quality of surfaces as parts of optical components.[197] The reflectors' polymer-based templates and supports, covered with ultra-thin metallic films calibrated to yield the desired reflective properties,[198–200] and protected from mechanical and chemical damaging by a silicon-based protective layer (see *Section 2.2* for further details),[201,202] are designed to guarantee the agreed performances and aesthetic features. As described in the previous sections, the purpose of the metallization process applied to automotive rear lamps is dual: on the one hand it must fulfill the photometric requirements imposed by the market regulation (functional value) while, on the other hand, it represents a style feature (aesthetical value). The presence of defects, which may arise during the several steps of every reflector's production process, hampers both the optical performance of the lamp and its appearance. In this section, the attention will be focused on the aesthetic defects that may also contribute from the beginning or during the work lifetime of the lamp with a performance degradation. Indeed, these defects invalidate both the optical requirements and the aesthetic aspect, and their origin must be investigated to prevent their occurrence.

The investigated samples come from standard production, thus representing a precise selection of effective issues occurring in the growth of metal-polymer heterojunctions. Selected surface defects were indeed identified and cut out from real reflectors by means of a bandsaw and blown with gaseous nitrogen, obtaining samples with appropriate size to fit the experimental laboratory requirements. In the specific case, plastic substrates are made of polycarbonate (PC) and metallized according to the process described in detail previously (*Section 2.2*).[203,204] Different kinds of defects involving the plastic substrate, the metallic film, and the HMDSO layer have been identified.

The observed faults have been classified in families, influencing the optical and aesthetic quality of the surfaces in an irreversible way: protrusions (PRTs), depressions (DPRs), opaque white spots (WSs), black spots (BSs) and scratches (SCRs). Representative images taken by means of a conventional camera of some examples of defects are shown in *Fig. 6.1.1*.

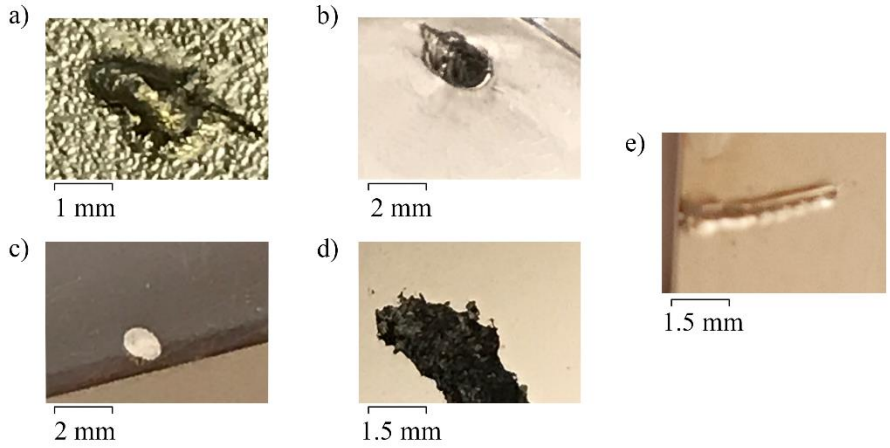

Fig. 6.1.1: images taken by a conventional camera of four examples of defects: (a) protrusions (PRT), (b) depressions (DPR), (c) white spots (WS), (d) black spots (BS), and (e) scratches (SCR).

The nature of these morphological and compositional alterations of the surfaces was thoroughly investigated by means of a combined microscopy-, spectroscopy-, and tomography-based approach. To this purpose, two complementary techniques have been exploited: Scanning Electron Microscopy (SEM) coupled with Energy Dispersive X-ray Spectroscopy (EDS), and Synchrotron-Radiation X-ray Computed Tomography (SR-CT), which has been recently demonstrated to play a fundamental role in the characterization of the internal structural failure (cracks) of materials.[205,206] The SEM and EDS measurements were performed at the CNR-IOM laboratory (Trieste) while the SR-CT analyses were performed at the SYRMEP beamline of the Elettra synchrotron (Trieste): access to the beamline was granted through proposal no. 20190256. The combined use of electron microscopy and SR-CT to perform a noninvasive investigation of this kind of multilayer films was exploited to indicate that both chemical contaminations, introduced in the fabrication process, and growth conditions strongly affect the presence of defects.

6.1.1 Protrusions (PRT)

3D defects induce strong structural and geometric distortions of the surface, thus being relevant from an optical point of view, together with chemical alterations that locally influence the reflectivity. The 3D reconstruction of a selected, typical protrusion (PRT) defect analyzed by means of X-ray SR-CT is shown in *Fig. 6.1.2*.

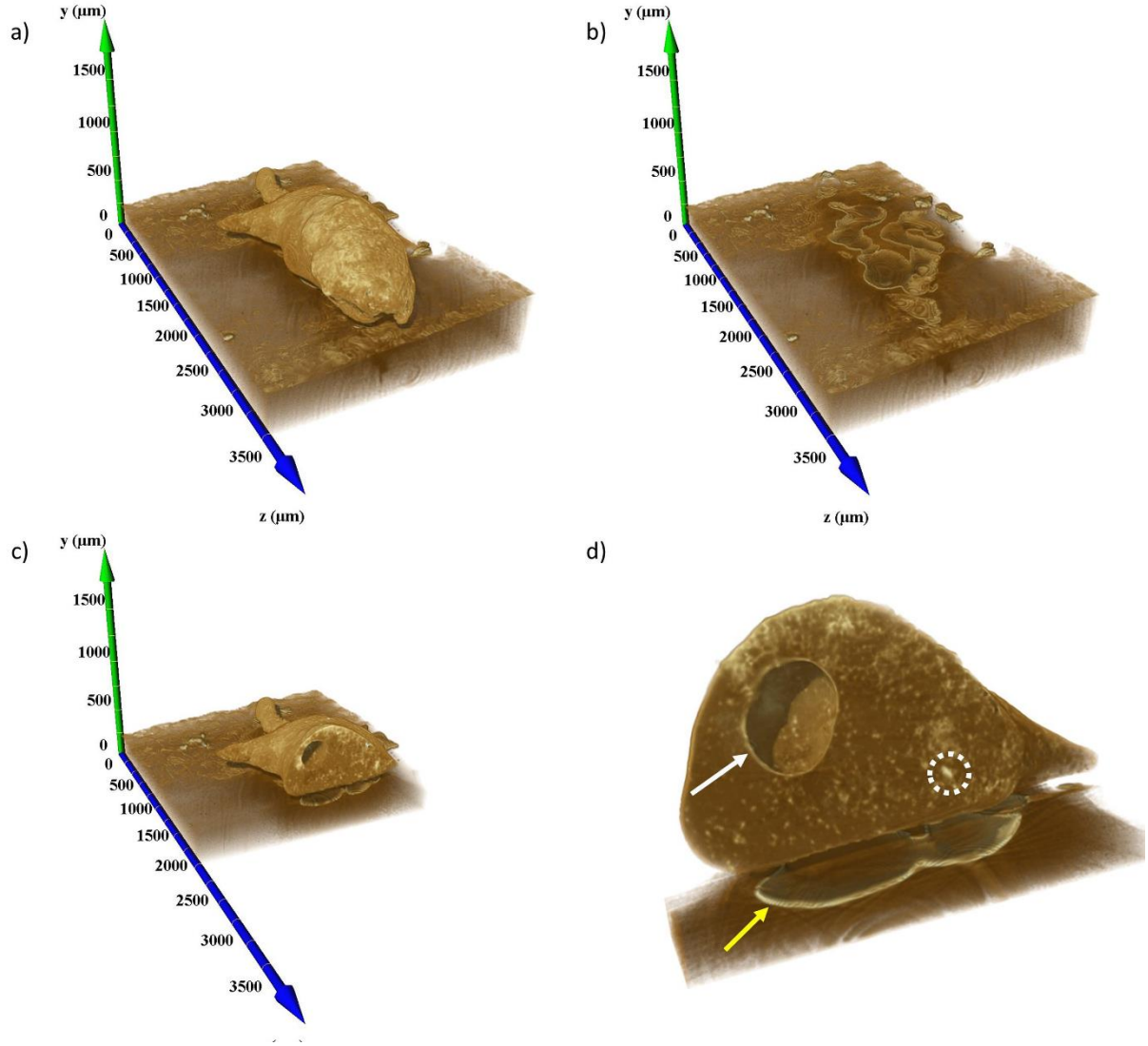

Fig. 6.1.2: (a) prospective view of a selected PRT defect imaged by means of SR-CT; (b) same prospective view where the PRT has been virtually removed, thus evidencing the cavities containing air at the interface with the substrate; (c) crosscut section, evidencing the cavities at the defect-substrate interface, as well as a large spheroidal air bubble trapped in the bulk of the PRT defect; (d) zoom of the clipped surface in (c), with arrows pointing to the air cavities inside (white arrow) and below (yellow arrow) the PRT. The dotted circle highlights a bright spot indicating strong X-ray attenuation. Parameters used for the acquisition and reconstruction: beam energy = 17 keV, pixel size = 2 µm, scintillator-sample distance = 100 mm, $\delta/\beta = 5$.

The size of the defect amounts to a maximum lateral extension of 3000 µm, with a depth of 800 µm (*Fig. 6.1.2a*). In *Fig. 6.1.2b*, the PRT has been virtually removed thus evidencing the presence of several air cavities in the substrate at the polymer-metal interface. A cut along the direction perpendicular to the surface is shown instead in *Fig. 6.1.2c* and in an enlarged front view in *Fig. 6.1.2d*. From the images, strong signal variations can be observed within the bulk of the defect, suggesting internal inhomogeneity due to the presence of different compositional phases. The latter cause contrast modulations, which are well captured in SR-CT images, allowing for image segmentation. Within the defect, the co-presence of non-miscible phases of Al (bright regions, high X-ray attenuation) and PC (dark regions, low X-ray

attenuations) uniformly dispersed inside the structure can be observed. Moreover, a spheroidal structure is visible inside the defect in panels (*c*) and (*d*, white arrow). The volume enclosed by this structure generates an extremely faint contrast, witnessing the presence of an air bubble in the protruding material. Similar, smaller cavities can be found also in the plastic substrate below the defect: they can be identified in panels (*b*), (*c*), and (*d*, yellow arrow) as a lack of substrate material with air trapped at the defective Al-PC interface. The results shown in *Fig. 6.1.2* suggest that both sintering of excess aluminum at the polycarbonate surface and, most probably, large Al clusters impinging at high temperature during the PVD process, can induce local melting of the interface. Polycarbonate and aluminum thus mix forming a porous structure that, upon venting of the deposition setup, is backfilled with air. Further support to this interpretation is provided by the depletion of PC under the defect due to inclusion in the 3D porous defective structure. A quantitative analysis has been performed to derive information about the internal composition of the Al-PC structure. The defect has a volume of 0.75 mm^3 and it is composed by PC (80.5%), Al (18.1%), and air (1.3%). Interestingly, a small fraction of the volume (0.1%) is occupied by a material with a significantly higher attenuation coefficient (see e.g. the small bright feature circled in *Fig. 6.1.2d*), suggesting the presence of a heavier contaminant in the Al evaporation flux used for the PVD process. A good contamination candidate is iron (Fe), which is indeed a stronger X-ray absorber than Al.

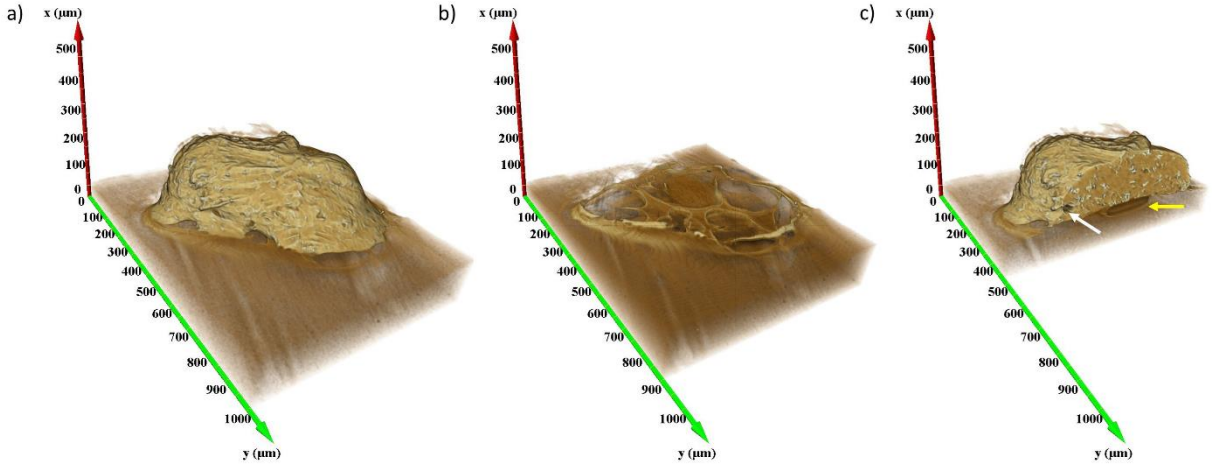

Fig. 6.1.3: (a) prospective view of a selected PRT defect imaged by means of SR-CT; (b) virtual removal of the PRT revealing air cavities in the underlying substrate; (c) crosscut view, with arrows pointing to the air cavities inside (white arrow) and below (yellow arrow) the PRT. Parameters used for the acquisition and reconstruction: beam energy = 17 keV, pixel size = 2 µm, scintillator-sample distance = 100 mm, $\delta/\beta = 5$.

The above situation is further confirmed by the analysis of a second, smaller defect (0.03 mm^3, *Fig 6.1.3a*). Again, the internal portion of the PRT (panel *c*) has a non-homogeneous filling, put in evidence by the color-scale contrast (polycarbonate darker than aluminum). Voids filled by trapped air can be observed in the bottom left corner of the crosscut section shown in *Fig 6.1.3c* (white arrow), together with cavities in the plastic substrate at the defect-substrate interface (yellow arrow). The interface is well imaged by virtually removing the PRT (*Fig 6.1.3b*). These observations consistently support the abovementioned interpretation, suggesting Al-induced melting of the plastic surface. The defect's composition is approximately the same

as in the previous case (79.4% of PC, 19.7% of Al, and 0.9% of air). At variance, no local contamination by heavier elements is observed here.

Another visualization of the same protrusion represented in *Fig. 6.1.3*, consisting on a set of 2D images taken at different penetration depth, is depicted in *Fig. 6.1.4*. The plane cutting the defect approaches the surface of the reflector from panel *a* (plane is outside the defect) to panel *k* (the plane is inside the reflector). With this set of images, the different internal composition is easily recognizable (panels *b-h*) as for the presence of void inside the protrusion (panels *d-f*). Moreover, the cavities in the substrate are clearly highlighted (panels *h-j*).

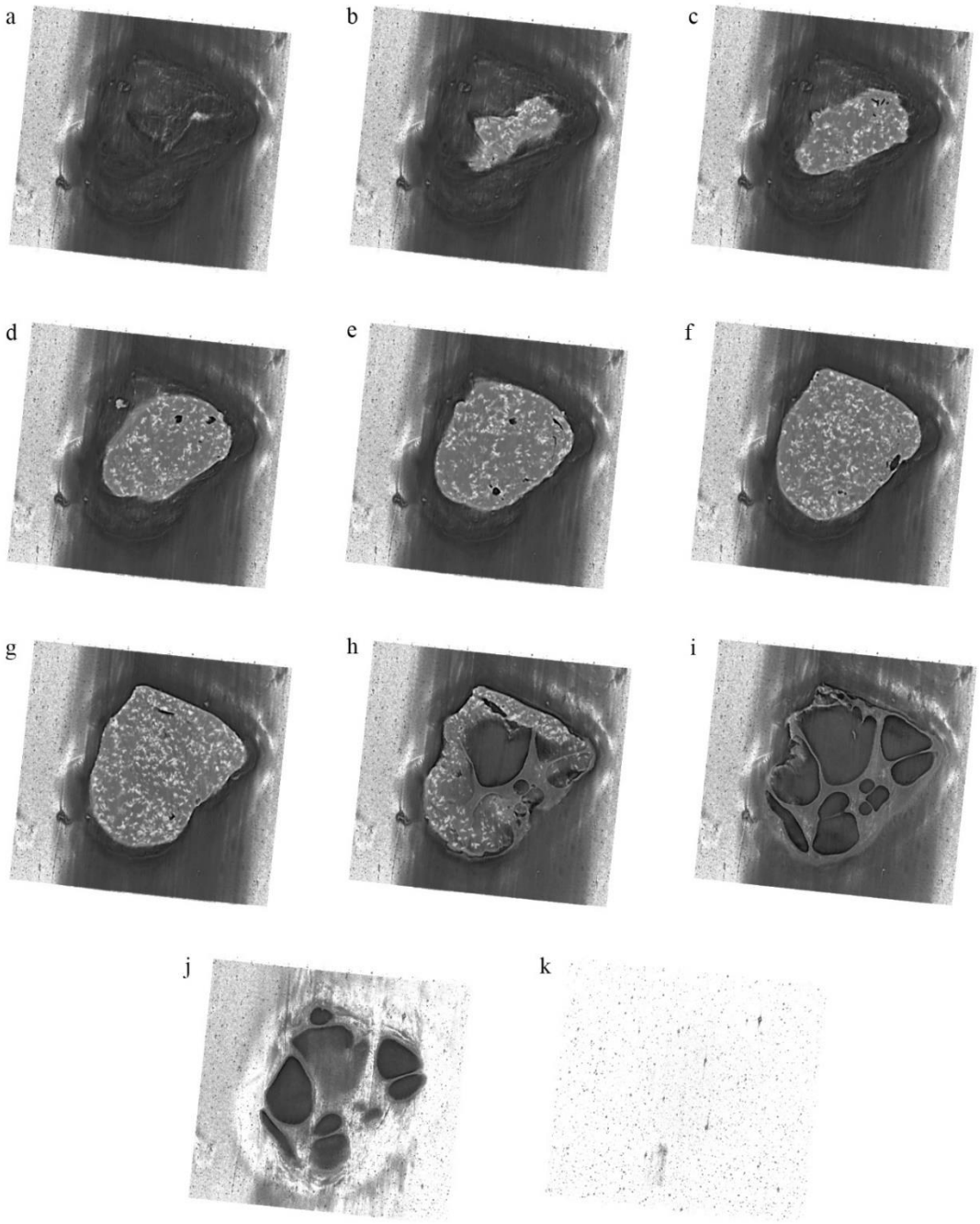

Fig. 6.1.4: set of bidimensional images representing the same protrusion depicted in figure 6.1.3. The penetration depth increases from panel (a, outside) to panel (k, completely inside the substrate) following a direction normal to the surface.

6.1.2 Depressions (DPR)

Concerning depression-like (DPR) defects, their lateral size allows the clear identification of the spots already to the naked eye. The tomographic reconstructions of two selected defects of this type are shown in *Fig. 6.1.5*.

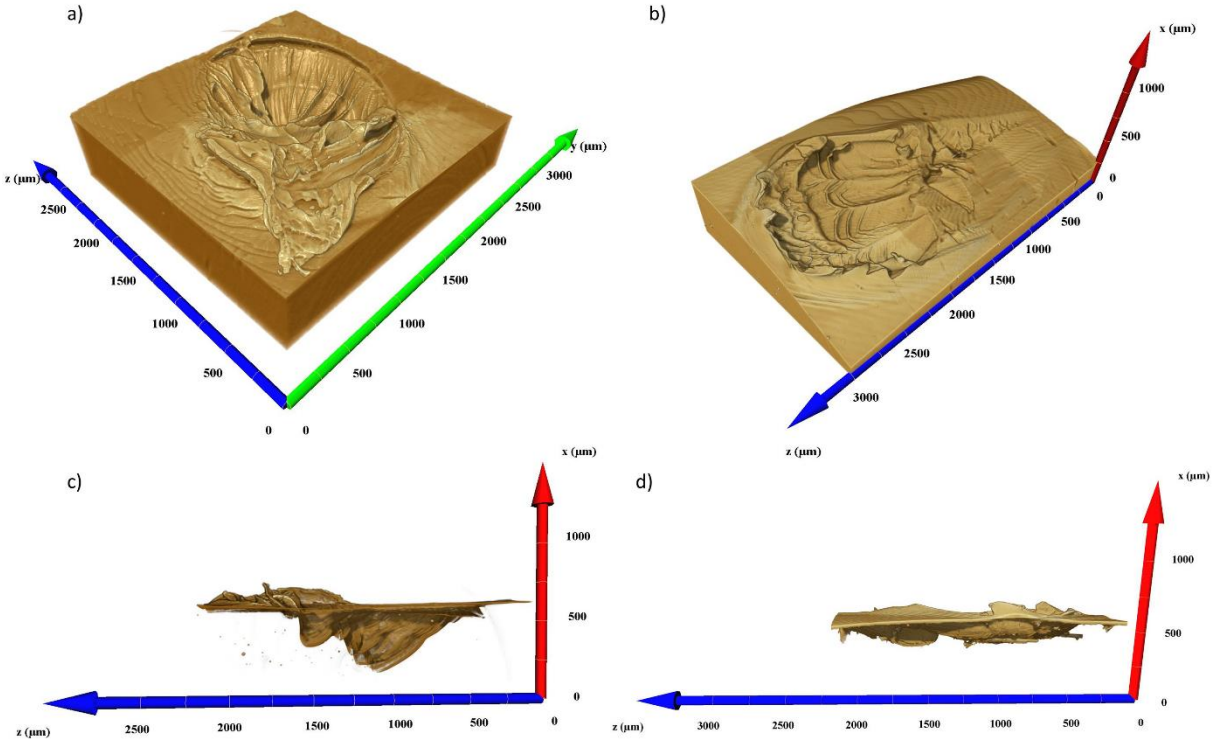

Fig. 6.1.5: a-b) prospective views of selected DPR defects; c-d) corresponding lateral views. Parameters used for the acquisition and reconstruction: beam energy = 17 keV, pixel size = 2 µm, scintillator-sample distance = 100 mm; δ/β = 100 for panels a) and c), δ/β = 150 for panels b) and d).

The concave DPRs have an irregular shape (panels *a-b*), both at the border and at the substrate-defect interface. Lateral views in panels (*c-d*) put in evidence the co-presence of structures protruding from the surface plane, which in one case can be as high as 200 µm (panel *c*). The morphology of the DPR defects suggests that the degree of melting at the Al-PC interface was so high that detachment of a PRT particle (similar to those previously shown) occurred. Indeed, melting induced by a hot Al cluster yields both formation of a porous 3D structure of mixed aluminum and polycarbonate composition, and the growth of a void at the defect-substrate interface due to PC migration. If the latter cavity is large enough, detachment of the whole structure may occur, possibly during the venting of the PVD setup, thus giving origin to the DPR defects. The comparable lateral size of the two typologies of defects further corroborates this interpretation.

6.1.3 White spots (WS)

Three additional flaw typologies have been identified by SEM. Two of them, based on their appearance and color upon simple optical inspection, have been classified as white and black spots (WS and BS, see *Section 6.1.4* for the BS defect description), while the last type is

called scratch (SCR, see *Section 6.1.5*). The WS is an extended region of approximately 1 mm width characterized by a geometrically regular superstructure growing on top of the metallized surface, altering both the reflectance properties of the film and its aesthetic appearance. When imaged by means of SR-CT, the defect does not show any 3D structure, affecting only the interface between the plastic substrate and the Al film, without any involvement of the underlying polycarbonate bulk (*Fig. 6.1.6*).

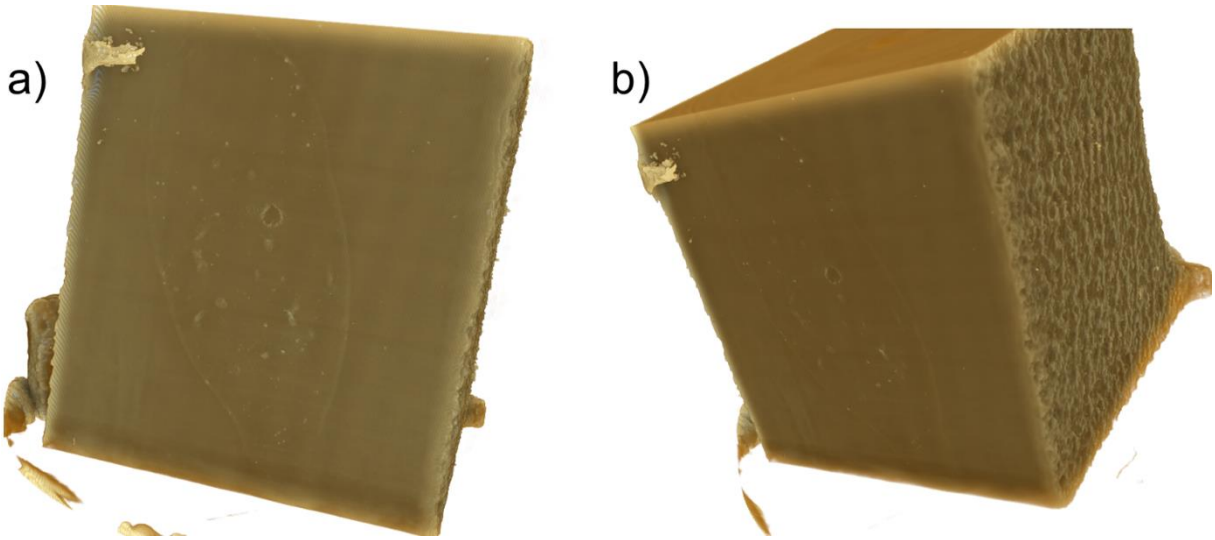

Fig. 6.1.6: (a) frontal and (b) lateral oblique views of a WS defect, as imaged by means of SR-CT (lateral dimensions 1.5 mm × 1.85 mm); parameters used for the acquisition and reconstruction: beam energy = 17 keV, pixel size = 2 µm, scintillator-sample distance = 100 mm, δ/β = 150.

Fig. 6.1.7 shows, instead, two SEM images of the same WS defect depicted in *Fig. 6.1.6*.

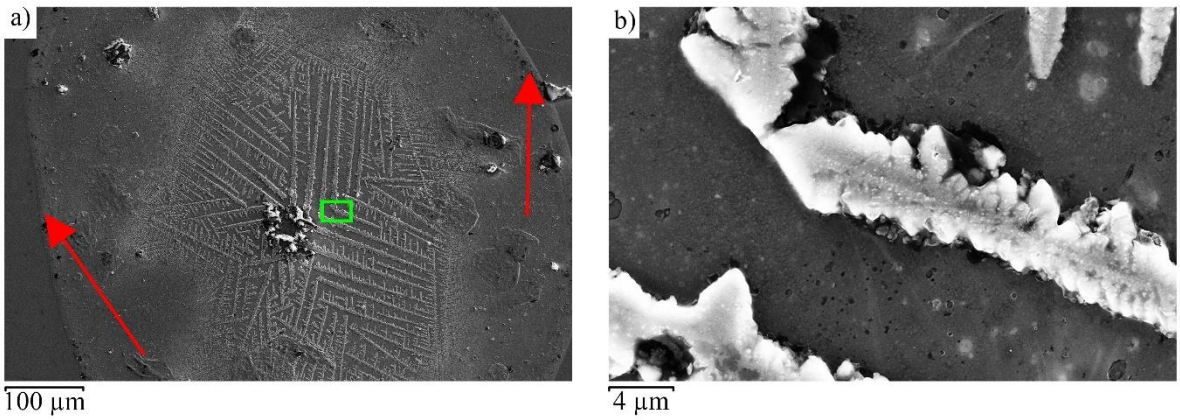

Fig. 6.1.7: SEM images representing the WS defect at different magnification; (a) overview taken with the Everhart-Thornley secondary electrons detector at 5 kV; (b) enlargement of the detail represented inside the green rectangle in panel (a) taken at 15 kV using the In-lens detector.

Fig. 6.1.7a represents an overview of the blemish which extends inside the borders pointed by the red arrows. The image is taken by the secondary electrons' detector, providing information on topographic details within the imaged area. Indeed, taking into consideration the defocusing difference between the borders and the outer substrate, it is quite clear that the flaw is

superimposed to the substrate, thus protruding from the surface. This is further confirmed (see again the red arrows) by the geometric shading effects due to the electron beam impinging direction. At the center of the defect a bright contrast, regular, geometric, fractal-like structure of about 200 μm × 300 μm is visible. A detail of the region enclosed in the green rectangle is depicted in *Fig. 6.1.7b* at higher magnification. The sharp edges and the regularity of the bright structure above the darker background support the thesis of a crystalline contamination. To yield better insight, the chemical composition of the structure was investigated by EDS elemental maps of the region shown in *Fig. 6.1.7b*, displayed in *Fig. 6.1.8*.

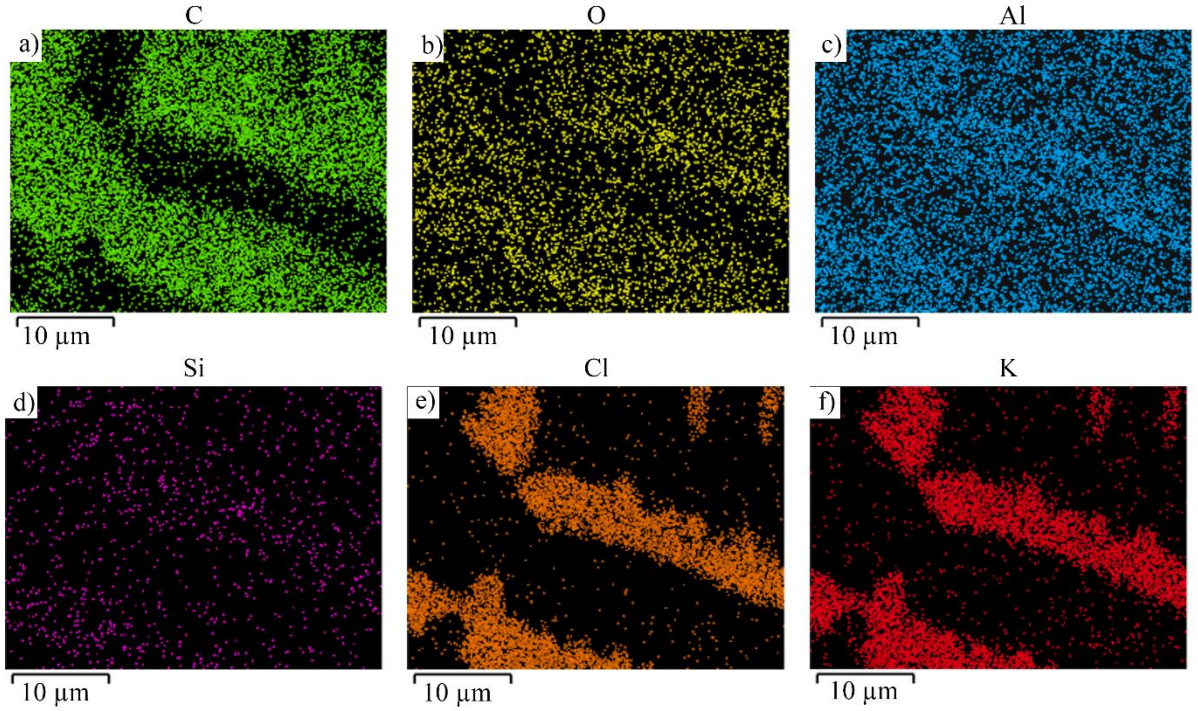

Fig. 6.1.8: EDS maps, collected at EHT = 5 kV showing the (a - green) carbon, (b - yellow) oxygen, (c - cyan) aluminum, (d - purple) silicon, (e - orange) chlorine, and (f - red) potassium spatial distributions over the detail represented in Fig. 6.1.7b.

It is important to recall that the presence of carbon and oxygen is associated with both the polymeric substrate and, to a smaller fraction, the HMDSO protective layer. The silicon signal is instead specific of the protector deposited on top of the sample to prevent oxidation, while aluminum is the only component of the metallic film. Thus, potassium and chlorine signals were not expected, therefore witnessing chemical contamination. The carbon map (*Fig. 6.1.8a*) is not homogeneous, indicating C depletion or screening at the dendritic defect, accompanied by complementary signals of both Cl and K (*Fig. 6.1.8e* and *Fig. 6.1.8f*, respectively). Same reasoning to be applied to the O signal (panel *b*), even if in this latter case the difference is less pronounced. The C and O signals are dominated by the contribution from the substrate polymer: this further support the idea that the defect is additive, thus covering the PC substrate. In addition, the homogeneity of the Al and Si maps (panels *c* and *d*, respectively) indicates that the K and Cl contaminations are not present over but below the metal thin film. It is possible to conclude that the contaminants are stacked at the interface between the plastic substrate and the Al film. Therefore, the surface has been contaminated prior to the PVD process, yielding incorporation and coating of the defect by Al. The high temperature of the metal Al atoms impinging at the surface upon metallization may also explain the induced recrystallization

promoted by heat. The structural and chemical characteristics of the defect are compatible with sweat, associated with human contamination of the polymer surface. The chemical composition of the sweat, showing the presence of both Cl and K elements, supports this scenario.[207,208]

6.1.4 Black spots (BS)

BSs are characterized by a dark region which does not reflect visible light and has a completely different appearance when compared to the surrounding, intact multilayer. In this case SR-CT does not provide useful insight into the nature of the defect, since the latter is limited to the very terminal layers and does not involve the underlying PC bulk. Interestingly, the SEM image collected over the BS (*Fig. 6.1.9a*) is affected, even when using the lowest EHT necessary to be surface sensitive, by strong electrostatic charging effects.

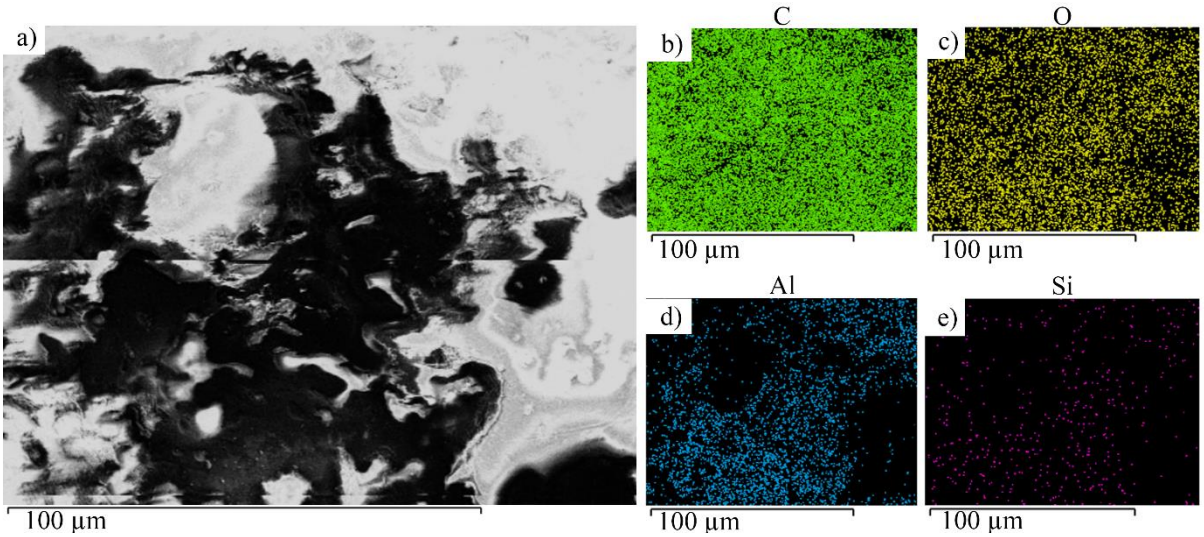

Fig. 6.1.9: SEM image taken at 5.0 kV with the In-Lens detector (a); EDS maps of (b - green) carbon, (c - yellow) oxygen, (d - cyan) aluminum, and (e - purple) silicon collected on the black spot defect shown in (a).

This indicates a locally poor electrical conductivity (white regions in *Fig. 6.1.9a*) associated with the exposure of the dielectric substrate and with a depletion of the conductive Al layer. Indeed, the corresponding EDS maps (panels *b-e* in *Fig. 6.1.9*) put in evidence the local depletion of both Al and Si in correspondence of the white dielectric areas, while C and O are homogeneously distributed over the entire map. Consequently, it is concluded that the metallized layer has detached completely from this portion of the surface, leaving the bare black polymeric substrate exposed.

6.1.5 Scratches (SCR)

SEM was fundamental also to yield better insight into the scratches, with the typical size of these latter defects ranges from hundreds of nanometers to several centimeters. These defects may, depending on their size, affect both appearance and optical properties of the reflectors. Two different samples have been examined by both imaging (*Fig. 6.1.10*, panels *a, e*) and chemical contrast (*Fig. 6.1.10*, panels *b-d, f-h*). The two scratches differ in size, but most importantly in chemical composition. Specifically, the most evident discrepancies can be

evidenced when comparing the carbon and the aluminum maps: panel *d* shows Al depletion, accompanied by an increment of the C signal (panel *b*). On the contrary, the C and Al distribution maps (panels *f, h*) referring to the second scratch (*e*) are homogeneous. This remarkable difference is fundamental to understand that the scratch may arise in different stages of the fabrication process: the scratch represented in panel (*a*) arose after the metallization process, more probably due to a human fault than to a machine problem, while the other defect (panel *e*) was present on the reflector surface prior to the metallization process and is therefore ascribable to the injection molding process. The O maps (panels *c* and *g*) are always homogeneous because oxygen is one of the constituted elements of polycarbonate-based substrate.

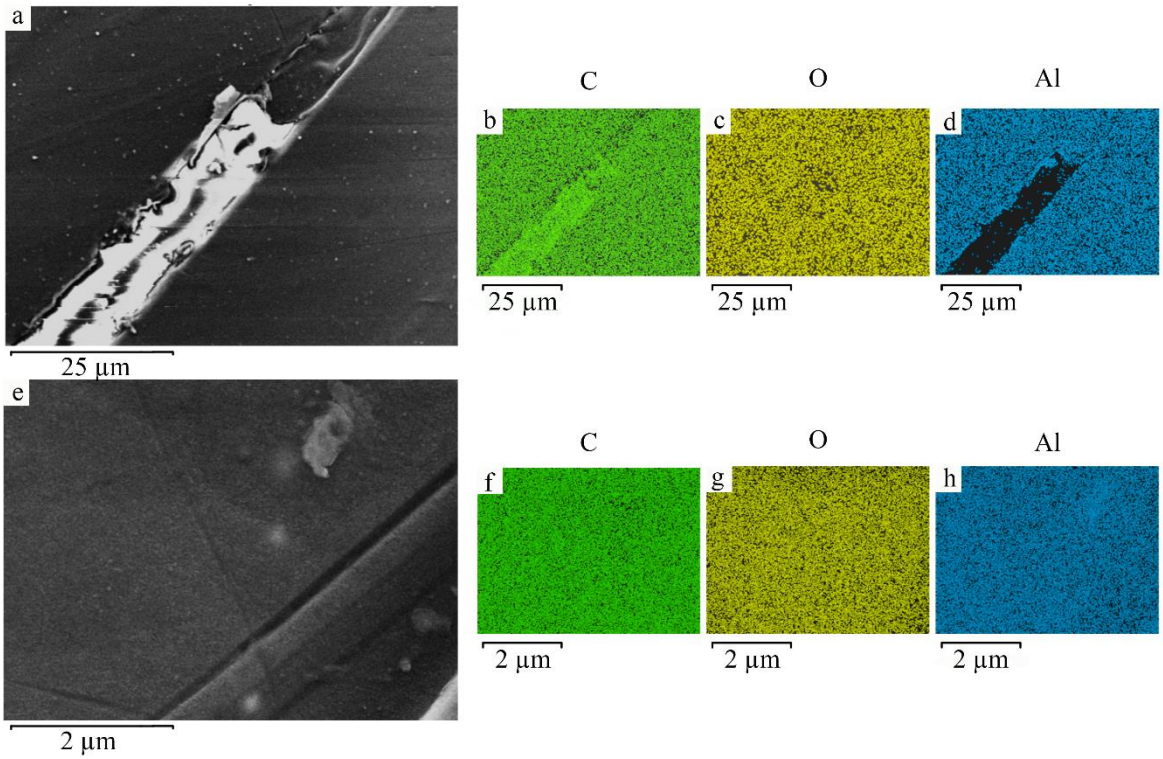

Fig. 6.1.10: SEM images taken at 5.0 kV with the In-Lens detector representing two different scratches (panels a and e). Panels b-d, EDS maps collected at EHT = 5 kV showing the (b - green) carbon, (c - yellow) oxygen, (d - cyan) aluminum spatial distributions of the defect represented in (a). Panels f-h, EDS maps collected at EHT = 5 kV showing the (f - green) carbon, (g - yellow) oxygen, (h - cyan) aluminum spatial distributions of the defect represented in (e).

6.1.6 Discussion

Morphological and compositional characterization details of defects appearing at the interface of aluminum-polymer heterostacks and originating from the growth process has been unraveled thanks to the measurements presented before. Insight into the buried interfaces has been achieved by means of SR-CT, while electron microscopy provided details about the role of contaminants. The SR-CT measurements were optimal to investigate the morphology of 3D defects involving interfaces between different materials, while on the contrary less information could be obtained when trying to characterize defects involving only the most terminal layer. On the other hand, SEM coupled with EDS enabled to obtain information on the morphology

and on the chemical distribution of the defects, thus complementing the information within a complete picture. Most of the issues arise from the Al deposition recipes, associated with the growth rate and the heat loading of the polymeric support. Thanks to these observations, the metal coating process recipes will be possibly revised, introducing lower deposition rates, compatibly with the pace required by the production chain. This will prevent aggregation of fused Al clusters that concurrently induces melting and damaging at the heterostack interfaces. Accordingly, external sources of chemical contamination will also be investigated, both environmental and due to the surface handling procedures. Moreover, the thorough characterization of the defects' morphology has already fostered the design and commissioning of an automated inspection system that allows for an in-line fast and efficient identification and classification of the defects, thus monitoring the optical and structural compliance in real time at the production stage. This prototype inspection system will be presented in *Section 6.3*.

6.2 Surface morphology and optical properties

In this chapter the attention will be focused on the functional parameters of the reflectors, characterizing the optical properties of the lamp, and correlating them to the different production recipes. A combined morphological and optical characterization of aluminum-coated thermoplastic polymer surfaces will be presented. The substrates used in this study consist of black polymeric plates of either polycarbonate (PC) or acrylonitrile butadiene styrene (ABS). A dedicated set of samples was specifically prepared following the same approach presented in the previous sections and adopted in the production chain, i.e. covering the plastic substrates with ultra-thin aluminum films evaporated *in vacuo*, on top of which a HMDSO protective layer was deposited. However, for this specific case the substrates have been chosen to be flat, in order to better perform the optical and structural analysis. Nevertheless, the production steps and instrumentations were exactly the same as the one exploited for the standard reflectors' preparations and described in *Section 2.2* (an image representing some of the samples is reported in *Fig. 6.2.1*).

Fig. 6.2.1: image representing some of the square-based (1.5 cm side) samples prepared for the optical and structural characterization.

Multiple, different preparation recipes were followed in order to investigate the role of several parameters, including the substrate material (PC or ABS), its thickness (1.5 or 3 mm), the thickness of the Al film and of the protective layer. The different treatments were studied in order to unravel the influence of substrate chemistry, substrate thickness, aluminum, protective layer thickness, and surface roughness on the actual optical reflectance properties. BRDF measurements, corroborated by surface morphological information obtained by means of AFM, correlate reflectance characteristics with the root mean square surface roughness, putting in evidence the role of the substrate and of the thin films' morphology. The results unravel information of interest within many applicative fields involving metal coating processes of plastic substrates in general, as well as for the specific case of automotive lighting. Optical properties, such as total reflectance, specular reflectance, and scattering behavior were examined at each step of the reflector coating process. Theoretical reflectance scattering models were sought to fit and parametrize the experimental data in a quantitative approach, with the future aim of implementing the results in a ray-tracing simulation environment as a tool for developers in the lighting design process. Finally, the possibility to extract information about

the surface roughness directly from the optical measurements has been verified, yielding remarkable results, in line with parallel, direct, but more expensive and time-consuming roughness measurements performed by means of AFM. The aforementioned measurements were performed within an international collaboration framework involving the KU Leuven Light&Lighting laboratory (for the BRDF measurements) and the Department of Engineering and Architecture of the Università degli Studi di Trieste (for the AFM measurements).

6.2.1 AFM analysis

The industrial evaporator is composed by 22 tungsten wires, which can contain a total of 44 Al spirals (see *Section 2.2* for details). Four configurations, depicted in *Fig. 6.2.2* and described in *Table 6.2.1*, have been selected to fill the evaporator in order to vary the amount of evaporated metal. In *Fig. 6.2.2*, the 2 wires-unit-cell configurations, repeated across the whole evaporator, are represented.

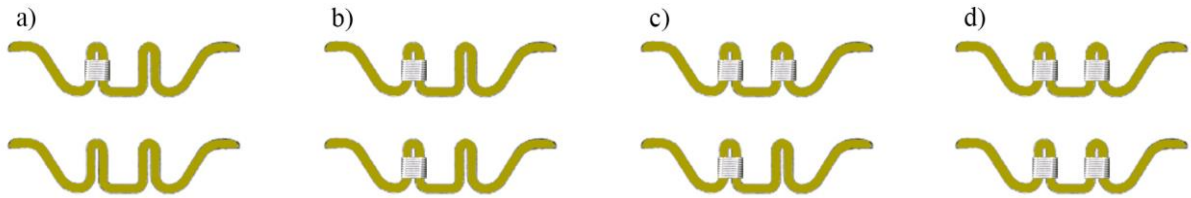

Fig. 6.2.2: schematic representation of the four possible configurations to fill the evaporators. The details are reported in table 6.2.1.

Case	Spirals on the 1st wire	Spirals on the 2nd wire	Total spirals	Al thickness (nm)
a)	1	0	11	68 ± 3
b)	1	1	22	81 ± 3
c)	2	1	33	115 ± 7
d)	2	2	44	166 ± 3

Table 6.2.1: quantitative description of the four cases used to fill the evaporator and depicted in Fig. 6.2.2. In the last column, the thickness measured with the AFM is reported.

No differences in the Al coating uniformity are observable if the spiral is placed in the meander on the left or in the one on the right of the wire due to the mechanic carousel rotation. These recipes result in different Al layer thickness values of (68 ± 3) nm, (81 ± 3) nm, (115 ± 7) nm, and (166 ± 3) nm, respectively. To determine the thickness values, the samples have been prepared by means of photolithography using a precise step-like photolithographic mask, which exhibited a sharp edge, useful to determine the thickness of the layers (as already shown in *Fig. 2.3.3*). Two HMDSO evaporation recipes were also tested, yielding (35 ± 4) nm and (61 ± 1) nm thick protective layers, respectively. For the silicon-based protective layer the photolithography was not exploited due to the solubility of the HMDSO in the acetone used for the lift-off. A physical masking with another glass slab over the main one was used. The AFM images showing the steps to determine the Al (panels *a-d*) and HMDSO (*e, f*) thickness are displayed in *Fig. 6.2.3*.

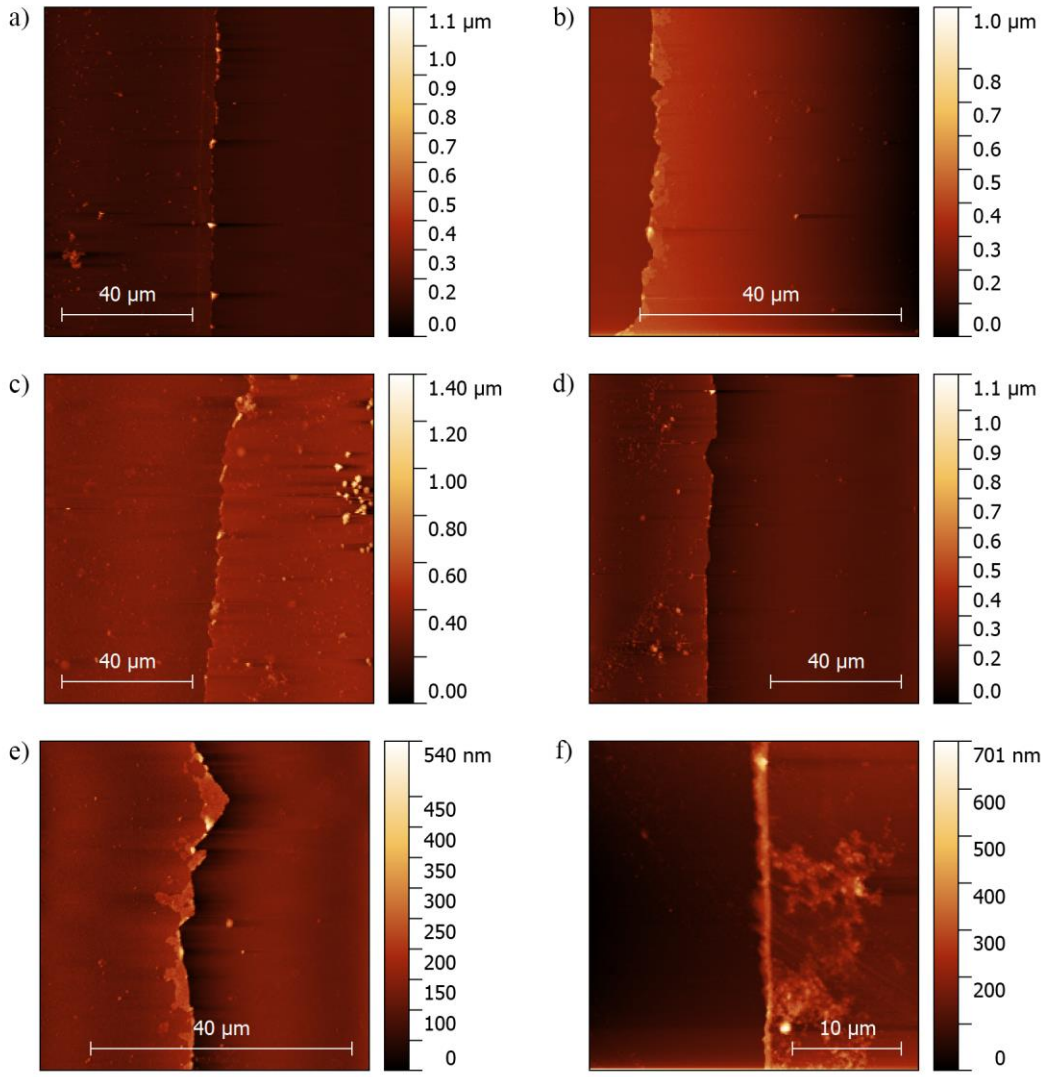

Fig. 6.2.3: AFM images representing the step between the coated and uncoated regions. Panels (a-d) refer to the metallization cycles listed in table 6.2.1, with layer thickness of (68 ± 3) nm, (81 ± 3) nm, (115 ± 7) nm, and (166 ± 3) nm, respectively. Panels (e, f) refers to the HMDSO thin film of thickness (35 ± 4) nm and (61 ± 1) nm respectively, generated starting from a 300 s and 600 s PECVD process.

Some environmental contaminations appear on the surfaces (for example on the right of the AFM image in panel *c* and *f*) because the measurements were performed in air and not in vacuo. On the raw AFM data, some height profiles in the direction perpendicular to the step have been taken (*Fig. 6.2.4*). The flat regions in the profile, corresponding to the bare glass substrate and the Al layer, have been fitted with a linear regression (light and dark green, respectively). The two regressions are not horizontal due to a possible tilt in the sample positioning. In the inflection point, a straight line perpendicular to the linear best fits has been taken (orange line, as highlighted in *Fig. 6.2.4a*) and the intersection of this line with the two regressions are represented with two red dots, named P_1 and P_2. In *Fig. 6.2.4b* an enlargement of the region of interest is represented: now the orange line is no more perpendicular to any of the two green lines, because the vertical dimensions are different with respect to the horizontal ones. Knowing the (P_1, P_2) distance, it is then possible to obtain the thickness of the deposited layer.

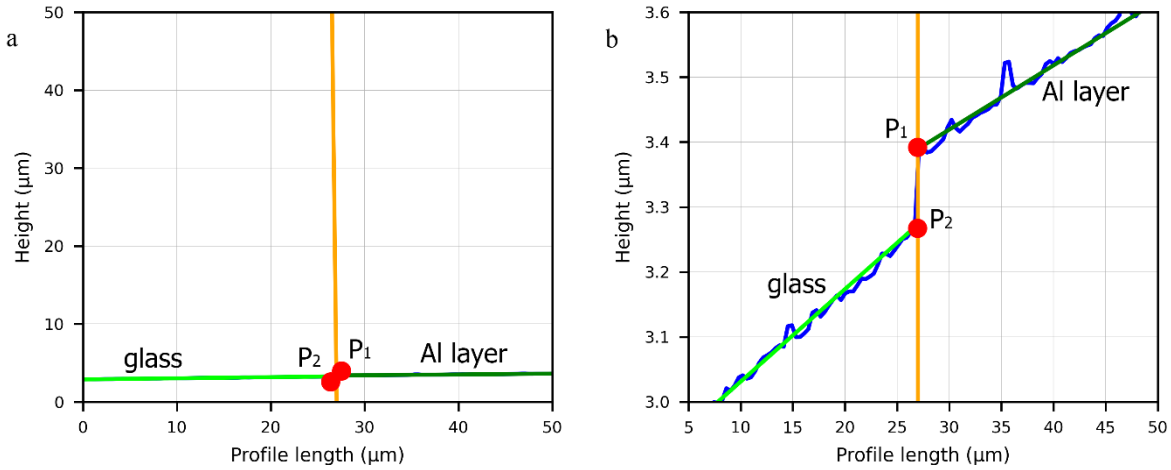

Fig. 6.2.4: example of data analysis for the thickness evaluation (note the scale expansion in panel b).

For every sample represented in *Fig. 6.2.3*, 20 height profiles have been acquired in regions not affected by environmental contaminations. The uncertainties are given by the statistical error originating from multiple measurements performed along the step edge. Concerning the surface roughness, it can be easily obtained from the AFM images as the root-mean-square (rms) (σ)

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}$$

where y_i is the difference between every local height with respect to the average height of the considered region. Both the actual roughness values and the associated uncertainties of the samples reported in *Fig. 6.2.5* were obtained statistically, i.e., by collecting and averaging local information from several regions on each sample (sampling window kept at a constant size of 20 µm × 20 µm). In particular, the central part (150 × 150 pixels) of each full-size image (256 × 256 pixels) was taken into account, excluding the external frame to avoid border and drift effects. The central region was divided in 9 subregions (50 × 50 pixels, around 4 µm × 4µm each). Mean values and standard deviations have been obtained performing the analysis on these 9 subregions (results are reported in *Table 6.2.2*). The rms surface roughness value for PC-based samples was found to be of the order of 5 nm, whereas ABS-supported films exhibited a higher surface roughness (20-30 nm), interestingly almost independent from the film thickness.

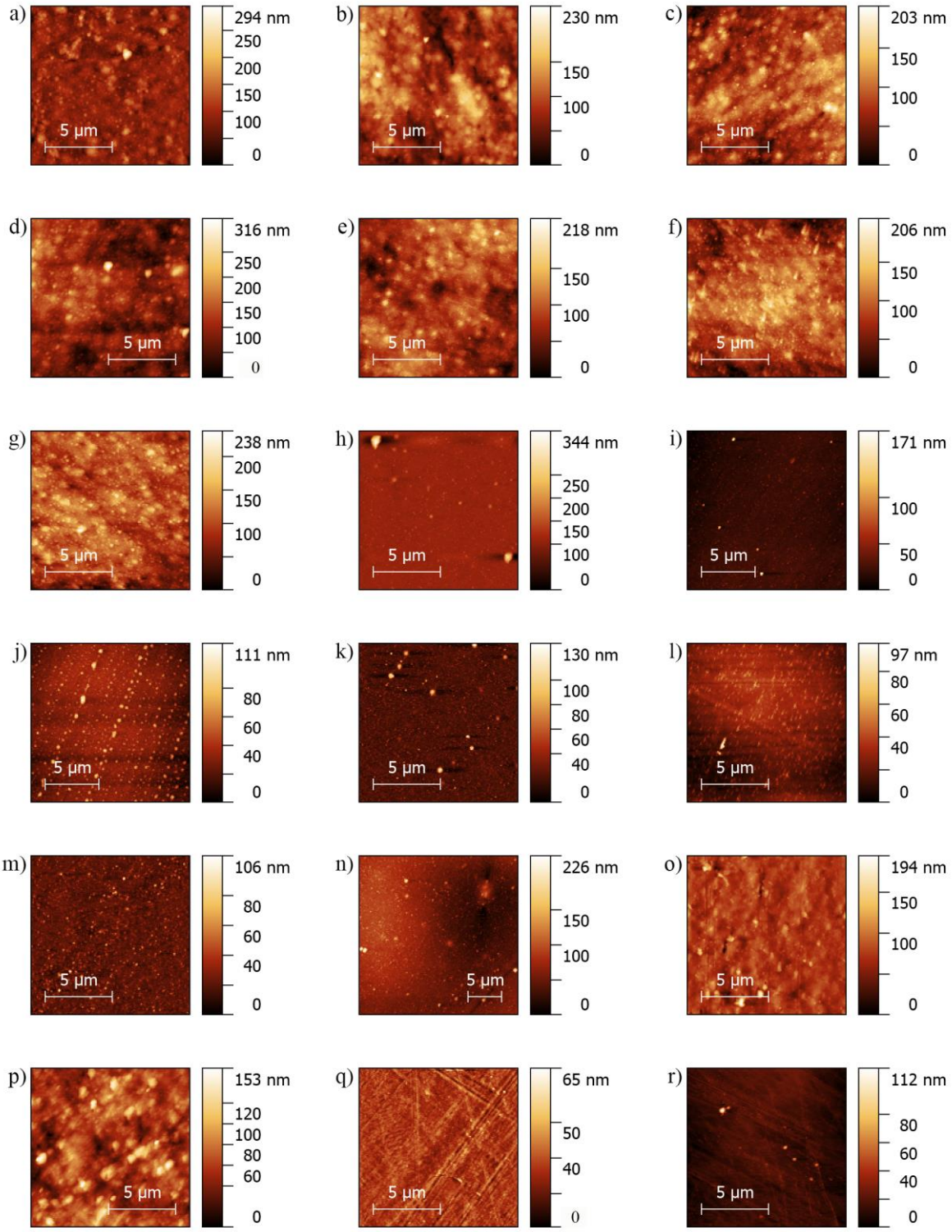

Fig. 6.2.5: selected AFM images of the samples listed in table 6.2.2

Id.	Substrate	Substrate thickness (mm)	Al thickness (nm)	Protector thickness (nm)	AFM rms (nm)
a)	ABS	1.5 ± 0.1	115 ± 7	35 ± 4	21 ± 4
b)	ABS	3.0 ± 0.1	68 ± 3	35 ± 4	26 ± 4
c)	ABS	3.0 ± 0.1	81 ± 3	35 ± 4	22 ± 3
d)	ABS	3.0 ± 0.1	115 ± 7	35 ± 4	30 ± 5
e)	ABS	3.0 ± 0.1	115 ± 7	0	20 ± 2
f)	ABS	3.0 ± 0.1	115 ± 3	61 ± 1	21 ± 2
g)	ABS	3.0 ± 0.1	166 ± 3	35 ± 4	23 ± 3
h)	PC	1.5 ± 0.1	115 ± 7	35 ± 4	6 ± 1
i)	PC	3.0 ± 0.1	68 ± 3	35 ± 4	5 ± 1
j)	PC	3.0 ± 0.1	81 ± 3	35 ± 4	8 ± 2
k)	PC	3.0 ± 0.1	115 ± 7	35 ± 4	7 ± 2
l)	PC	3.0 ± 0.1	115 ± 7	0	6 ± 1
m)	PC	3.0 ± 0.1	115 ± 7	61 ± 1	6 ± 1
n)	PC	3.0 ± 0.1	166 ± 3	35 ± 4	13 ± 2
o)	ABS	1.5 ± 0.1	0	0	14 ± 3
p)	ABS	3.0 ± 0.1	0	0	19 ± 4
q)	PC	1.5 ± 0.1	0	0	3 ± 1
r)	PC	3.0 ± 0.1	0	0	4 ± 1

Table 6.2.2: surface roughness values and uncertainties extracted from AFM measurements.

6.2.2 BRDF analysis

Role of the substrate

Since the rms surface roughness seemed to be primarily determined by the type of substrate (PC vs ABS), the first investigation regards the role of the substrate on surface reflectance. The spectral BRDF of coated PC and ABS samples of 1.5 mm and 3 mm thickness was therefore measured (*Fig. 6.2.6*). In the case shown in the figure, the thicknesses of the Al and HMDSO were the same for all samples, i.e. (115 ± 7) nm and (35 ± 4) nm, respectively. This specific configuration was selected because it best represents what is actually implemented at the applicative industrial level. The spectral BRDF, measured in the 45°,0°: 45°,180° geometry, is shown in *Fig. 6.2.6a*.

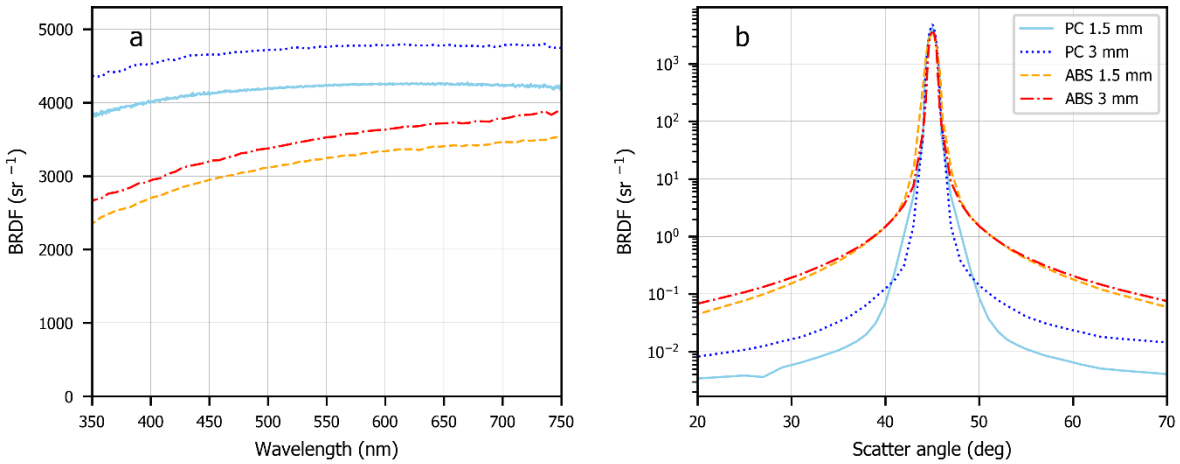

Fig. 6.2.6: measured BRDF for the different substrates (PC and ABS, 1.5 and 3 mm) coated with (115 ± 7) nm of Al and (35 ± 4) nm hexamethyldisiloxane; (a) spectral dependence (45°: 45° specular geometry); (b) angular (555 nm) dependence. Angles refer to the sample surface normal.

Major differences in the scattering properties of the two sample types can be evidenced readily at a first glance. The trend is very different for the two types of substrates: for PC, the spectral BRDF slightly increases with wavelength (~ 10%) for both the 1.5 mm (3874 sr^{-1} and 4209 sr^{-1}) and the 3 mm (4348 sr^{-1} and 4740 sr^{-1}) thick substrates, while for ABS this wavelength-dependence is more pronounced (~ 35%) for the two thicknesses (1.5 mm: 2372 sr^{-1} and 3503 sr^{-1}; 3 mm: 2669 sr^{-1} and 3918 sr^{-1}). The angular distribution of the BRDF, measured in the 45°,0°: x,180° geometry with x in the interval [20°,70°] at a wavelength of 555 nm, is depicted in *Fig. 6.2.6b* (angles refer to the sample surface normal). PC samples show a narrower specular distribution than ABS samples, corresponding to a greater capacity of PC to act as a mirror-like material. For both sample types, the BRDF increases with substrate thickness. Indeed, for the ABS samples, the specular BRDF peak at 555 nm for the 3 mm specimen is significantly higher than the value of the 1.5 mm one (approximately 9% difference, 3590 sr^{-1} vs. 3307 sr^{-1}, respectively), while for the PC specimens the relative difference is even larger (13%, 4703 sr^{-1} vs. 4112 sr^{-1}). This trend was found for all the different recipes exploited to prepare the samples, verifying the conclusion that rougher surfaces exhibit a much more diffuse behavior with respect to smoother samples. Profilometry measurements revealed that the 1.5 mm substrate displays long-range surface height deformations of about 12 μm in a 20 mm range, while for the 3 mm substrate the recorded value is an order of magnitude lower (*Fig. 6.2.7*). This finding confirms the information gained by the BRDF data, indicating that thinner substrates undergo deformation at the injection molding production stage, thus before the coating process.

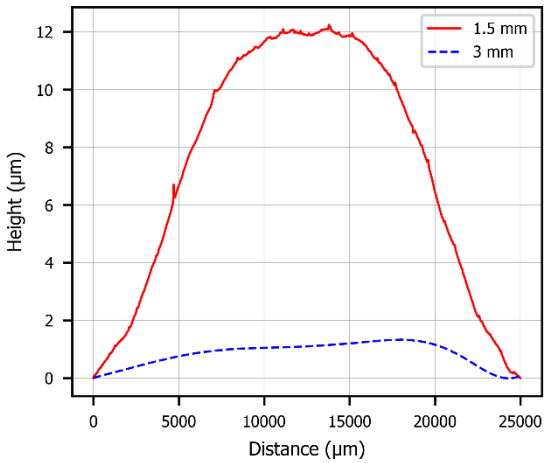

Fig. 6.2.7: profilometer measurements on the plastic plates; 1.5 mm PC (red full line), 3 mm PC (blue dashed line). Similar results were obtained for the ABS case.

Role of the reflective layer

Both PC and ABS substrates (1.5 and 3.0 mm thick) were coated with an Al layer of thickness between 70 and 170 nm, depending on the evaporation recipe. A (35 ± 4) nm HMDSO protective layer was post-deposited on top. The BRDF functions of the 3 mm PC samples with an Al coating thickness of (68 ± 3) nm, (81 ± 3) nm, (115 ± 7) nm, and (166 ± 3) nm, respectively, are shown in *Fig. 6.2.8*.

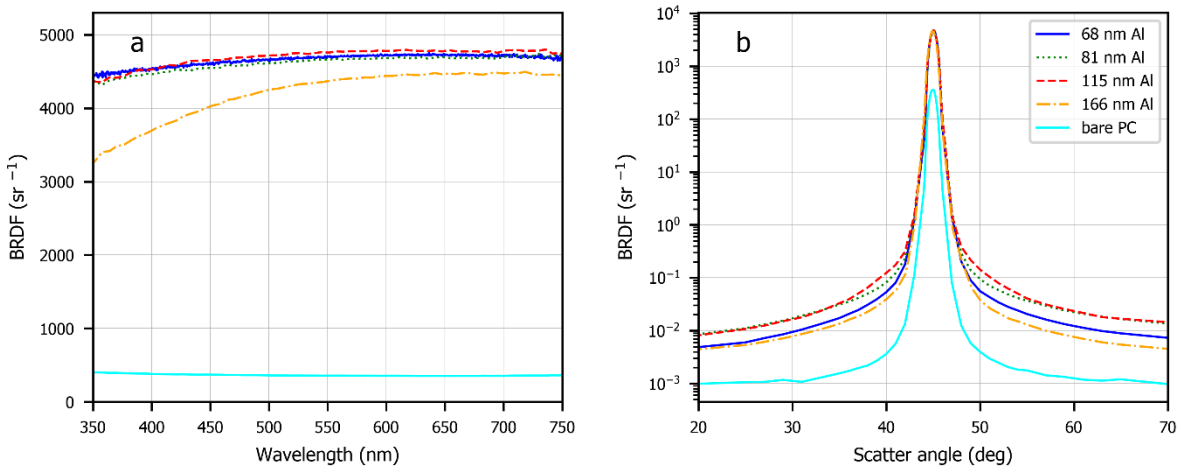

Fig. 6.2.8: measured dependence of the BRDF on the Al film thickness for a 3 mm PC substrate; (a) spectral (45° with respect to the sample surface normal) and (b) angular (555 nm) BRDF. The Al thicknesses reported in the figure are: 68 nm (blue continuous line), 81 nm (green dotted line), 115 nm (red dashed line), 166 nm (orange dot-dashed line) and bare PC (continuous cyan line). Apart from the bare PC, all the samples have a 35 nm of HMDSO over the Al layer.

The BRDF data of a 3 mm thick bare PC sample is also shown for direct comparison (no Al and no HMDSO on top, cyan full line). As expected, the addition of the Al layer increases the BRDF value of about one order of magnitude, both in the specular and off-specular directions. Interestingly, the Al layer film thickness does not seem to affect the BRDF peak value, apart from the case of the thickest film (166 nm), which yields a significantly lower BRDF intensity. As for the spectral dependence of the BRDF, measured in the 45°,0°: 45°,180°

geometry (*Fig. 6.2.8a*), it is observed that the overall shape of the function is similar for all Al films, showing increasing reflectance with wavelength. Clear differences are observed for what concerns the angular dependence of the BRDF (*Fig. 6.2.8b*), indicating best specular and off-specular reflectance for an Al layer thickness between 81 and 115 nm. The total reflectance of the samples was measured by means of a HunterLab Ultrascan PRO spectrophotometer (8°: d geometry). In particular, we found that the total reflectance (specular included) amounts to (82 ± 2) % for Al thickness in the range of 68-115 nm, slightly decreasing to (79 ± 1) % for the 166 nm Al film, in contrast with only (5 ± 1) % for the bare PC substrate. Almost similar conclusions are drawn from the results obtained for the thinner (1.5 mm) PC substrate, apart from the effects ascribed to the substrate deformation, as discussed above.

Similar to the PC case, the Al-covered 3 mm thick ABS samples yield an almost one order of magnitude higher BRDF than the bare ABS surface (*Fig. 6.2.9*). Regarding the spectral distribution (*Fig. 6.2.9a*), the ABS sample with an 81 nm thick Al film yields the highest signal, while the lowest BRDF is observed for the thickest coating (166 nm Al), in analogy with the PC case. The spectral trend is also similar, yielding progressively increasing BRDF values for increasing wavelength. Concerning the BRDF angular dependence (*Fig. 6.2.9b*), equivalently to the PC-based samples, the highest BRDF can be obtained with a coating thickness ranging between 81 nm and 115 nm. The total reflectance ranges from (70 ± 1) % to (69 ± 1) % for the samples characterized by 68 nm, 81 nm, and 115 nm Al layer, respectively, while it drops to (66 ± 1) % for the 166 nm Al sample. For the bare ABS substrate, the total reflectance is (5 ± 1) %. Therefore, it is found that for both ABS and PC substrates the Al coating thickness yielding best spectral and spatial (both in the specular and off-specular directions) properties ranges between 81 and 115 nm, corresponding to best integrated reflectance.

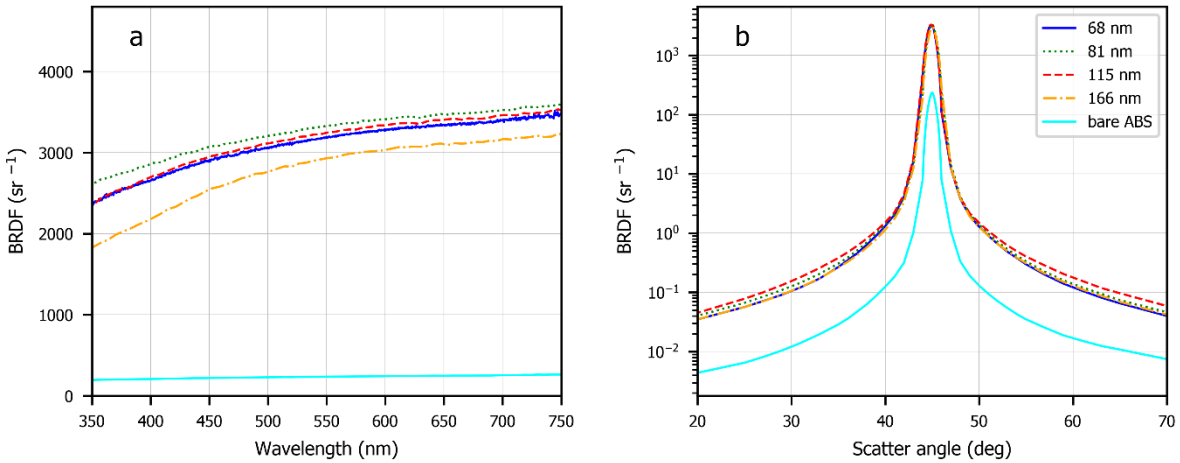

Fig. 6.2.9: measured dependence of the BRDF from the Al film thickness for a 3 mm ABS substrate; (a) spectral (45° with respect to the sample surface normal) and (b) angular (555 nm) BRDF. The Al thicknesses reported in the figure are: 68 nm (blue continuous line), 81 nm (green dotted line), 115 nm (red dashed line), 166 nm (orange dot-dashed line) and bare ABS (continuous cyan line). Apart from the bare ABS, all the samples have a 35 nm of HMDSO over the Al layer.

Role of the protective layer

As already mentioned, the HMDSO protective layer is deposited by means of PECVD over the Al film in order to prevent both oxidation and mechanical damage. The influence of

this necessary protection on the surface scattering properties was analyzed for six different types of samples. Different HMDSO evaporation recipes were adopted, yielding HMDSO films of (35 ± 4) and (61 ± 1) nm as quantified by means of AFM, respectively, on top of both PC and ABS substrates covered by a (115 ± 7) nm thick Al film. A bare Al film without protective layer was adopted as a reference. The measured spectral BRDF in the 45°,0°: 45°,180° geometry and the spatial distribution at 555 nm for the 3 mm thick PC substrates is depicted in *Fig. 6.2.10*. The total reflectance measured with the HunterLab spectrophotometer varies from (85 ± 1) %, referred to the sample without protective layer on top, (82 ± 1) % for the specimen with the 35 nm thick protective layer, and to (76 ± 1) % for the specimen with the 61 nm thick protective layer.

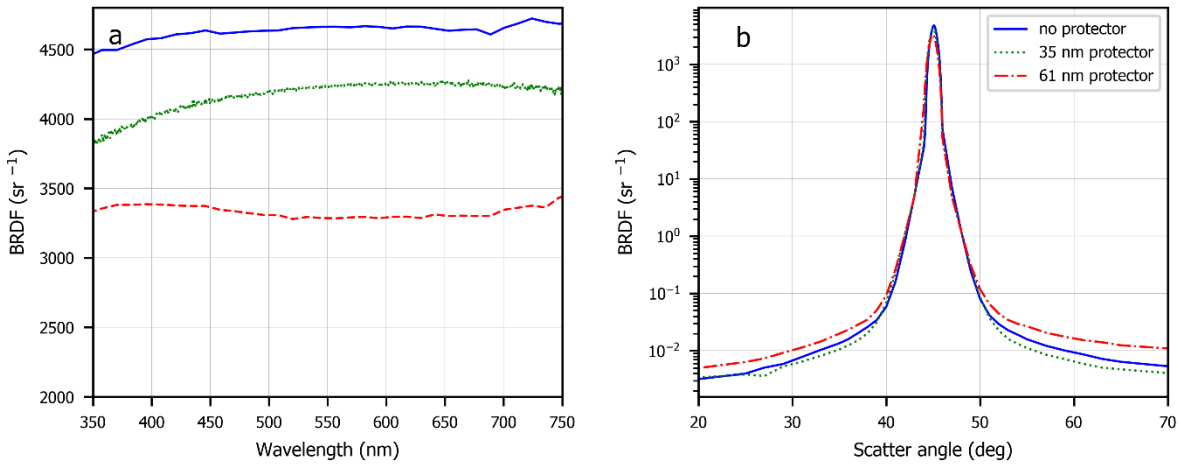

Fig. 6.2.10: measured dependence of the BRDF from the thickness of the protective layer deposited on 115 nm Al on 3 mm PC; (a) spectral (45° with respect to the sample surface normal) and (b) angular (555 nm) BRDF for samples with 0, (35 ± 4) and (61 ± 1) nm thick protective layer on top of a 115 nm thick Aluminum film (respectively blue continuous line, green dotted line and red dashed line).

As expected, the protective layer affects the spatial BRDF. The BRDF peak at 555 nm drops from 4732 sr^{-1} for the uncoated film to 4246 sr^{-1} and 3340 sr^{-1} for the samples covered by 35 and 61 nm of HMDSO, respectively. Interestingly, a 35 nm thick protective layer yields the lowest signal for angles far from the specular direction, while a thicker coating (61 nm) introduces a broader angular dispersion. Summarizing, the protective layer deposited on the Al film supported by PC hampers the reflectance in the specular direction, while it introduces an angular spreading.

Data for the ABS substrates metallized with 115 nm of Al and covered with different HMDSO films are shown in *Fig. 6.2.11*. The bare Al film displays the largest BRDF peak value at all wavelengths (*Fig. 6.2.11a*) while, at variance with the PC samples case, the sample with an intermediate protector layer thickness (35 nm) has the lowest specular reflectance. Remarkably, the thickest protective layer (61 nm) behaves like the bare aluminum at wavelengths below 430 nm, drastically changing its trend for wavelengths larger than 430 nm, resembling the 35 nm HMDSO sample profile. This is ascribed to the role of the CH species in HMDSO which are known to contribute with transition lines at a wavelength of 430 nm.[209,210] The contribution of the CH species is also imperceptibly visible on PC substrates (*Fig. 6.2.10a*) as a soft difference in the trend, starting from 430 nm, of the spectrum characterized by 61 nm of HMDSO on top of it. The different behaviors could be explained by diverse interaction with

the substrate. The angular spread is weakly affected, all three samples exhibit very similar properties (*Fig. 6.2.11b*). The total reflectance values range from (70 ± 1) % to (65 ± 1) % (the higher the HMDSO thickness, the lower the reflectance).

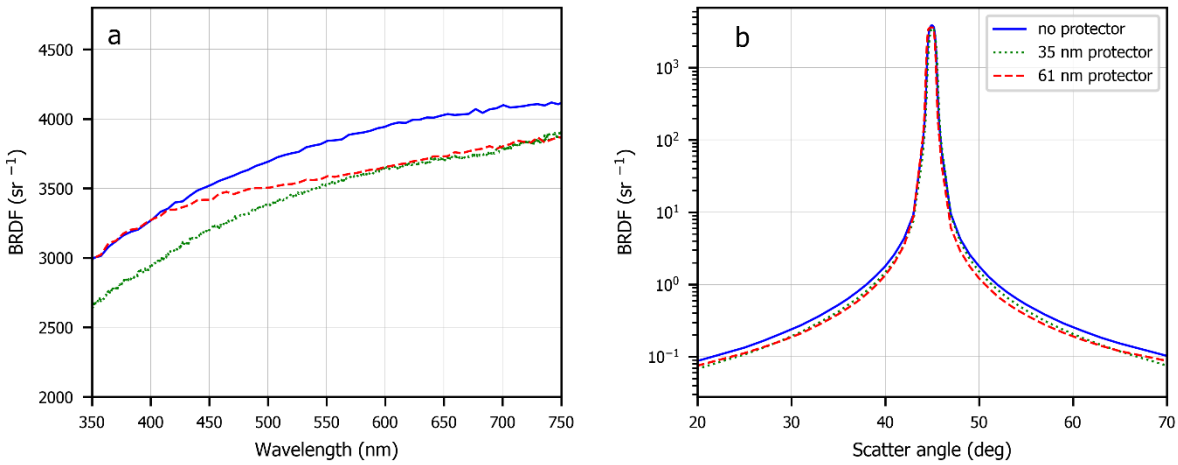

Fig. 6.2.11: measured dependence of the BRDF from the thickness of the protective layer deposited on 115 nm Al on 3 mm ABS; (a) spectral (45° with respect to the sample surface normal) and (b) angular (555 nm) BRDF for samples with 0, (35 ± 4) and (61 ± 1) nm thick protective layer on top of a 115 nm thick Aluminum film (respectively blue continuous line, green dotted line and red dashed line).

6.2.3 Data evaluation and modelling

The preparation of the samples, characterized by the PVD and PECVD processes, yields surfaces with isotropic roughness, as demonstrated by AFM images (*Fig 6.2.5*). Moreover, it became obvious from the BRDF data that incident light is reflected within a narrow scattering angle which brings us to compare the samples with perfect mirrors. These two observations lead us to employ the empirical *ABg* model, which is widely used in many fields,[211,212] to fit the BRDF data.[213] To analyze the specimens with the *ABg* model, we switched to the Harvey Shack representation; the most important coordinate in this reference system is $|\beta - \beta_0|$ that is the difference between the projection of the scattered beam ($\beta = sin\theta_{scatter}$) and the projection of the specular direction ($\beta_0 = sin\theta_{specular}$) in the plane of incidence. The *ABg* formula to fit the data is defined as

$$BRDF = \frac{A}{B + |\beta - \beta_0|^g} \qquad (6.2.1)$$

where the parameters (*A*, *B* and *g*) used in this model come from a rearrangement of the Harvey fitting function.[214] The BRDF spectra, taken at a fixed wavelength of 555 nm, have been plotted in a log-log graph with respect to $|\beta - \beta_0|$. The *ABg* model is not able to fit the experimental data in the specular reflection direction, in fact ray-tracer software platforms require the specular reflection as an additional parameter. With the experimental setup used, it is not possible to precisely measure the specular component due to the convolution of the incident and reflected beams' angular distributions, which affects a finite angular region that will be called, from now on, convolution-region (the angular distribution of the incident beam can be seen in *Section 3.4.3*). This, combined with the substrates showing small geometric deformations due to the molding process of the supports, with the finite size of the probe light

beam (diameter = 1.2 cm), and with the aperture size of the detector (diameter = 1.5 cm), results in a region where specular Fresnel reflection cannot be easily distinguished and decoupled from small angle scattering.[215] In a first approximation, it has been assumed that the signal contribution in the convolution-region is all ascribable to the specular Fresnel reflection. In fact, for the model fitting procedure, our interest is focused on the off-specular component of the reflected beam (i.e. pure surface scattering). In order to select properly the region of interest, a scan of the incident beam has been performed. This defines the angular region that must be excluded during the fitting of the model. An example of the data fitting to the *ABg* model is shown in *Fig. 6.2.12*, highlighting that the only scatter region is considered (i.e. convolution-region excluded).

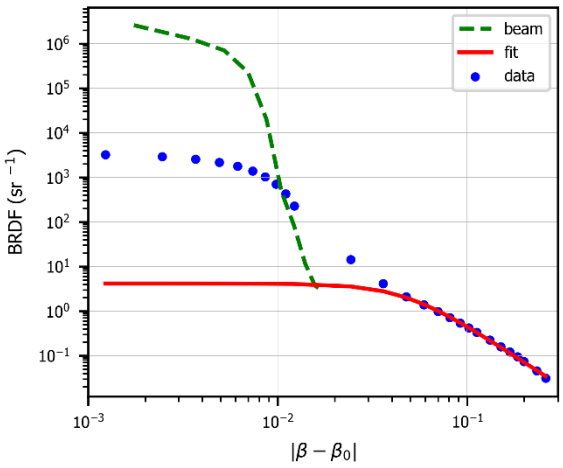

Fig. 6.2.12: Harvey Shack representation of the BRDF data (blue dots), together with the best fit according to the ABg model (red full line); the green dashed line represents the measured incident beam.

Quantitative information on the morphological properties can be obtained from the empirical *ABg* model parameters. In particular, the reconstruction of the BRDF function is attained by exploiting the fitted parameters. Integration over the space yields then the Total Integrated Scattering (*TIS*), defined as:

$$TIS = \int_{\varphi_s=0}^{2\pi} \int_{\theta_s=0}^{\pi/2} BRDF(\theta_i,\, \varphi_i,\, \theta_s,\, \varphi_s,\, \lambda) sin\theta_s cos\theta_s d\theta_s d\varphi_s. \qquad (6.2.2)$$

In our case, we calculated the *TIS* integral substituting the *ABg* expression for the BRDF and operating a simple change of variable expressing $|\beta - \beta_0|$ in terms of θ_i, θ_s, φ_s

$$TIS = \int_{\varphi_s=0}^{2\pi} \int_{\theta_s=0}^{\pi/2} \frac{A}{B + [(sin\theta_s \cdot sin\varphi_s)^2 + (sin\theta_s \cdot cos\theta_s - sin\theta_i)^2]^{g/2}} sin\theta_s cos\theta_s d\theta_s d\varphi_s, \qquad (6.2.3)$$

excluding the angles in both θ_s and φ_s representing the specular direction. The wavelength of the exploited incident light is larger than the rms surface roughness (λ > rms) and it is possible to assume that scattering (outside the convolution-region) is mainly ascribable to the surface roughness of the specimen. Moreover, as demonstrated by AFM measurements, samples exhibit an isotropic surface roughness. These motivations permit to obtain the surface roughness (σ) from the *TIS* using the formula

$$TIS = 1 - e^{-(\frac{4\pi cos\theta_i \sigma}{\lambda})^2} \quad (6.2.4)$$

where θ_i is the incident beam angle with respect to the normal of the sample, and λ is the wavelength of the incident light. The *TIS* data obtained from the spectrophotometer measurements and from the BRDF fits are reported in *Table 6.2.3*.[216] It is important to remark that the zone chosen to perform the BRDF fit does not include the convolution-region, being the interest focused on the surface scattering only. However, for the spectrophotometer measurements it has been possible to consider the *TIS* both taking into account and avoiding the specular reflectance (third and fourth column in *Table 6.2.3*, respectively).

Sample	TIS from BDRF data fit	TIS from total reflectance	TIS from diffuse reflectance
a)	0.083 ± 0.009	0.113 ± 0.005	0.095 ± 0.005
b)	0.097 ± 0.008	0.108 ± 0.005	0.091 ± 0.005
c)	0.094 ± 0.009	0.111 ± 0.005	0.092 ± 0.005
d)	0.092 ± 0.009	0.119 ± 0.005	0.099 ± 0.005
e)	0.115 ± 0.010	0.155 ± 0.005	0.136 ± 0.005
f)	0.088 ± 0.009	0.142 ± 0.005	0.113 ± 0.005
g)	0.094 ± 0.009	0.154 ± 0.005	0.118 ± 0.005
h)	0.004 ± 0.002	0.007 ± 0.005	0.006 ± 0.005
i)	0.007 ± 0.002	0.007 ± 0.005	0.006 ± 0.005
j)	0.013 ± 0.002	0.013 ± 0.005	0.011 ± 0.005
k)	0.012 ± 0.002	0.010 ± 0.005	0.009 ± 0.005
l)	0.006 ± 0.002	0.008 ± 0.005	0.007 ± 0.005
m)	0.010 ± 0.002	0.017 ± 0.005	0.014 ± 0.005
n)	0.004 ± 0.002	0.006 ± 0.005	0.005 ± 0.005
o)	0.031 ± 0.003	0.050 ± 0.005	0.042 ± 0.005
p)	0.043 ± 0.003	0.060 ± 0.005	0.047 ± 0.005
q)	0.003 ± 0.002	0.011 ± 0.005	0.005 ± 0.005
r)	0.002 ± 0.002	0.012 ± 0.005	0.005 ± 0.005

Table 6.2.3: TIS values obtained from the BRDF data (second column), from the total (third column) and diffuse (fourth column) reflectance measured by means of spectrophotometer. Samples' names in the first column are the same as the ones reported in Table 6.2.2.

In most of the cases, the *TIS* values from the fit are smaller than the ones obtained by means of total reflectance spectrophotometer measurements. As discussed above, the BRDF data belonging to the convolution-region cannot be fitted; instead, the *TIS* from the spectrophotometer can be calculated considering the diffuse (i.e. not specular) reflectance only. In this case (fourth column in *Table 6.2.3*) the data are slightly closer to the one obtained from the BRDF measurements. The deviations between the *TIS* extrapolated from the BRDF data and the *TIS* calculated using the spectrophotometer data can be imputable to the different experimental setup, in fact the incident angle for the spectrophotometer measurements is equal to 8°, while for the BRDF it is 45°.

The roughness values obtained from the BRDF measurements have been compared with the corresponding values coming from the AFM experiments and the ones coming from the reflectance measurements, yielding remarkable agreement. They can be visually seen in *Fig. 6.2.13* and the *Table 6.2.4*.

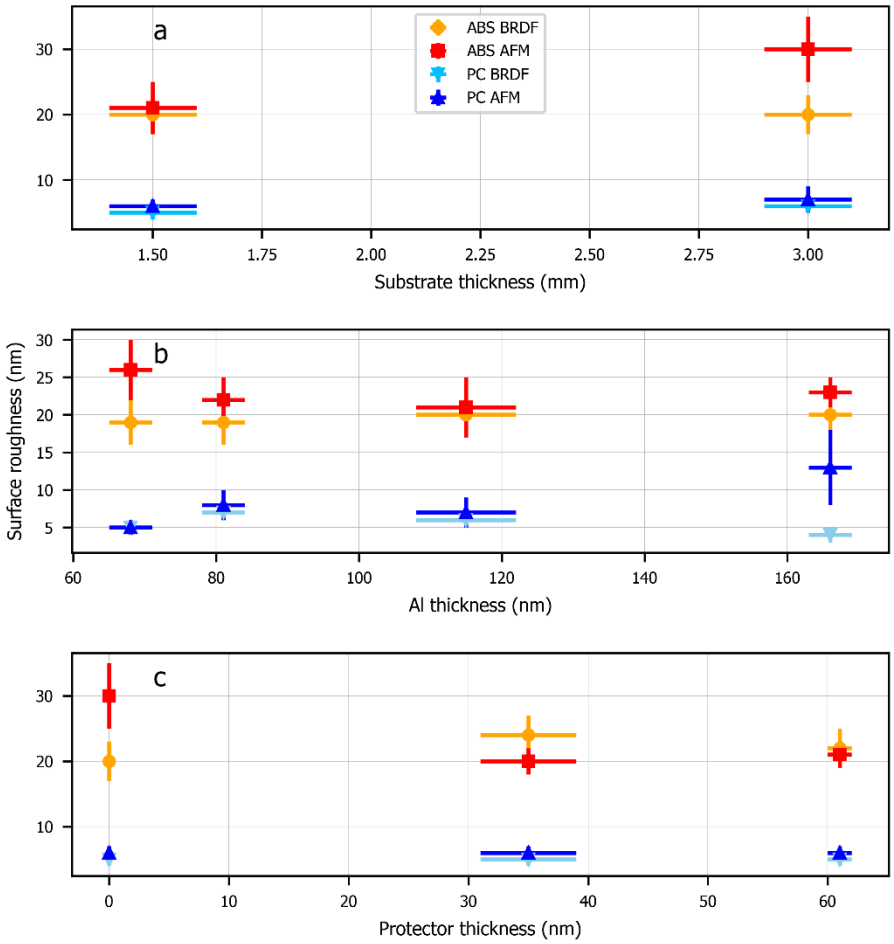

Fig. 6.2.13: root mean square surface roughness extracted from AFM and BRDF measurements for samples prepared on different substrates (PC and ABS): in panel (a), the substrate thickness dependence is represented, in panel (b) the influence of the Al film thickness is showed, while in panel (c) the influence of the HMDSO film thickness is highlighted.

Id.	Substrate	Substrate thickness (mm)	Al thickness (nm)	Protector thickness (nm)	BRDF rms (nm)	AFM rms (nm)	Total reflectance rms (nm)	Diffuse reflectance rms (nm)
a)	ABS	1.5 ± 0.1	115 ± 7	35 ± 4	20 ± 3	21 ± 4	14 ± 1	15 ± 1
b)	ABS	3.0 ± 0.1	68 ± 3	35 ± 4	19 ± 3	26 ± 4	17 ± 2	19 ± 2
c)	ABS	3.0 ± 0.1	81 ± 3	35 ± 4	19 ± 3	22 ± 3	16 ± 2	16 ± 2
d)	ABS	3.0 ± 0.1	115 ± 7	35 ± 4	20 ± 3	30 ± 5	15 ± 2	16 ± 2
e)	ABS	3.0 ± 0.1	115 ± 7	0	24 ± 3	20 ± 2	17 ± 2	18 ± 2
f)	ABS	3.0 ± 0.1	115 ± 3	61 ± 1	22 ± 3	21 ± 2	17 ± 2	18 ± 2
g)	ABS	3.0 ± 0.1	166 ± 3	35 ± 4	20 ± 3	23 ± 3	17 ± 2	19 ± 2
h)	PC	1.5 ± 0.1	115 ± 7	35 ± 4	5 ± 1	6 ± 1	4 ± 1	4 ± 1
i)	PC	3.0 ± 0.1	68 ± 3	35 ± 4	5 ± 1	5 ± 1	4 ± 1	4 ± 1
j)	PC	3.0 ± 0.1	81 ± 3	35 ± 4	7 ± 1	8 ± 2	5 ± 1	5 ± 1
k)	PC	3.0 ± 0.1	115 ± 7	35 ± 4	6 ± 1	7 ± 2	4 ± 1	5 ± 1
l)	PC	3.0 ± 0.1	115 ± 7	0	5 ± 1	6 ± 1	4 ± 1	4 ± 1
m)	PC	3.0 ± 0.1	115 ± 7	61 ± 1	5 ± 1	6 ± 1	5 ± 1	5 ± 1
n)	PC	3.0 ± 0.1	166 ± 3	35 ± 4	4 ± 1	13 ± 2	3 ± 1	4 ± 1
o)	ABS	1.5 ± 0.1	0	0	12 ± 1	14 ± 3	20 ± 2	20 ± 2
p)	ABS	3.0 ± 0.1	0	0	14 ± 2	19 ± 4	21 ± 2	21 ± 2
q)	PC	1.5 ± 0.1	0	0	2 ± 1	3 ± 1	9 ± 1	8 ± 1
r)	PC	3.0 ± 0.1	0	0	2 ± 1	4 ± 1	9 ± 1	9 ± 1

Table 6.2.4: surface roughness values and uncertainties extracted from BRDF, AFM measurements, and total reflectance measurements. As highlighted by the table, the rms surface roughness is primarily determined by the substrate type.

A better match between the rms values extracted from BRDF and AFM measurements with respect to both the reflectance spectrophotometer values (total and diffuse) is evident. This could be ascribable to the different incident angle of the impinging light exploited for the BRDF and spectrophotometer measurements (45° and 8° respectively) or to the deformation of the substrate (stressed in *Fig. 6.2.7*) which could have played a crucial role in the acquisition of the signal with the two different experimental techniques. The investigation on the dependence of the surface roughness on different substrate, Al and HMDSO thickness shows that increasing the film thickness of both the Al and HMDSO layers has no influence on the rms values, while the main difference is in the plastic type used as substrate for the growth of the layers. In particular, the roughness values obtained on ABS-based samples are always greater than the PC-based counterparts, showing that the roughness is mainly determined by the substrate type (PC or ABS).

It has been demonstrated that is therefore possible to determine the surface roughness of the samples by means of a BRDF analysis based on the *ABg* model. Concerning the trends, as reasonably expected, the higher the surface roughness, the lower the specular and the higher the off-specular BRDF values. When, as stated before, this information is associated with the fact that the roughness is mainly influenced by the substrate type and not by the different metallization treatments (*Fig. 6.2.13*), it is possible to conclude that films supported by ABS act more as diffusers, while PC-based films behave more mirror-like. The selection of the substrate material must be defined according to the scope of the reflector: if the purpose of the reflective component is to spatially spread the light over a wider angular interval (without inserting any diffusive element inside the lamp, but only exploiting the optical properties supplied by different substrates), an ABS substrate could be envisaged. If the main goal of the reflector is to focus light in a selected direction, PC should be put forward. Optimal values for

the metallic film thickness were determined in order to maximize the total integrated reflectance (81-115 nm Al). For a thickness around 68 nm the transmission through the Al film affects the reflectance,[217] while for thicker films the accumulation of contaminants in the deposition process may start to play a role, thus decreasing the overall reflectance.[218,219] Concerning the HMDSO, its presence is mandatory to prevent the sample from oxidation, even if it hampers the optical properties of the metal, as shown in the literature.[220] The 35 nm film is more performant than the 61 nm one, while still being sufficient for adequate protection, which has been confirmed by internal validation tests performed inside the company.

To perform a more complete and rigorous analysis, the specular component must be kept into account when performing the fit. For this task, the

6.2.4 Discussion

The experimental measurements and the subsequent data analysis reported in this chapter helped in gaining a deeper insight into the optical properties of Al-coated thermoplastic polymer surfaces. The most important discovery was that the bare ABS substrate is rougher than the PC one and the actual surface roughness of the multilayer coating is dominated by the substrate material. Consequently, metallic films grown on ABS act more as diffusers, while PC substrates yield mirror-like characteristics. In an applicative view, these differences already play a relevant role in the choice of the proper substrate material for each element of the reflective part of a car head or rear lamp. As a function of the different metallization layers, the BRDF analysis revealed small spectral and angular distribution differences, so that it has been possible to optimize the metallization process parameters. More specifically, an Al coating of 80-115 nm optimizes the surface reflectance. Regarding the protective HMDSO layer, a 35 nm thick film proved to be the best solution to achieve the desired optical properties and, simultaneously, it prevents both surface oxidation and possible mechanical damage. Thicker coatings reduce the reflectance, affecting the optical performance of the device. The BRDF and total reflectance analysis performed within the *ABg* framework provided the retrieval of quantitative parameters that will allow modeling of the multilayer films in a simulation software to design and engineer automotive lighting devices. The advantages of the BRDF measurements over the total reflectance ones can be found in a more detailed geometrical description of the off-specular scattered component of the reflected beam, which turns out to be a key-factor for the ray-tracer simulations. In fact, by means of the measurements described in this section new and more precise models have been introduced in the ray-tracing simulators, to be used by optical engineers to better simulate the different metallization treatments.

6.3 Automatic quality check for reflectors

Automatic inspection is a fundamental task in a production process in order to increase quality standards and reduce associated time and costs. As already discussed in *Chapter 5*, human supervision can be inefficient. In addition, in the specific case, if not promptly detected, a defective component could be assembled on a lamp and could lead to discard the final product, together with all its compliant components. To reduce the possibilities of error, an automated computer-based inspection system yielding a greater reliability of the quality check results is required. In this section, the design and validation of an innovative and unique machine, able to both recognize the defects arising on the reflectors' surfaces and to check the reflective properties of the metal thin film is presented. The double aim of the reflectors, i.e. giving the lamp the desired aesthetic aspect and supplying the precise optical properties demanded by the photometric specifications, requires a precise and thorough quality inspection. The inspection system has been specifically tailored for the reflectors because, in addition to the defects detection, it has been designed to check the local reflectance properties of the aluminum layer: in fact, the geometric shape of these components, characterized by paraboloidal cavities, together with finer superstructures exploited to focus light, makes a standard reflectance measurement impossible to perform. Indeed, the most common experimental instruments such as spectrophotometers, reflectometers or ellipsometers need precise geometries and possibly flat samples.[221–223] Moreover, such techniques are usually cost and time demanding, making it impossible to check every single component with such methodologies. The prototype machine that will be presented in this chapter can, instead, perform a fast and reliable investigation to determine if the local and global reflective properties of the deposited Al coating satisfy the quality requirements by means of an optical inspection. Moreover, exploiting Support Vector Machines (SVMs) supervised learning models, the prototype tool is able to recognize autonomously the defects arising on the reflectors and classify them.[224,225] The role of the machine learning models – the SVM in the present case – allows the prototype to progressively fine tune its precision with every scan, leading to a satisfactory detection accuracy for both the structural and optical defects.

6.3.1 Why optical properties?

As stated above, the complex geometries of the reflectors make it impossible to adopt standard reflectance measurements setups. BRDF measurements, as discussed in *Section 6.2*, have been performed on almost flat samples, to be able to contain the angular spreading of the reflected light. Some other methodologies to determine the optical properties of the metal-polymer heterostacks have been investigated and have been classified as inadequate. One approach (intensively exploited by many packaging companies) involves instead the evaluation of the electrical conductivity of the metal layer measured by sheet resistance.[226] In brief, sheet resistance consists in a measure of the resistance of thin films uniform in thickness. Knowing the geometrical positions of the two pins used to generate a potential difference and of the two exploited to measure the current intensity, it is possible to determine the resistance of the layer. Thanks to tabulated values of resistivity (reported in the literature for several elements)[227] it is possible to extrapolate the thickness of the film. This approach demonstrates some weaknesses: the aluminum layer deposited over a polymeric template has a thickness in the nm range, which is not the correct operation range for such technique (best operating range ~ μm). Moreover,

the technique implies a uniform contact of the pins with the sample, and this is extremely difficult to guarantee when the samples are not flat. Lastly, the thickness of a thin layer is not to the only parameters that determines the actual optical properties of the film itself. In fact, chemical composition (contamination) also contributes. In *Table 6.3.1*, the comparison between Atomic Force Microscopy and Sheet resistance results is reported for the present case: four flat samples with different thickness values have been measured and no agreement can be seen from the results.

AFM thickness (nm)	Sheet resistance thickness (nm)
68 ± 3	40 ± 10
81 ± 3	50 ± 11
115 ± 7	80 ± 15
166 ± 3	120 ± 25

Table 6.3.1: measurements of the thicknesses of four different samples by means of AFM (left column) and sheet resistance (right column). No agreement is evidenced.

Therefore, an innovative approach to check the optical properties of the reflectors exploiting an optical-based-system has been implemented. Using a luminance camera (LMK 5 by Technoteam, example of luminance image in *Fig. 6.3.1*), it is possible to highlight the differences resulting from different metallization treatments.

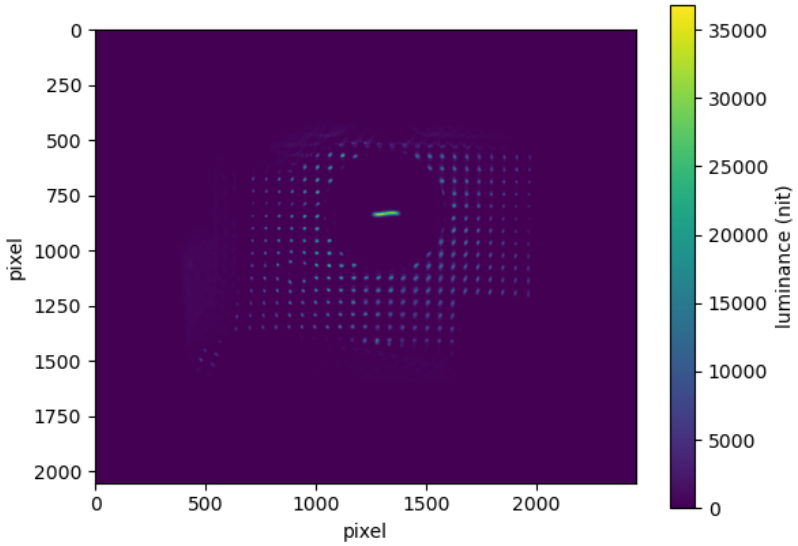

Fig. 6.3.1: example of luminance image of reflector and light source. The color bar on the right expresses the luminance value of each pixel (in nit).

This method does not give information regarding the absolute reflectance value, while once a correct threshold has been determined it yields a Boolean discrimination factor, thus permitting to distinguish compliant and non-compliant reflectors. Six different recipes have been considered to deposit the aluminum on the plastic substrate: they differ only in the numbers of aluminum spirals evaporated, thus yielding layers with different Al thickness. A recap of the recipes and the relative number of Al spirals used is reported in *Table 6.3.2*.

Cycle number	Aluminum spirals used
1	2
2	5
3	8
4	11
5	14
6	17

Table 6.3.2: cycle number (left column) and total Al spirals used in each cycle (right column).

Together with the quality check team, the reflectors metallized exploiting the different cycles have been classified as compliant or non-compliant by means of visual inspection: cycle 1 and 2 yield non-compliant coatings. In *Fig. 6.3.2* it is possible to visualize these differences: *Fig. 6.3.2a* represents the luminance image coming from a reflector metallized with "cycle 1" treatment, while the luminance image reported in *Fig. 6.3.2b* is obtained using 17 Al spirals (cycle 6) during the evaporation process. They both show the luminance images of the reflector with the subtraction of the bulb filament contribution: to perform a correct analysis, in fact, the light source must be excluded to focus the attention on the reflector only. The color bar on the right of the figure is essential to comprehend the different outputs.

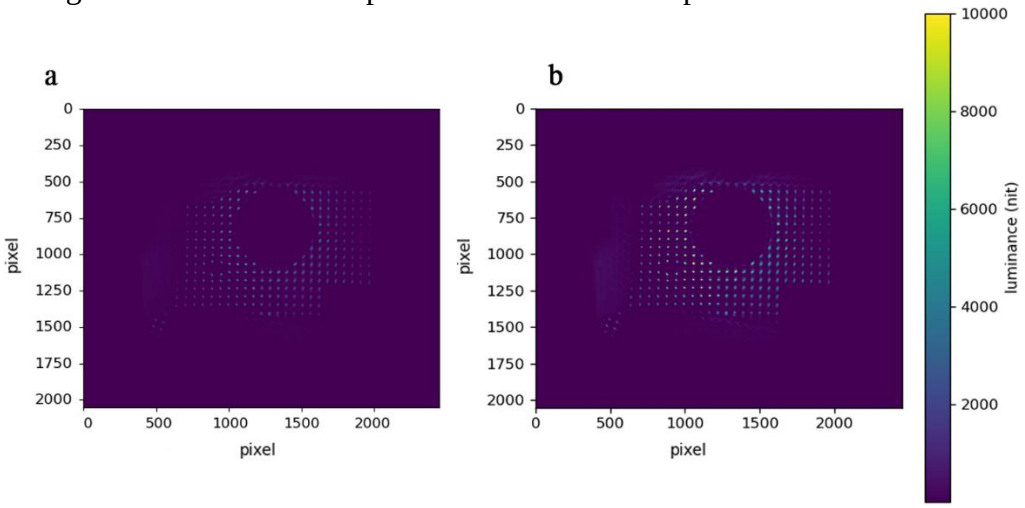

Fig. 6.3.2: two luminance images of different reflectors. Panel (a) represents a reflector metallized using cycle number 1, while (b) one metallized by means of cycle 6.

The introduction of an opal material positioned between the reflector and the camera is useful to gain information regarding the overall reflective properties of the samples. Clearly, the local information is lost by the introduction of the slab, which homogenously scatters light. It is therefore impossible in this latter case to disentangle the intensity contributions ascribable to the bulb from the one coming from the reflector. On the other side, the total integration of the signal is a key-information to unravel the spatial distribution of light and therefore detect possible failures in the manufacturing or coating process. Moreover, the introduction of the diffusive material supplies a fast and robust chance to check the correct placement of the reflector on the holder (the control performed also by checking the pins positions) (*Fig. 6.3.3*) without having saturation problems.

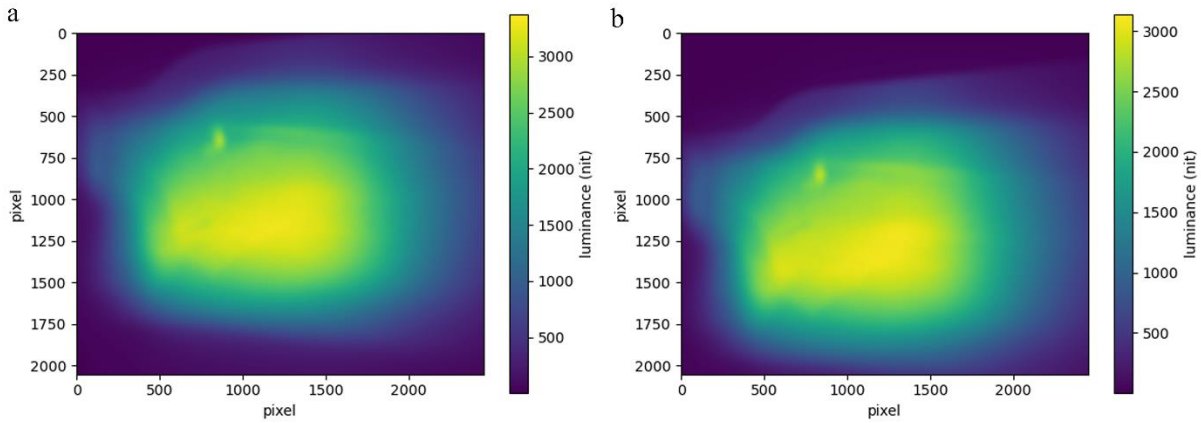

Fig. 6.3.3: visual representation of bad positioning: (a) right positioning of reflector, (b) same reflector rotated with an angle of 2.5° with respect to the normal direction. The difference is evidenced by a 300 pixel shift of the image in the vertical direction.

Unluckily, the luminance camera is an expensive and precious object that cannot be installed on a prototype machine, placed in the productive chain to scan a high number of components. However, similar information can be achieved by means of a standard camera: exploiting a photopic filter and acquiring multiple images at different integration times, it is possible to mimic the functionality of the luminance camera and to become sensitive to the extremely wide range of luminance values that can arise at different locations on the reflector. The photopic filter (transmission spectrum depicted in *Fig. 6.3.4*) is necessary to simulate the human light perception and, therefore, replicate the luminance camera working principle.

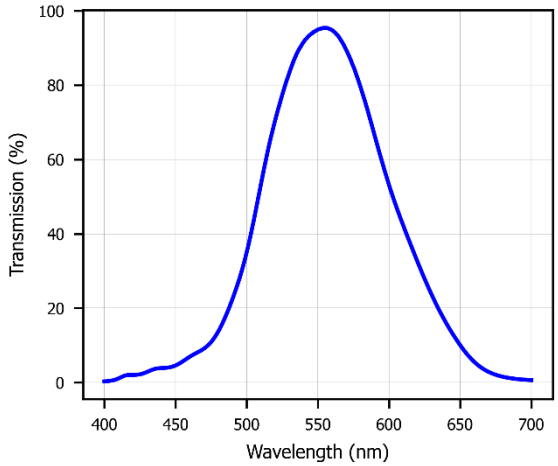

Fig. 6.3.4: characteristic transmission spectrum of the photopic filter.

The results are highlighted in *Fig. 6.3.5*, where a comparison between data acquired by luminance camera (*Fig. 6.3.5a*) and standard camera (*Fig. 6.3.5b*) are represented. The trend of the two set of data is similar and in both it has been possible to set a threshold in order to distinguish compliant components from defective ones. As stated above, cycle 1 and 2 produce scrap components, while the other four cycles lead to a correct Al coverage.

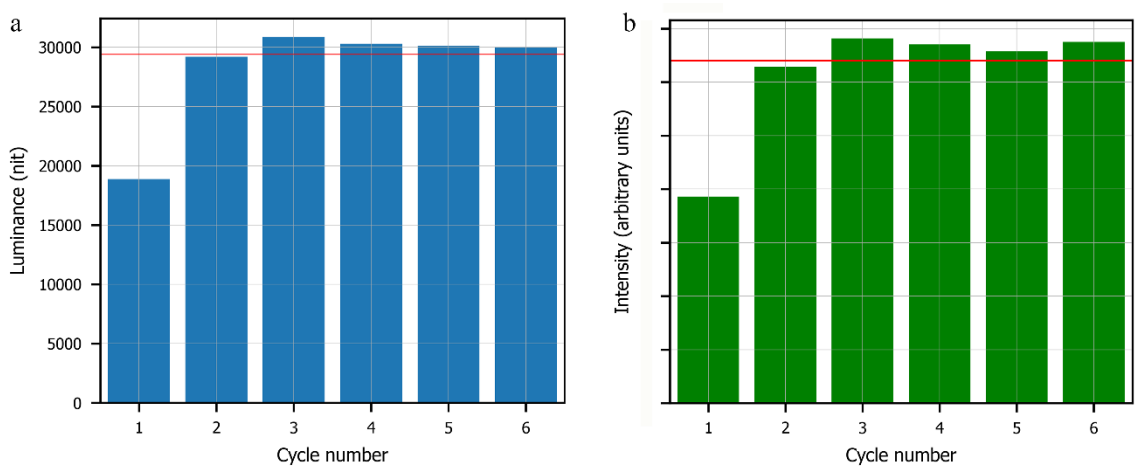

Fig. 6.3.5: comparison between luminance camera (a) and standard camera (b) signals. The red line is the threshold distinguishing compliant and non-compliant reflectors; it has been applied in collaboration with quality operators.

Moreover, by means of this prototype machine it will be possible to give a real time feedback regarding some process failures, affecting both the injection molding process and the coatings deposition. In fact, the SVM-based machine decreases the detection time required for each component and subsequently performs the analysis, allowing to have a prompt response, which can be supplied to technicians to fine tune the metallization and molding processes parameters involved in the production chain.

6.3.2 Setup of the machine

The machine structure can be described as a 1.5 m high parallelepiped having a square basis (side = 1.2 m): a sketch can be observed in *Fig. 6.3.6*, representing the isometric view (panel *a*), the lateral (*b*), top (*c*), and front (*d*) views.

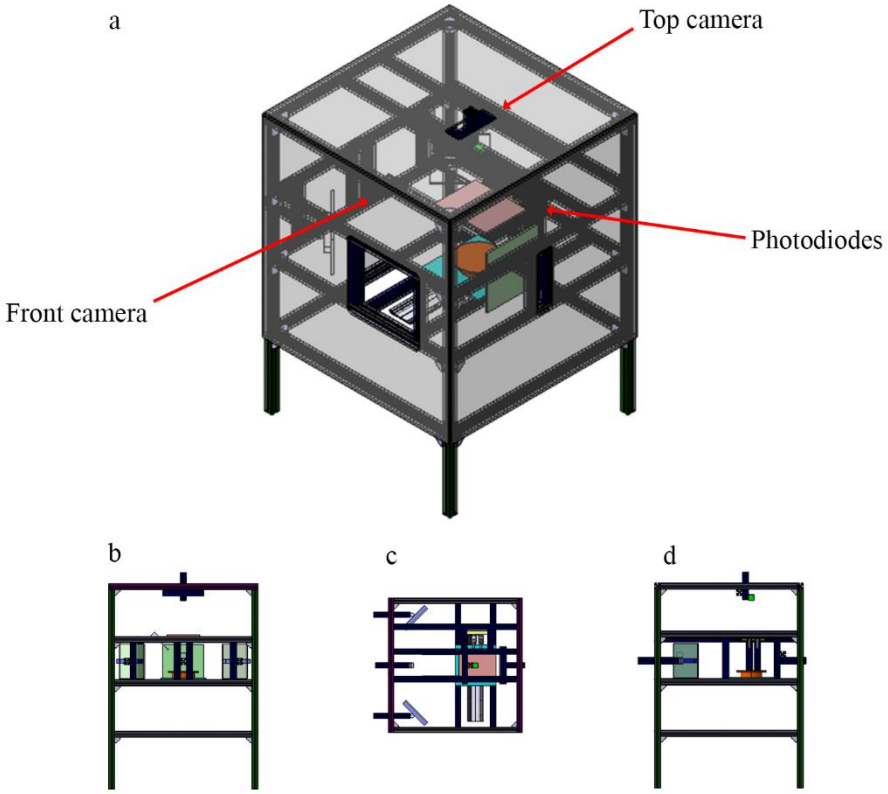

Fig. 6.3.6: CAD model of the prototype machine. Isometric (a), lateral (b), top (c), and front (d) views.

The reflector is placed on a rotating footboard in the middle of the structure (*Fig. 6.3.6*). Two GigE "mvBlueCOUGAR-X104ic" cameras, with resolution of 2464x2056 pixels acquiring greyscale images, are installed: the top camera is equipped with a photopic filter and is dedicated to the "luminance-like" images acquisitions. No illuminator is needed to evaluate the reflectance of the sample because the bulb lamp (spectrum in *Fig. 6.3.7*) placed inside the housing is exploited. Anyway, for the structural and color evaluations the illuminators are still needed.

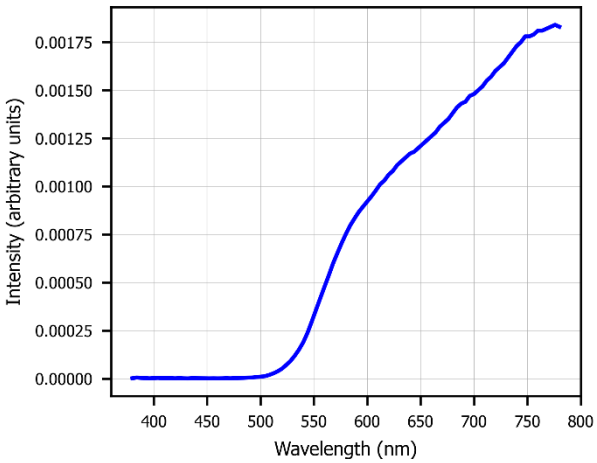

Fig. 6.3.7: emission spectrum of the light bulb mounted on the chosen reflector.

The aim of the front camera is to scan the components looking for structural defects: standard 8-bits greyscale images are acquired at different incident angles, exploiting the rotating

board. An array of photodiodes, coupled with a diffractive prism, is exploited to evaluate the color of the metallization treatments performed on the reflector. Four illuminators are alternately turned on to shine light from precise directions to highlight possible defects. One (on the opposite side of the front camera) helps detecting the incompleteness, two (on either side of the front camera) illuminate from different directions the reflectors to better detect flaws on its surface, and the last one (located below the photodiode) shine light to perform the color evaluation. A picture representing the machine is reported in *Fig. 6.3.8*.

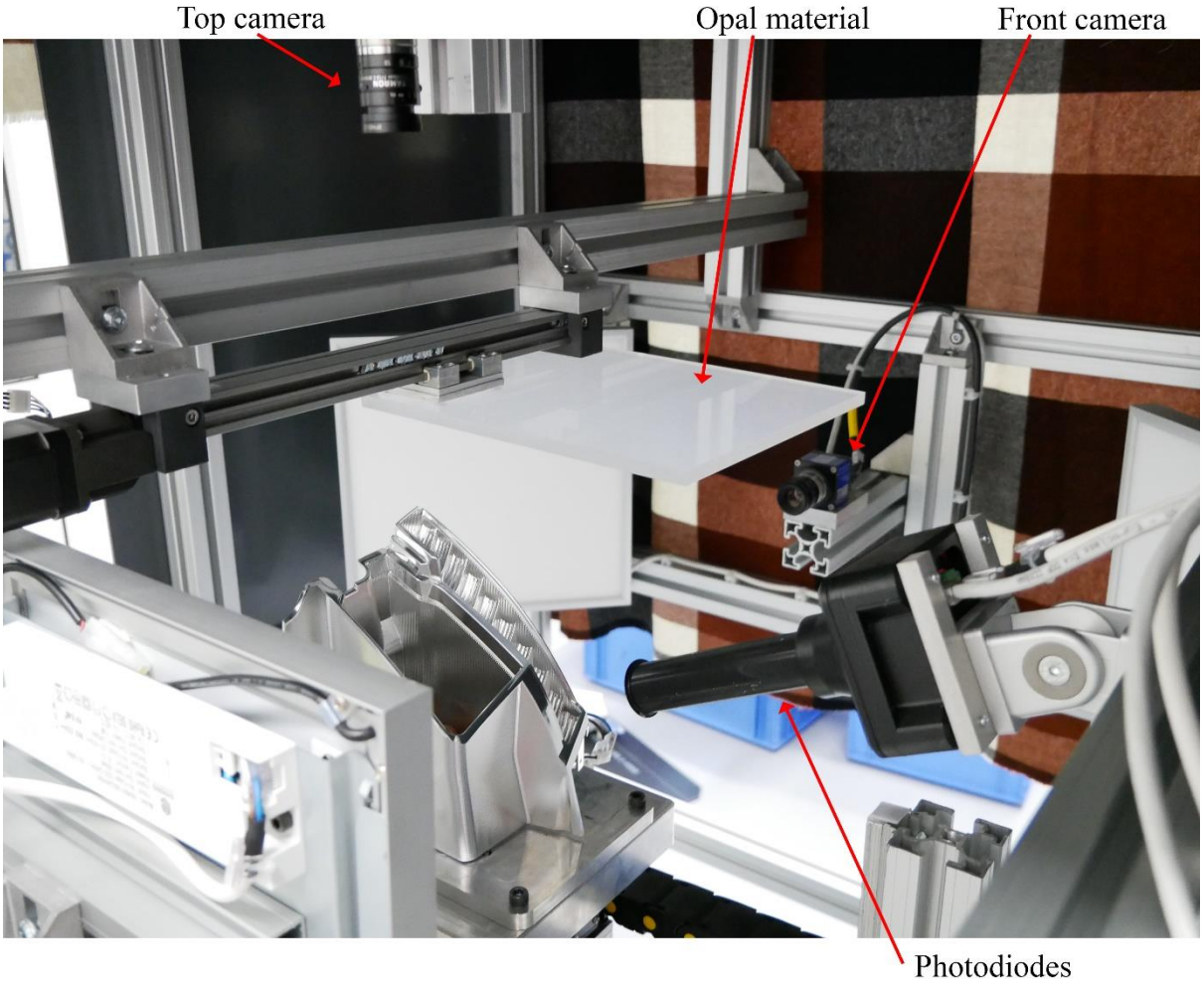

Fig. 6.3.8: prototype machine picture, representing the front camera, the top camera, the photodiode and the opal slab. The reflector is placed on a rotating footboard to be correctly scanned.

The design of a software architecture (coded in C#) composed by several algorithmic modules to correctly perform the optical and structural evaluations has been developed. The main aspects that must be accommodated to preserve the functional characteristics of the machine are the following:

- Configurability: the software must allow for an easy adaptation to different reflectors models.
- Objectivity of the measurements: the software must associate to each sample some univocal quantities (luminance, color, dimensions) which must be reproducible and easily interpretable.

- Trained classification and prediction: the system must be able to find the correct parameters to predict the presence of future anomalies coherently with the instructions given by the operators.

To satisfy all these requirements, a three-layers-architecture has been designed, composed by:

- Acquisition layer: constituted by modules, sensors, illuminators, and motors, it defines in a standard and repeatable way the operations to be performed during the acquisition process. These operations generate raw files of different nature as for example images or text files.
- Analysis layer: its aim is to analyze the heterogeneous data generated by the acquisition layer and provide, as output, series of numbers (features) which represent the main characteristics of the item scanned. All these features represent the whole description of the item which must be compared to the reference in order to find some defects.
- Applicative layer: it is made by three main processes:
  A) Training: the users can set the system defining the defects' classes and associating them a severity. A measurement campaign has to be performed, manually labelling all the items in order to teach the machine how to recognize the different defects.
  B) Prediction: the predictive models designed can be implemented to scan each reflector, indicating the presence and the location of defects.
  C) Validation: allows the users to evaluate the predictive performance of the system on a new dataset, where both defective and compliant components can be scanned and classified. The overall evaluation metrics (number of defective components, false positives, false negatives, …) are given.

The abovementioned details are summarized in *Fig. 6.3.9*.

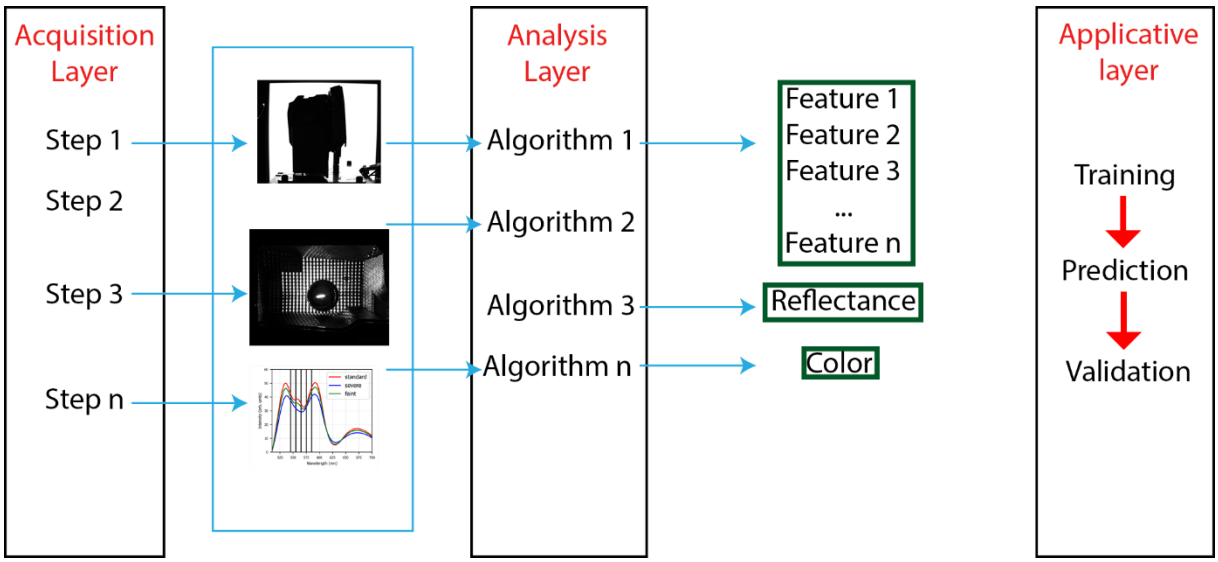

Fig. 6.3.9: schematic representation of the three-layer architecture designed for the prototype machine.

6.3.3 Working principle

The main defects arising on reflectors ascribable to the injection molding process (incompleteness, flash) or to the deposition process (color) are depicted in *Fig. 6.3.10*. An incorrect filling of the mold results in an incompleteness, which is evident for the farthest region with respect to the injection point (shape incompleteness – *Fig. 6.3.10a* and pin incompleteness – *Fig. 6.3.10b*). If the temperature of the melt PC is not high enough, some "scars" will be visible on the reflector's aspect. This defect is denominated flash and is represented in *Fig. 6.3.10c*. Finally, if the metal evaporation rate is too high, the metal layer acquires a yellowish aspect, not shiny enough for the correct optical performances to be fulfilled (*Fig. 6.3.10d*).

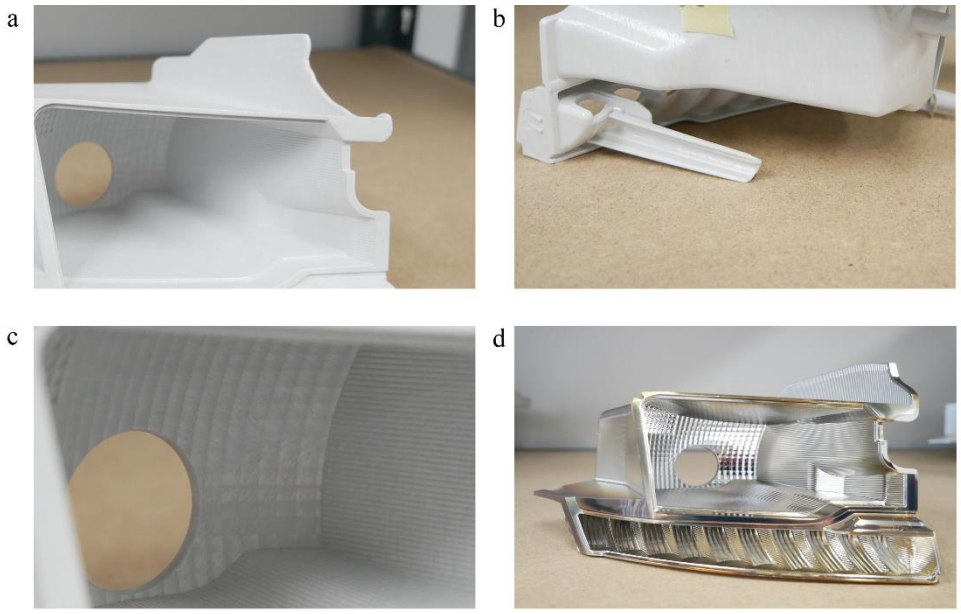

Figure 6.3.10: images showing defects appearing on the reflectors: shape incompleteness (a), pin incompleteness (b), trace (c) and yellowish (d).

A training, aiming on building precise classification models, has been performed. To create the models, 106 components, as a training set, have been scanned and peculiar features have been extracted. During this process, all the scanned reflectors have been manually labelled as compliant/non-compliant. Based on the training, for each defect we have selected the more suitable features to describe it. These features are used, by means of SVM, to generate a model allowing for an automatic classification of the components.

Component Scanning

During the scanning process, the rotating footboard, the illuminators, the cameras and the photodiodes are synchronized to perform a fast data acquisition. During a 180° rotation of the footboard, 10 images are acquired to perform the shape and pins incompleteness evaluation and to look for potential flashes. In addition, color spectrum is acquired exploiting the photodiode and 10 images taken using different integration times (20, 40, 80, 160, 320, 640, 1280, 2560, 5120, 10240 μs) are collected by the top camera to perform the reflectance evaluation. Lastly, the opal slab slides between the reflector and the top camera to take additional 5 images (integration times: 1500, 3000, 4500, 6000, 7500 μs). The whole acquisition process lasts 72 seconds, with most of the time dedicated in footboard rotations and slab sliding.

Upgrading the moving motors, it will be possible to reduce the acquisition time below 10 seconds, thus faster than the human check in the production line (12 seconds).

Features extraction

The heart of the detection process is the extraction of features. The features, which are translated into quantitative values, are crucial for the SVM learning models and for the subsequent classification of the components. All the extracted features can be grouped in five big families, involving the five main defects typologies under examination:

a) Color
b) Reflectance
c) Trace
d) Shape incompleteness
e) Pin incompleteness

Color

The color evaluation is performed acquiring the signal emitted by the component in the visible region by means of the photodiode installed in the machine.

Fig. 6.3.11: example of yellowish. The sample on the left is compliant, while the middle (faint yellowish) and the right (severe yellowish) ones are scraps.

When not correctly metallized, reflectors tend to change their aesthetic appearance, they lose their brilliant and gleamy aspect and exhibit a yellowish color (*Fig. 6.3.11*). Five intensity values are measured, acquired at 545, 555, 565, 575, 585 nm (yellow region of the visible spectrum) to perform the distinction between compliant and non-compliant reflectors. The spectra extracted from the analysis of the three reflectors in *Fig. 6.3.11* are reported in *Fig. 6.3.12*. They are represented with different colors: red for the sample with a standard coverage of Al, green for the sample with a faint lack of coverage while blue for the sample with a severe absence of Al on the plastic surface. The spectra have been measured in the 500-700 nm range. In *Fig. 6.3.12* there are 5 black vertical lines highlighting the wavelengths at which the analysis is performed. From the graph, a clear difference between the three spectra is evincible: in particular, the intensity signal is lower for samples with more pronounced yellow aspect.

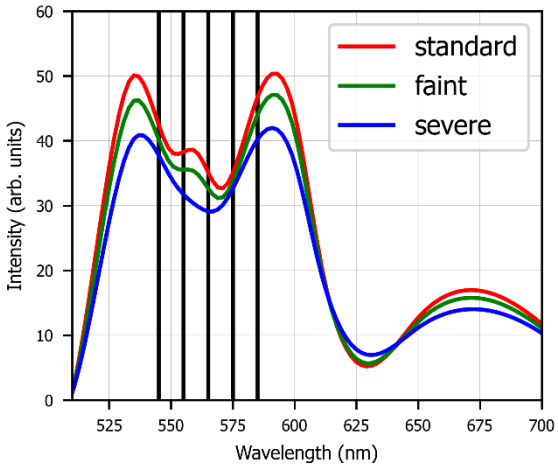

Fig. 6.3.12: reflectance spectra of the three samples showed in picture 6.3.11 (red – compliant reflector, green and blue – defective components). The vertical lines represent the wavelengths at which the features are extracted.

Reflectance

As described in the previous section, a standard camera is exploited to check the reflectance properties of the coated components, acquiring several images with different integration time (IT). The role of the photopic filter is essential to collect data which are similar to the luminance ones. An example of a poorly coated reflector is visible on the right side of *Fig. 6.3.13*.

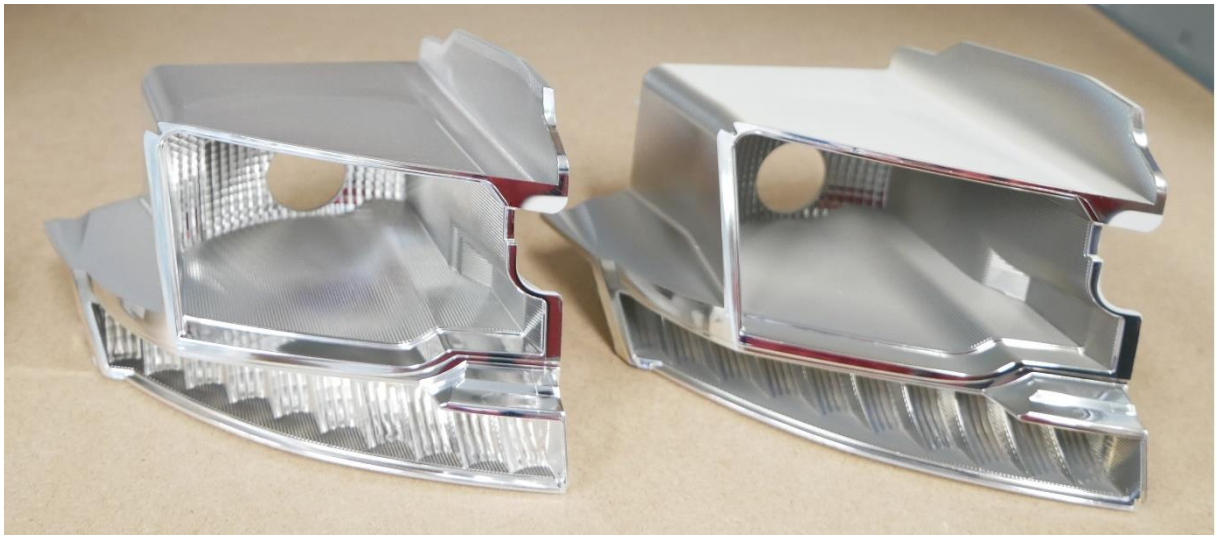

Fig. 6.3.13: example of different metallization treatments on different reflectors: one with correct aluminum thickness (left) and one with insufficient Al on top (right).

The variation of the IT is fundamental to be sensitive to different regions of the reflector: lower ITs permit to better detect shiny regions, while higher ITs allow for a better analysis of darker zones, i.e. zones with low Al coverage. To compare images acquired with different ITs, the value of every pixel is "normalized" to a common 10 ms integration time (t_{common}) by multiplying each pixel intensity for the ratio between the common exposure time and the measured one. For example, for an image acquired using an integration time equals to 2560 μs

(i.e. $t_{measure}$ = 2560 μs), the single pixel intensity calculation to obtain the common intensity value (I_{common}) will be

$$I_{common} = I_{measure} \frac{t_{common}}{t_{measure}} \tag{6.3.1}$$

This normalization procedure is performed for each pixel of every collected image. The images are then merged by taking the pixel-wise maximum and an output image is generated with the same size of the original ones (2464 x 2056 pixels). The resulting grayscale images can have intensity values greater than 255, therefore a 16-bit encoding is used to extend the range of possible value. If, for a precise integration time, the pixel intensity is saturated, the algorithm automatically considers the acquisition with less integration time and performs again the saturation check. Five features are extracted on the resulting images:

a) Maximum pixel value intensity I_{MAX}
b) Horizontal pixel position of the maximum x_{MAX}
c) Vertical pixel position of the maximum y_{MAX}
d) Total intensity (i.e. sum of the single pixels intensities) I_{TOT}
e) Total intensity using predefined masks which exclude the light source and the lateral faces of the reflector $\tilde{I}_{TOT}$

The same process described above has been applied to the images taken with the opal material interposed between the reflector and the camera. In this case, the local properties of the sample are lost while more information regarding the global reflective properties and the positioning accuracy of the component on its holder are earned. Obviously, only 4 features are extracted: in fact, the introduction of the opal causes the uncontrolled refraction of light and the subsequent inability to separate the signal component ascribable to the reflector from the one attributable to the light source. In *Fig. 6.3.14*, the images involving two samples showing non-compliant (*a*-*e* and *k*-*o*) and compliant (*f*-*j* and *p*-*t*) reflectance properties, taken with increasing integration time, are represented. As perceivable from the figure, incrementing the integration time leads to a larger recorded intensity. However, the algorithm abovementioned is fundamental to avoid saturation (as the case represented in panel *t*) which may bring to incorrect conclusions.

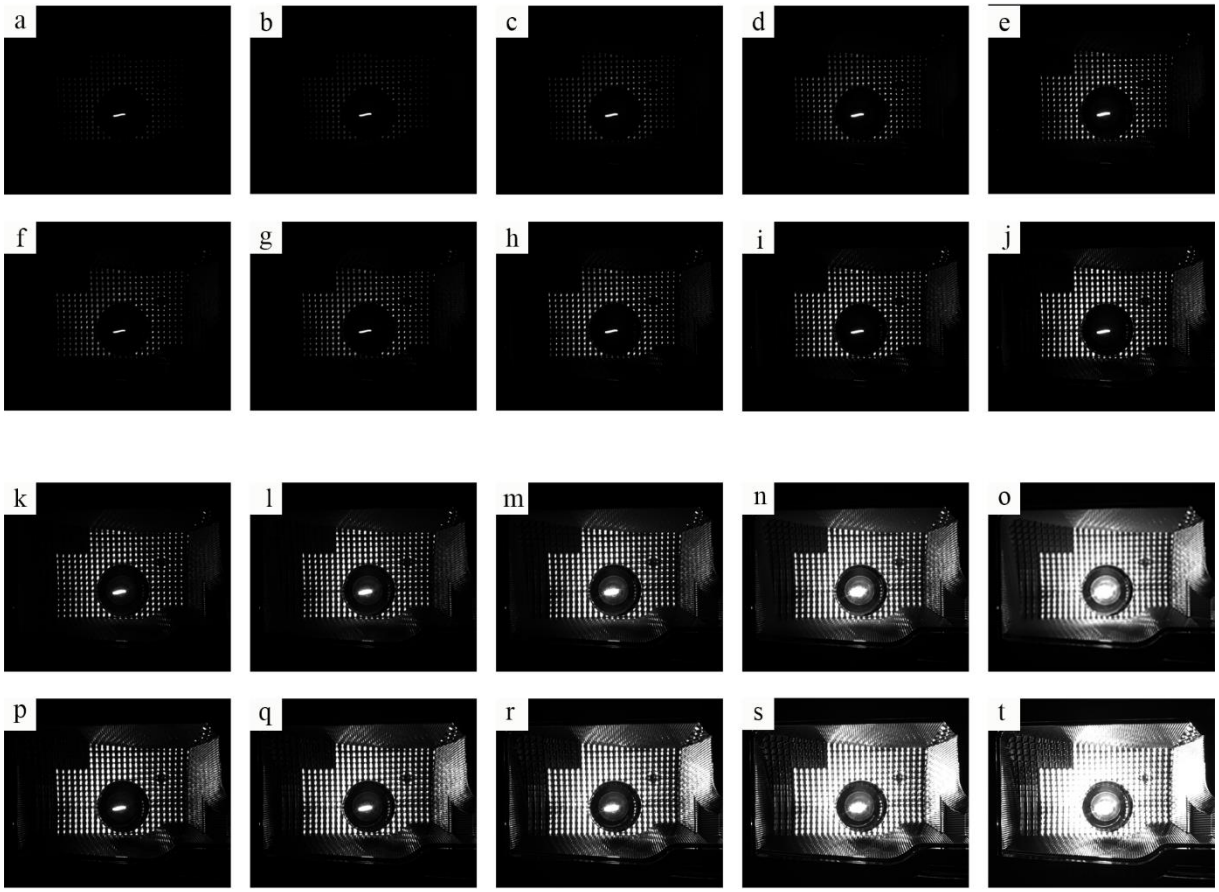

Fig. 6.3.14: 20 images representing acquisitions performed with different integration times. The IT increases from 20 (a and f) to 10240 (o and t) μs by doubling in every image. Panels (a-e) and (k-o) are representative of the cycle 1, while images (f-j) and (p-t) are showing cycle 5. Differences are clearly visible.

Flash

Flashes can arise in many different locations on the reflector. In addition, their size range is wide: some of them cover the whole reflector (approximately 15 cm), while others are just some fractions of mm long. To be more sensible to the correct location of the flashes, the whole reflector surface has been divided in 6 regions (represented in *Fig. 6.3.15*) and a different model has been trained for each of them. Moreover, every region has been subsequently divided in sub-regions, denominated zones, in order to increase the sensitivity of the feature extraction process. As an example, it is possible to consider the upper part of the reflector, the region labelled with number 2 in *Fig. 6.3.15*.

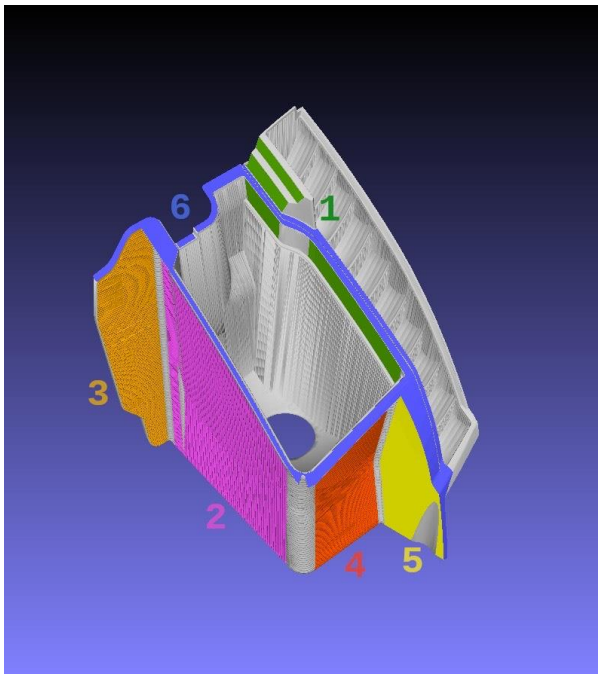

Fig. 6.3.15: enumeration of different reflector's regions.

The entire surface of region 2 has been divided in 25 zones, and for each zone the algorithm extracts the same variety of texture features. Two sample zones are depicted in *Fig. 6.3.16.*

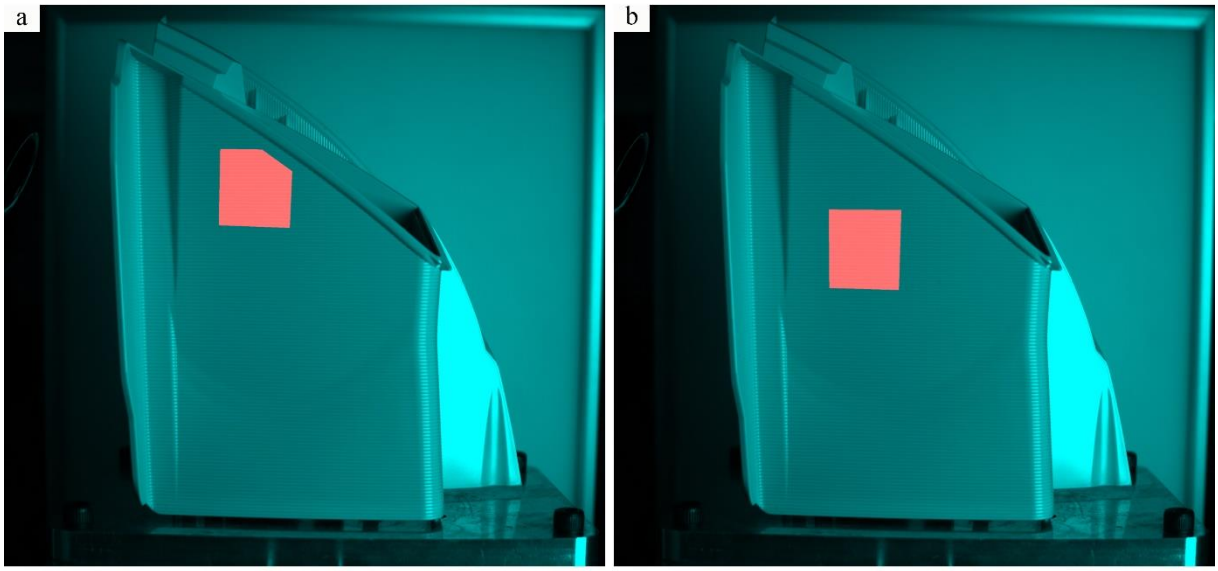

Fig. 6.3.16: representation of two zones in which the region 2 has been divided.

In the specific, two main groups of quantities have been utilized: First Order Statistics (FOS),[228–230] that concerns the distribution of gray level intensities, and Gray Level Co-occurrence Matrix (GLCM),[231–233] which takes information about spatial relation between the gray-level intensity values. 19 FOS features have been implemented, and one of the most suitable to detect anomalies on the selected region was found to be the so-called Entropy

$$entropy = -\sum_{i=1}^{N} p(i) log_2[p(i) + \epsilon], \tag{6.3.2}$$

where N is the number of greyscale levels in the selected region, $p(.)$ the normalized first order histogram and ϵ an arbitrarily small positive number.[234,235] This quantity specifies the randomness in the image pixels intensities, it therefore measures the average amount of information required to encode the image values. A GLCM of size N × N describes the second-order joint probability function of an image region constrained by the mask and is defined as $P(i, j|\delta, \theta)$. The $(i, j)^{th}$ element of this matrix represents the number of times the combination of levels i and j occur in two pixels in the image, that are separated by a distance of δ pixels along angle θ (δ defined according to the infinity norm). μ_x is the mean gray level intensity of p_x (which is the normalized first order histogram for the x-component of the image) and is defined as $\mu_x = \sum_{i=1}^{N} p_x(i) i$ while μ_y is the mean gray level intensity of p_y and is defined as $\mu_y = \sum_{j=1}^{N} p_y(j) j$. σ_x and σ_y are the standard deviation of μ_x and μ_y, respectively. A GLCM feature that revealed to well distinguish defective items from compliant ones is the Correlation, a value ranging between -1 (inversely correlated) and 1 (perfectly correlated) which shows the linear dependence of gray level values to their respective pixels in the GLCM (correlation equals to 0 means pixels are uncorrelated)[236,237]

$$correlation = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} p(i, j) ij - \mu_x \mu_y}{\sigma_x(i) \sigma_y(j)}. \tag{6.3.3}$$

Apart from the correlation, other 24 features are extracted on every zone of each reflector. In *Fig. 6.3.17* are represented the differences between a compliant component (*d-f* and *j-l*) and a non-compliant one (*a-c* and *g-i*) due to the present of an extended flash. The rotation of the footboard permits to correctly shine light on the different regions of the reflector and therefore to perform the analysis in the best conditions. Exploiting the different reflection induced by the flash, by means of the abovementioned features it is possible to detect and classify the blemish.

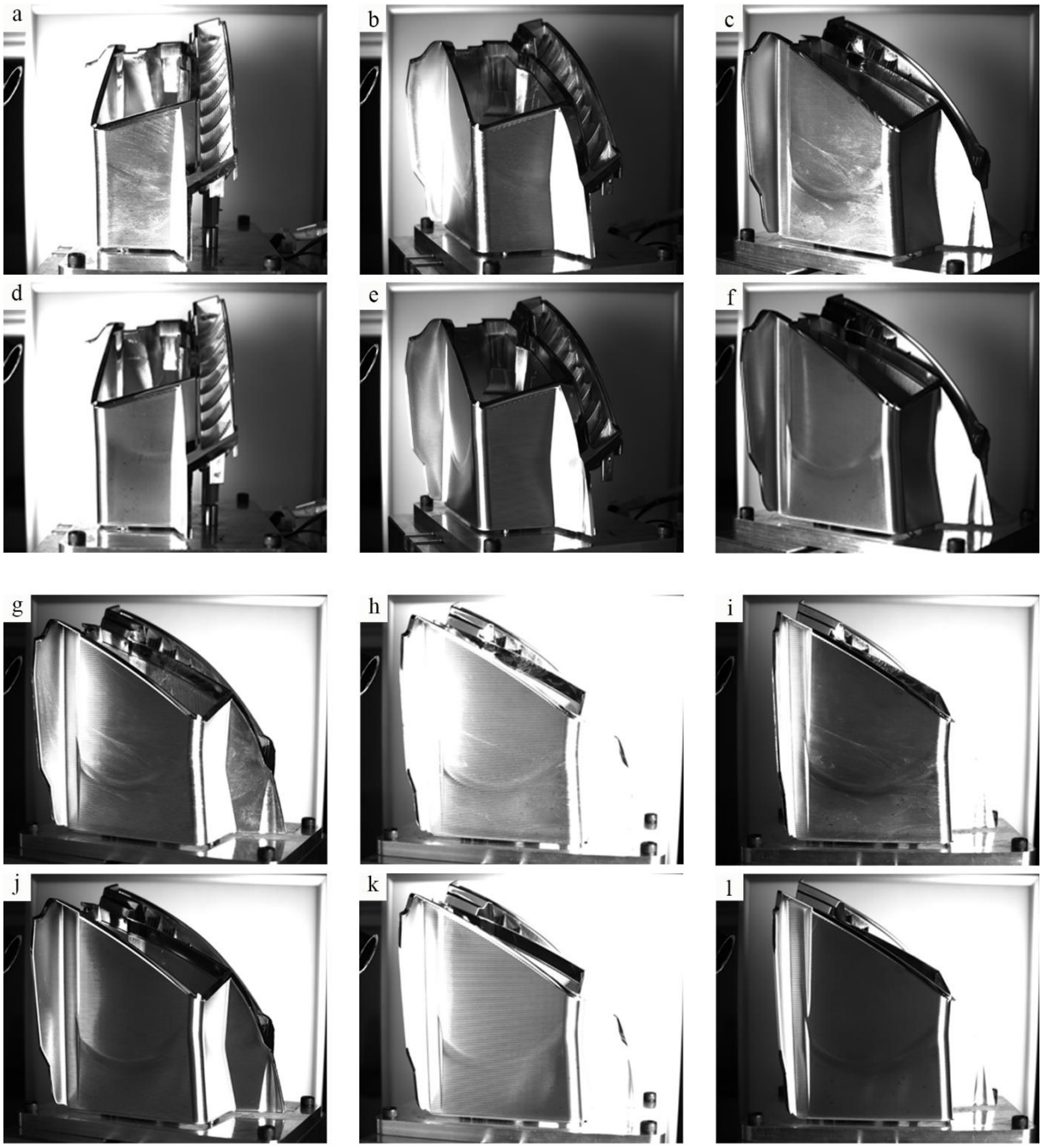

Fig. 6.3.17: images representing the acquisitions performed to detect the trace defect. A compliant component (d-f and j-l) and a non-compliant piece (a-c and g-i) are showed.

Shape incompleteness

Regions furthest to the injection point are filled lastly by the liquid PC inside the mold. This aspect is extremely important to easily determine the incompleteness defects locations on the reflectors. Two regions, highlighted by the red circles in *Fig. 6.3.18*, are the ones showing the concerning defects more frequently.

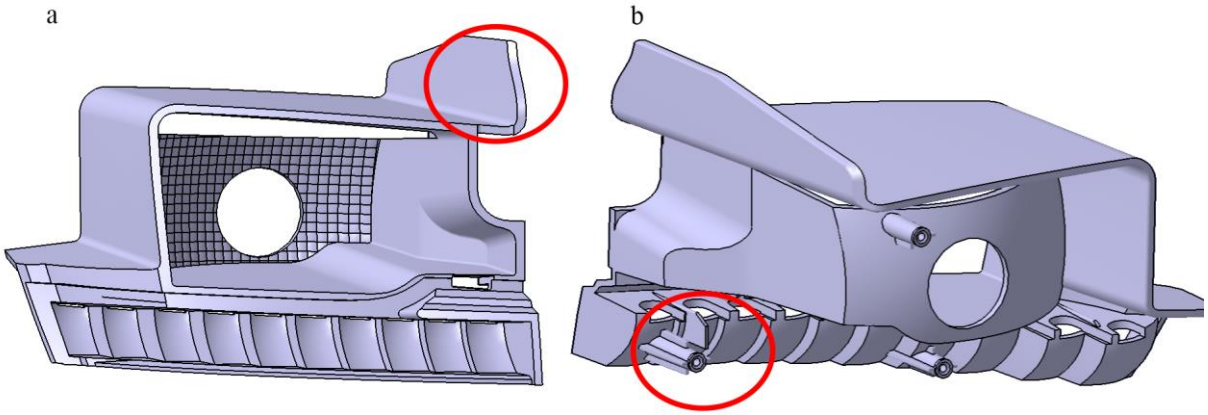

Fig. 6.3.18: CAD model illustrating the regions where the shape (a) and pin incompleteness (b) may manifest.

In this section the attention will be focused on the shape incompleteness, represented in *Fig. 6.3.18a*, while in the next section the pin incompleteness (*Fig. 6.3.18b*) will be treated. A real example of shape incompleteness is represented in *Fig. 6.3.19*. Here, one compliant reflector (panel *a*) is compared to two scrap components, showing a small (*b*) and a severe (*c*) lack of material.

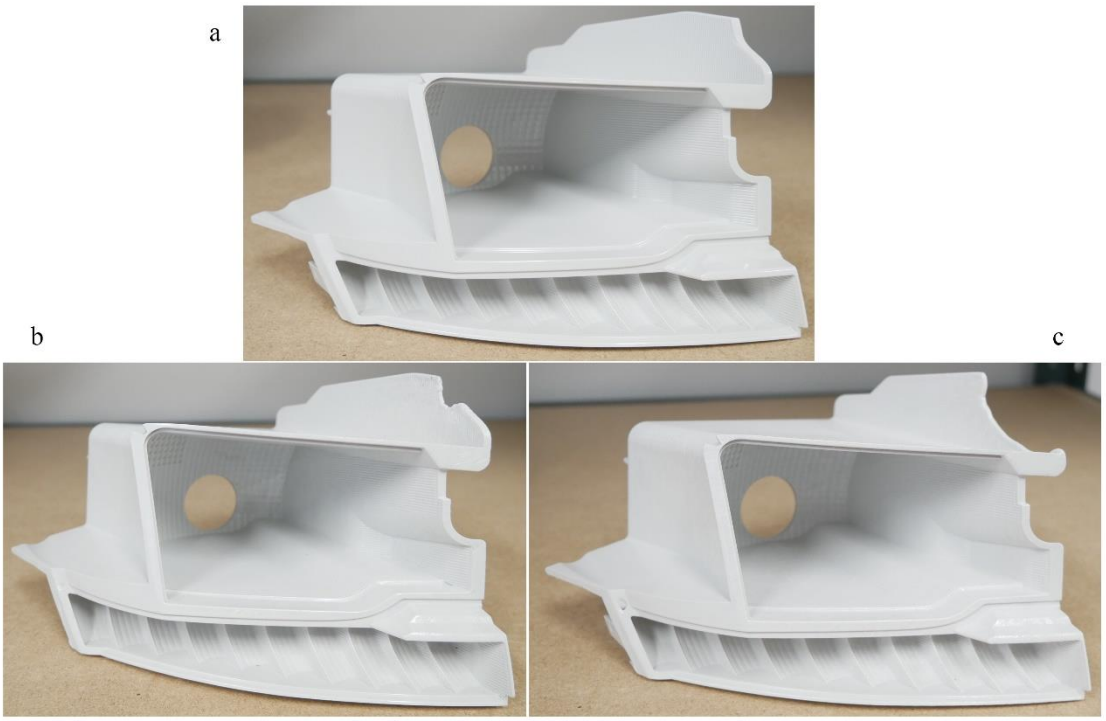

Fig. 6.3.19: a compliant reflector (a) compared to two components showing a faint shape incompleteness (b) and a severe incompleteness (c).

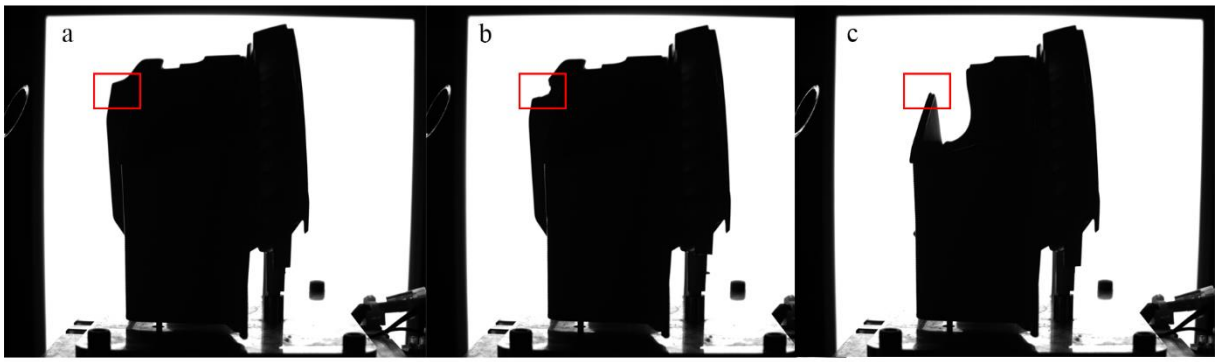

Fig. 6.3.20: visual representation of the shape incompleteness detection process. The red rectangle represents a possible ROI used to spot the flaw in the different cases: compliant component (a), faint defect (b), severe blemish (c).

To enhance its shape, the reflector is placed directly in front of a uniform and Lambertian light source, providing both spatial and angular uniform illumination. The algorithms work by comparing the acquired image with a reference, taken from a compliant sample. Once the image has been collected, rigid rotations and translations are performed to superpose it to the reference one. Marching squares algorithm is applied in a predefined region of interest (ROI), giving as output the profile contour of both reference and current image. Four features are extracted:

a) Discrete reference-sample mean distance
b) Discrete sample-reference mean distance
c) Area of the symmetric difference of the contours
d) Maximum distance sample-reference

Reference-sample and sample-reference mean distances are not the same in discrete geometry, because the contour is a discrete set of points: this is the reason why both of them are considered. As perceivable from *Fig. 6.3.20*, the differences between compliant reflectors (panel *a*) and defective ones (panel *b* and *c*) is evident focus the attention on the red rectangles.

Pin incompleteness

The other region which is lastly filled by the fluid PC is represented by the pin, highlighted in *Fig. 6.3.21*. The pins are fundamental for fixing the reflector to the housing and their shape must be perfect to avoid the detachment of the component. In *Fig. 6.3.21* a perfect-shape pin is represented (*a*), compared to one showing a tenuous flaw (*b*) and one representing a serious blemish (*c*).

Fig. 6.3.21: example of pin incompleteness. A compliant piece (a) is compared to a pin showing a faint incompleteness (b) and one presenting a severe defect (c).

The reflector has three pins: during the description only one will be considered, being the procedure the same for the other two. The algorithm selects a well-defined ROI in the proximity of the final section of the pin, highlighted by the red square in *Fig. 6.3.22*. In this region dark and bright zones, ascribable to the illumination direction, form a well-defined regular pattern. A "response" to this pattern can be estimated by the sum of intensities of all bright pixels minus the sum of intensities of all dark pixels

$$response = \sum\{I(bright) + [(-1) \cdot I(dark)]\}. \quad (6.3.4)$$

If the pin is absent or its shape is not ideal, the light/dark pattern is completely different with respect to the expected one. This is stressed by the abovementioned formula, leading to different numerical values in presence of incompleteness. The accuracy of this detection process is extremely high, being able to recognize defects even smaller than 0.5 mm. In fact, the minimal absence of plastic represented in *Fig. 6.3.22b* and *Fig. 6.3.22c* is easily identified by the prototype machine.

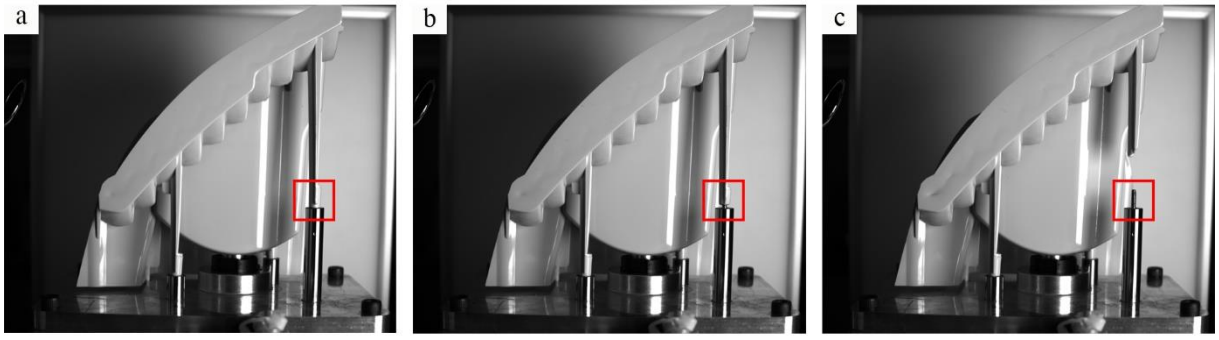

Fig. 6.3.22: example of detection of pin incompleteness. The red rectangle represents an enlargement of the ROI exploited for the pixel intensity analysis, which otherwise would have been impossible to be visualized. It is therefore possible to distinguish compliant (a) and faulty pieces (b, c).

6.3.4 Preliminary results

By means of an accurate supervised training process, it is possible to teach the machine to autonomously detect the scraps. The training consists in scanning both defective and non-defective samples manually labelling them as compliant or non-compliant. This process is fundamental to find the correct hyperplane in the feature space to separate the experimental data in two categories (corresponding to compliant/non-compliant classification), as described in *Section 4.2*. During the validation process, performed after the training, the reflectors have been scanned and classified autonomously by the prototype: the output produced by the machine has been then compared with the response of the quality operators, which represents the inspection standard. The results are summarized in the two tables below, for both the metallized and non-metallized samples (*Table 6.3.3* and *Table 6.3.4*).

Non-metallized samples					
Defect type	No. of Samples	False positives	False positives (%)	False negatives	False negatives (%)
Pin incompleteness	147	0	0 %	0	0 %
Figure incompleteness	147	0	0 %	0	0 %
Flash	147	9	6 %	1	0.7 %

Table 6.3.3: results of the scanning process on non-metallized samples.

Regarding the non-metallized samples, the setup chosen, and the algorithms implemented allowed the machine to perfectly detect the incompleteness defects (both affecting the figure and the pins). Concerning the flash, a 6% of the total components scanned have resulted in false positives (compliant components incorrectly discarded by the machine), while a 0.7% of reflectors have been classified as false negatives (defective components not recognized by the machine). When the analysis is performed on metallized reflectors, the flash flaw is more evident and the accuracy increases: in fact, as reported in *Table 6.3.4*, all the flashes are correctly detected. For the yellowish, both false positive and false negatives are present. The rates are low, and their incorrect classification is ascribable to their borderline color appearance which is extremely difficult to be correctly tracked. The outcomes of the analysis performed to evaluate the reflectance of the Al layer deposited over the plastic substrate demonstrated that no false positive has been found, meaning that not a single component has been incorrectly discarded by the prototype. However, 4% of the total components analyzed for the reflectance are false negatives.

Metallized samples					
Defect type	No. of Samples	False positives	False positives (%)	False negatives	False negatives (%)
Flash	98	0	0 %	0	0 %
Yellowish	103	3	3 %	4	4 %
Reflectance analysis	148	0	0 %	6	4 %

Table 6.3.4: results of the scanning process on metallized samples.

6.3.5 Discussion

A machine vision- and learning-based prototype setup to autonomously detect defects on the reflectors and to monitor their reflective properties has been designed and validated. The preliminary studies performed exploiting a visual inspection-based approach showed encouraging results, driving to the use of a luminance camera to evaluate the reflectance of the metallized reflectors. The prototype machine has been designed with two standard CCD cameras to monitor the structural and optical characteristics of the sample and with a photodiode able to detect the color appearance of the Al layer deposited. One of the two standard camera has been equipped with a photopic filter which, in combination with the constant variation of the integration times, permits to obtain information coherent with the luminance camera one. The preliminary results are promising, with an overall accuracy around 95% which can be even higher for some specific defects.

Moreover, the prototype has been designed in order to be easily implemented along the production lines chain to perform the scanning and give a real time feedback to the operators. At the end of every scan the software displays a 3D model of the reflector indicating, by means of a color code, the typologies of defects have been observed during the acquisition and their location, as represented in *Fig. 6.3.23*.

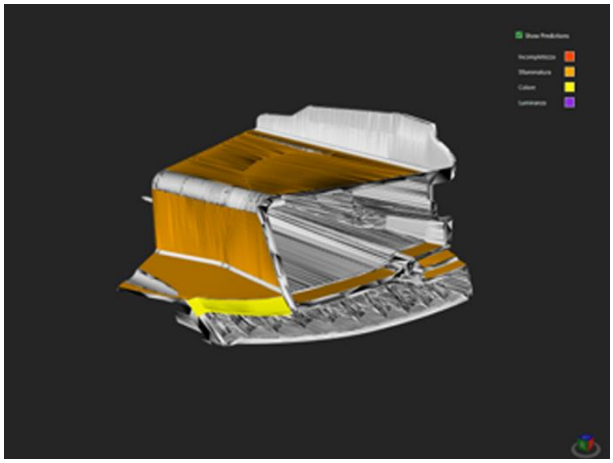

Fig. 6.3.23: example of the 3D model as displayed by the software after an acquisition. The colored regions indicate the presence of a blemish, whose family is described by the color-coded legend (top-right corner).

The main future perspective is to increase the detection accuracy of the machine until it reaches (ideally) the 100%. This can be performed by increasing the number of training scans, particularly for the yellowish and luminance blemishes to better recognize the borderline flaws. In addition to the fine tuning of the detection procedures, the next planned implementation is the realization of a communication system between the prototype detection machine, the metallization machine and the simulation software, working on automatic feedback to tune in real time the working parameters.

7. Conclusions and perspectives

The PhD project I have been involved in was a collaboration between the University of Trieste and Marelli Automotive Lighting Italy. The aim of this collaboration was the exploitation of surface physics and material science methodologies to give solutions to several industrial issues, associated to both the design and productive processes of a rear lamp. During the three-years-long-project, several answers to many questions have been provided.

The combined SEM and SR-CT analysis, performed on defective reflectors, allowed for a deep insight into the nature of surface and interface defects that may arise during the molding or the metallization production phases. In fact, by means of SEM it has been possible to gain topographic and chemical information regarding surface defects (white spots, black spots, scratches) with nanometric resolution, while exploiting the SR-CT the deep nature of the interface defects has been unraveled. Thanks to this study, the origins of several surface and interface defects have been understood, thus allowing to take some preventive actions to avoid their future formation.

The characterization of the optical properties of reflector-like samples has been performed by means of BRDF measurements: spectral and angular distributions of the light scattered from several samples' surfaces have been acquired to determine the single production parameter's effect on the overall optical properties. Exploiting the Harvey Shack representation and the *ABg* model, the experimental data have been fitted and the parameters A, B, and g have been extracted for each sample. Furthermore, by means of BRDF measurements, the rms surface roughness of the samples can be estimated. These values, obtained by optical measures, have been compared to the one extracted using AFM and excellent agreement has been found. Moreover, exploiting a photolithographic process, the thickness evaluation of both the Al and HMDSO thin films has been performed. These measurements have a prime importance because they permit to better represents the diverse metallization treatments with the ray-tracing simulations during the concept design of a rear lamp, thus allowing for a more reliable and more precise representation of the diverse configurations available.

Finally, two detection concept machines have been designed and validated during the doctoral project. The first, performing a machine vision-based analysis, grants a reliable quality check on the outer lenses. The second one is based on both machine-vision and machine-learning algorithms and allows for a fast and solid detection of scrap reflectors. The analysis is accomplished to spot both structural defects arising from the injection molding process and to check the correct optical properties given by the Al layer. The uniqueness of this machine consists in implementing a luminance-like analysis on the reflector to check if the Al coverage is adequate by upgrading a standard CCD camera. The results coming from the validation of both the machine are encouraging, with an overall accuracy of 95%. The two prototypes represent a big step into the "Industry 4.0" framework, leading the possibility to perform real time and automatic quality analysis on each component.

One of the main perspectives is to perform BRDF measurements on glass flat samples, thus being insensitive to the deformation of the plastic substrate. Those measurements are fundamental to unravel the fine influence of the Al and HMDSO layers on the overall optical properties of the samples. Moreover, exploiting transparent substrates (glass) the Bidirectional Transmittance Distribution Function (BTDF) of the samples can be studied, therefore being able to fully characterize all the optical properties (reflection, transmission and absorption) of

the metal and protective layers. Those measurements were initially scheduled in March-April 2020, but the covid-19 pandemic made it impossible to complete the task.

Regarding the prototype machine dedicated to the automatic scan of reflectors, the next and most important upgrade is to connect it to the molding and metallizing software, generating a closed feedback loop. In case of defects, the feedback loop will give the possibility to tune in real time and following a computer-based way the production parameters to correct the faults. This will be, for the company, the definitive jump in the “Industry 4.0” environment, allowing for a safer and more controlled parameters selection.

8. Scientific production

Publications

- Fontanot, T., Ermacora, D., Simonetti, G., Raducci, S., Vesselli, E., & Paroni, S. "An automatic visual inspection system to scan outer lenses of automotive rear lamps". *Optical Measurement Systems for Industrial Inspection XI* (Vol. 11056, p. 1105624). International Society for Optics and Photonics (2019).
- Fontanot, T., Audenaert, J., Hanselaer, P., Pecorari, I., Lughi, V., Vesselli, E., Paroni, S. & Leloup, F. B. "BRDF characterization of Al-coated thermoplastic polymer surfaces". *J Coat Technol Res* 17, 1195-1205 (2020).
- Fontanot, T., Donato, S., Brombal, L., Di Lillo, F., Dreossi, D., Rigon, L., Longo, R., Dal Zilio, S., Ciancio, R., Del Linz, M., Paroni, S., & Vesselli, E. "On the origin of surface and interface defects associated with the growth of al-coated thermoplastic heterostacks". *Surface and Interface Analysis* (2020) *in press.*

Patents

- European Patent Application No. 20209318.3, *pending*.
- European Patent Application No. 20209308.4, *pending*.

Conferences

- Speaker: SPIE Optical Metrology – Optical Measurements Systems for Industrial Inspection XI", 24-27 June 2019, Munich (Germany), with the work entitled "*An automatic visual inspection system to scan outer lenses of automotive rear lamps*".
- Speaker: GDR Appamat Workshop, part of the Color and Imaging Conference (CIC)", 21-25 October 2019, Paris (France), with the work entitled "*BRDF characterization of metallized plastics*".

9. Bibliography

1. Frank, A. G., Dalenogare, L. S. & Ayala, N. F. Industry 4.0 technologies: Implementation patterns in manufacturing companies. *Int. J. Prod. Econ.* **210**, 15–26 (2019).
2. Lasi, H., Fettke, P., Kemper, H. G., Feld, T. & Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **6**, 239–242 (2014).
3. Wollschlaeger, M., Sauter, T. & Jasperneite, J. The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0. *IEEE Ind. Electron. Mag.* **11**, 17–27 (2017).
4. Vaidya, S., Ambad, P. & Bhosle, S. Industry 4.0 - A Glimpse. *Procedia Manuf.* **20**, 233–238 (2018).
5. Guizani, M. *The industrial internet of things*. *IEEE Network* vol. 33 (2019).
6. Guariente, P., Antoniolli, I., Ferreira, L. P., Pereira, T. & Silva, F. J. G. Implementing autonomous maintenance in an automotive components manufacturer. *Procedia Manuf.* **13**, 1128–1134 (2017).
7. Bagheri, B., Yang, S., Kao, H. A. & Lee, J. Cyber-physical systems architecture for self-aware machines in industry 4.0 environment. *IFAC-PapersOnLine* **28**, 1622–1627 (2015).
8. Azizi, A. Evaluation Improvement of Production Productivity Performance using Statistical Process Control, Overall Equipment Efficiency, and Autonomous Maintenance. *Procedia Manuf.* **2**, 186–190 (2015).
9. Shrouf, F., Ordieres, J. & Miragliotta, G. Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm. *IEEE Int. Conf. Ind. Eng. Eng. Manag.* **2015-Janua**, 697–701 (2014).
10. Witkowski, K. Internet of Things, Big Data, Industry 4.0 - Innovative Solutions in Logistics and Supply Chains Management. *Procedia Eng.* **182**, 763–769 (2017).
11. Lee, J., Davari, H., Singh, J. & Pandhare, V. Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manuf. Lett.* **18**, 20–23 (2018).
12. Xu, L. Da, Xu, E. L. & Li, L. Industry 4.0: State of the art and future trends. *Int. J. Prod. Res.* **56**, 2941–2962 (2018).
13. Penumuru, D. P., Muthuswamy, S. & Karumbu, P. Identification and classification of materials using machine vision and machine learning in the context of industry 4.0. *J. Intell. Manuf.* **31**, 1229–1241 (2020).
14. Candanedo, I. S., Nieves, E. H., González, S. R., Martín, M. T. S. & Briones, A. G. Machine learning predictive model for industry 4.0. *Commun. Comput. Inf. Sci.* **877**, 501–510 (2018).
15. Pérez, L., Rodríguez, Í., Rodríguez, N., Usamentiaga, R. & García, D. F. Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors (Switzerland)* **16**, (2016).
16. Wan, J. *et al.* Software-Defined Industrial Internet of Things in the Context of Industry 4.0. *IEEE Sens. J.* **16**, 7373–7380 (2016).
17. Wildes, R. P. *et al.* A machine-vision system for iris recognition. *Mach. Vis. Appl.* **9**, 1–8 (1996).
18. Keenan, S. J. *et al.* An automated machine vision system for the histological grading of cervical intraepithelial neoplasia (CIN). *J. Pathol.* **192**, 351–362 (2000).
19. Blasco, J., Aleixos, N. & Moltó, E. Machine vision system for automatic quality grading of fruit. *Biosyst. Eng.* **85**, 415–423 (2003).
20. Luo, X., Jayas, D. S. & Symons, S. J. Identification of damaged kernels in wheat using a colour machine vision system. *J. Cereal Sci.* **30**, 49–59 (1999).
21. Bulanon, D. M. & Kataoka, T. Fruit detection system and an end effector for robotic harvesting of Fuji apples. *Agric. Eng. Int. CIGR J.* **12**, 1–14 (2010).
22. Kurada, S. & Bradley, C. A machine vision system for tool wear assessment. *Tribol. Int.* **30**, 295–304 (1997).
23. Foo, S. Y. A rule-based machine vision system for fire detection in aircraft dry bays and engine compartments. *Knowledge-Based Syst.* **9**, 531–540 (1996).
24. Oliver, N., Rosario, B. & Pentland, A. A Bayesian computer vision system for modeling human interactions. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **1542**, 255–272 (1999).
25. Lahajnar, F., Bernard, R., Pernuš, F. & Kovačič, S. Machine vision system for inspecting electric plates. *Comput. Ind.* **47**, 113–122 (2002).
26. Yachida, M. & Tsuji, S. A Versatile Machine Vision System for Complex Industrial Parts. *IEEE Trans. Comput.* **C**–**26**, 882–894 (1977).
27. Huang, W. & Kovacevic, R. Development of a real-time laser-based machine vision system to monitor and control welding processes. *Int. J. Adv. Manuf. Technol.* **63**, 235–248 (2012).
28. Romeo, L. *et al.* Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0. *Expert Syst. Appl.* **140**, (2020).

29. Coffey, V. C. Machine Vision: The Eyes of Industry 40. *Opt. Photonics News* **29**, 42 (2018).
30. Vieira, A. A. C., Dias, L. M. S., Santos, M. Y., Pereira, G. A. B. & Oliveira, J. A. Setting an industry 4.0 research and development agenda for simulation – A literature review. *Int. J. Simul. Model.* **17**, 377–390 (2018).
31. Uriarte, A. G., Ng, A. H. C. & Moris, M. U. Supporting the lean journey with simulation and optimization in the context of Industry 4.0. *Procedia Manuf.* **25**, 586–593 (2018).
32. Rodič, B. Industry 4.0 and the New Simulation Modelling Paradigm. *Organizacija* **50**, 193–207 (2017).
33. Fujimoto, A., Tanaka, T. & Iwata, K. Arts: Accelerated Ray-Tracing System. *IEEE Comput. Graph. Appl.* **6**, 16–26 (1986).
34. Kay, T. L. & Kajiya, J. T. Ray tracing complex scenes. *Proc. 13th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1986* **20**, 269–278 (1986).
35. Wald, I., Slusallek, P., Benthin, C. & Wagner, M. Interactive rendering with coherent ray tracing. *Comput. Graph. Forum* **20**, 153–165 (2001).
36. Hart, J. C., Sandin, D. J. & Kauffman, L. H. Ray tracing deterministic 3-D fractals. *Proc. 16th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1989* 289–296 (1989) doi:10.1145/74333.74363.
37. Segovia, B., Iehl, J. C. & Peroche, B. Coherent Metropolis Light Transport with Multiple-Try Mutations. *Tech Rep. LIRIS* 12 (2007).
38. Liu, Z., Wang, K., Luo, X. & Liu, S. Precise optical modeling of blue light-emitting diodes by Monte Carlo ray-tracing. *Opt. Express* **18**, 9398 (2010).
39. Veach, E. & Guibas, L. J. Metropolis light transport. *Proc. 24th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1997* 65–76 (1997) doi:10.1145/258734.258775.
40. Govaerts, Y. M. & Verstraete, M. M. Raytran: a monte carlo raytracing model to compute light scattering in threedimensional heterogeneous media. *IEEE Trans. Geosci. Remote Sens.* **36**, 493505 (1998).
41. Azodi, H., Siart, U. & Eibert, T. F. A fast 3-D deterministic ray tracing coverage simulator including creeping rays based on geometry voxelization technique. *IEEE Trans. Antennas Propag.* **63**, 210–220 (2015).
42. Kazanjian, A. R. Review of properties of polycarbonate resins. *Polym. Plast. Technol. Eng.* **2**, 123–160 (1973).
43. Greco, R. & Sorrentino, A. Polycarbonate/ABS blends: A literature review. *Adv. Polym. Technol.* **13**, 249–258 (1994).
44. Kumar, R. *et al.* Acrylonitrile Butadiene Styrene (ABS) plastic-based low cost tissue equivalent phantom for verification dosimetry in IMRT. *J. Appl. Clin. Med. Phys.* **11**, 24–32 (2010).
45. Rodriguez, J. Maximizing the strength of fused-deposition ABS plastic parts. *10th Solid Free. …* 335–342 (1999).
46. Ali, U., Karim, K. J. B. A. & Buang, N. A. A Review of the Properties and Applications of Poly (Methyl Methacrylate) (PMMA). *Polym. Rev.* **55**, 678–705 (2015).
47. Pawar, E. A Review Article on Acrylic PMMA Eshwar Pawar. *ISOR J. Mech. Civ. Eng.* **13**, 1–4 (2016).
48. Schanda, J. *Colorimetry: understanding the CIE system*. (John Wiley & Sons, 2007).
49. https://nbertoa.wordpress.com/2017/08/02/lighting-series-part-5-photometry/.
50. Sagawa, K. & Takeichi, K. Spectral luminous efficiency functions in the mesopic range. *J. Opt. Soc. Am. A* **3**, 71 (1986).
51. CIE (1926). Commission Internationale de l'Éclairage. in *Colourimetry (2nd ed.). Vienna: CIE Publication.* (Cambridge University Press, 1996).
52. CIE (1932). Commission Internationale de l'Eclairage. in (Cambridge University Press).
53. http://greendisplay.com.cn/Technology/BACKLIGHT/CIE.html.
54. Agrawal, A. R., Pandelidis, I. O. & Pecht, M. Injection-molding process control—A review. *Polym. Eng. Sci.* **27**, 1345–1357 (1987).
55. Bryce, D. *Plastic Injection Molding: mold design and construction fundamentals*. vol. III (1998).
56. Seaman, C. M., Desrochers, A. A. & List, G. F. Multiobjective Optimization of a Plastic Injection Molding Process. *IEEE Trans. Control Syst. Technol.* **2**, 157–168 (1994).
57. Dang, X. P. General frameworks for optimization of plastic injection molding process parameters. *Simul. Model. Pract. Theory* **41**, 15–27 (2014).
58. Tsai, C. C. & Lu, C. H. Multivariable self-tuning temperature control for plastic injection molding process. *IEEE Trans. Ind. Appl.* **34**, 310–318 (1998).
59. Coleman, B. D. & Noll, W. The thermodynamics of elastic materials with heat conduction and viscosity. *Arch. Ration. Mech. Anal.* **13**, 167–178 (1963).
60. Huang, G., Chan, P. K. & Kamal, M. R. Phase diagram for liquid crystalline polymer/polycarbonate blends. *Can. J. Chem. Eng.* **81**, 243–257 (2003).
61. Kim, W. N. & Burns, C. M. Phase behavior of blends of polycarbonate with partially miscible polymers. *J. Appl. Polym. Sci.* **41**, 1575–1593 (1990).

62. Hatzius, K., Li, Y., Werner, M. & Jungnickel, B. J. Phase diagram and crystallization of polycarbonate/poly-ε-caprolactone blends. *Angew. Makromol. Chemie* **243**, 177–187 (1996).
63. Harris, M., Potgieter, J., Ray, S., Archer, R. & Arif, K. M. Acrylonitrile butadiene styrene and polypropylene blend with enhanced thermal and mechanical properties for fused filament fabrication. *Materials (Basel).* **12**, (2019).
64. Chen, X., Lam, Y. C. & Li, D. Q. Analysis of thermal residual stress in plastic injection molding. *J. Mater. Process. Technol.* **101**, 275–280 (2000).
65. Seow, L. W. & Lam, Y. C. Optimizing flow in plastic injection molding. *J. Mater. Process. Technol.* **72**, 333–341 (1997).
66. Bryce, D. M. *Plastic Injection Molding Fundamentals*. vol. II (1997).
67. http://www.tecnovacuum.it/tecnomet-combi-.html.
68. Djennas, F., Prack, E. & Matsuda, Y. Investigation of Plasma Effects on Plastic Packages Delamination and Cracking. *IEEE Trans. Components, Hybrids, Manuf. Technol.* **16**, 919–924 (1993).
69. Kaplan, S. L. & Rose, P. W. Plasma surface treatment of plastics to enhance adhesion. *Int. J. Adhes. Adhes.* **11**, 109–113 (1991).
70. Kuhr, M., Bauer, S., Rothhaar, U. & Wolff, D. Coatings on plastics with the PICVD technology. *Thin Solid Films* **442**, 107–116 (2003).
71. Mahan, J. ~E. Physical Vapor Deposition of Thin Films. *Phys. Vap. Depos. Thin Film. by John E.~Mahan, pp.~336.~ISBN 0-471-33001-9.~Wiley-VCH , January 2000.* 336 (2000).
72. Sproul, W. D. Physical vapor deposition tool coatings. *Surf. Coatings Technol.* **81**, 1–7 (1996).
73. Mattox, D. M. *Deposition (PVD) Processing Second edition Dedication To my wife Vivienne*. (2009). doi:10.1016/B978-0-8155-2037-5.00025-3.
74. Wang, Y. A study of PVD coatings and die materials for extended die-casting die life. *Surf. Coatings Technol.* **94**–**95**, 60–63 (1997).
75. Pierson, H. O. *the Chemistry of Cvd. Handbook of Chemical Vapor Deposition* (1992). doi:10.1016/b978-0-8155-1300-1.50009-x.
76. Vasudev, M. C., Anderson, K. D., Bunning, T. J., Tsukruk, V. V. & Naik, R. R. Exploration of plasma-enhanced chemical vapor deposition as a method for thin-film fabrication with biological applications. *ACS Appl. Mater. Interfaces* **5**, 3983–3994 (2013).
77. Meyerhofer, D. Characteristics of resist films produced by spinning. *J. Appl. Phys.* **49**, 3993–3997 (1978).
78. Bornside, D. E., Macosko, C. W. & Scriven, L. E. Spin coating: One-dimensional model. *J. Appl. Phys.* **66**, 5185–5193 (1989).
79. Mack, C. *Fundamental Principles of Optical Lithography: The Science of Microfabrication*. *Fundamental Principles of Optical Lithography: The Science of Microfabrication* (2007). doi:10.1002/9780470723876.
80. Reiser, A. *et al.* The molecular mechanism of novolak-diazonaphthoquinone resists. *Eur. Polym. J.* **38**, 619–629 (2002).
81. Che-Ping, L., Hsiharng, Y. & Ching-Kong, C. A new microlens array fabrication method using UV proximity printing. *J. Micromechanics Microengineering* **13**, 748–757 (2003).
82. Lin, T. H., Yang, H. & Chao, C. K. Concave microlens array mold fabrication in photoresist using UV proximity printing. *Microsyst. Technol.* **13**, 1537–1543 (2007).
83. Walker, E. J. Reduction of Photoresist Standing-Wave Effects by Post-Exposure Bake. *IEEE Trans. Electron Devices* **22**, 464–466 (1975).
84. Erdman, N., Bell, D. C. & Reichelt, R. *Scanning Electron Microscopy*. (2019).
85. Jablonski, A., Salvat, F. & Powell, C. J. NIST Electron Elastic-Scattering Cross Section Datbase #64 Version 3.1. (2003).
86. Murata, K. Spatial distribution of backscattered electrons in the SEM and electron microprobe. *J. Appl. Phys.* **45**, 4110 (1974).
87. Drescher, H., Reimer, L. & Seidel, H. Ruckstreukoeffizient und Sekundarelektronenausbeute von 10–100 keV-Elektronen und Beziehungen zur Raster-Elektronenmikroskopie. *Z. Angew. Phys.* **29**, 331 (1970).
88. Napchan, E. Backscattered Electrons in the SEM. *Microsc. Anal.* 9–11 (2001).
89. Heinrich, K. F. J. Optiques de rayons X et microanalyse. in *Fourth International Congress on XRay Optics and Microanalysis* (eds. Castaing, R., Deschamps, P. & Philibert, J.) (Hermann, 1966).
90. Hawkes, P. W. & Spence, J. C. *Springer Handbook of Microscopy*. (Springer Nature, 2019).
91. Kramers, H. A. On the theory of x-ray absorption and of the continuous x-ray spectrum. *Philos. Mag.* **46**, 836 (1923).
92. Reimer, L. *Scanning Electron Microscopy*. (1985).
93. Stephenson, S. T. The continuous x-ray spectrum. in *Handbuch der Physik* 337–370 (Springer, 1957).
94. Bambynek, W. *et al.* X-ray fluorescent yields, Auger, and Coster–Kronig transition probabilities. *Rev.*

Mod. Phys. **44**, 716 (1972).
95. Xing, Q. Information or resolution: Which is required from an SEM to study bulk inorganic materials? *Scanning* **38**, 864–879 (2016).
96. Nixon, W. C. Scanning electron microscopy. *Microelectron. Reliab.* **4**, 55–56 (1965).
97. Mohammed, A. & Abdullah, A. Scanning Electron Microscopy (Sem): a Review. *Proc. 2018 Int. Conf. Hydraul. Pneum. - HERVEX* 85 (2018).
98. Mironov, V. L. Fundamentals of scanning probe microscopy. 207 (2004) doi:10.1007/978-3-540-34315-8.
99. Binnig, G.; Quate, C.F.; Gerber, C. Atomic Force Microscope. *Pysical Rev. Lett.* **56**, 5-1-5–9 (1986).
100. Maragliano, C., Glia, A., Stefancich, M. & Chiesa, M. Effective AFM cantilever tip size: Methods for in-situ determination. *Meas. Sci. Technol.* **26**, (2015).
101. Song, Y. & Bhushan, B. Coupling of cantilever lateral bending and torsion in torsional resonance and lateral excitation modes of atomic force microscopy. *J. Appl. Phys.* **99**, (2006).
102. Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64**, 1868–1873 (1993).
103. Stark, R. W., Drobek, T. & Heckl, W. M. Thermomechanical noise of a free v-shaped cantilever for atomic-force microscopy. *Ultramicroscopy* **86**, 207–215 (2001).
104. Albrecht, T. R., Akamine, S., Carver, T. E. & Quate, C. F. Microfabrication of cantilever styli for the atomic force microscope. *J. Vac. Sci. Technol. A Vacuum, Surfaces, Film.* **8**, 3386–3396 (1990).
105. Lee, J. S. *et al.* Multifunctional hydrogel nano-probes for atomic force microscopy. *Nat. Commun.* **7**, 1–14 (2016).
106. Martin, Y., Williams, C. C. & Wickramasinghe, H. K. Atomic force microscope-force mapping and profiling on a sub 100-Å scale. *J. Appl. Phys.* **61**, 4723–4729 (1987).
107. Meyer, G. & Amer, N. M. Novel optical approach to atomic force microscopy. *Appl. Phys. Lett.* **53**, 1045–1047 (1988).
108. Tichy, J., Erhart, J., Kittinger, E. & Privratska, J. *Fundamentals of piezoelectric sensorics: mechanical, dielectric, and thermodynamical properties of piezoelectric materials*. (2010).
109. Baselt, D. R., Clark, S. M., Youngquist, M. G., Spence, C. F. & Baldeschwieler, J. D. Digital signal processor control of scanned probe microscopes. *Rev. Sci. Instrum.* **64**, 1874–1882 (1993).
110. Hölscher, H., Schwarz, U. D. & Wiesendanger, R. Calculation of the frequency shift in dynamic force microscopy. *Appl. Surf. Sci.* **140**, 344–351 (1999).
111. Paulo, Á. S. & García, R. Tip-surface forces, amplitude, and energy dissipation in amplitude-modulation (tapping mode) force microscopy. *Phys. Rev. B - Condens. Matter Mater. Phys.* **64**, 1–4 (2001).
112. Cleveland, J. P., Anczykowski, B., Schmid, A. E. & Elings, V. B. Energy dissipation in multifrequency atomic force microscopy. *Appl. Phys. Lett.* **72**, 2613–2615 (1998).
113. Anczykowski, B., Gotsmann, B., Fuchs, H., Cleveland, J. P. & Elings, V. B. How to measure energy dissipation in dynamic mode atomic force microscopy. *Appl. Surf. Sci.* **140**, 376–382 (1999).
114. https://scientificservices.eu/item/atomic-force-microscope---solver-pro/15.
115. L. Rigon. X-Ray Imaging with Coherent Sources. in *Comprehensive Biomedical Physics* (ed. Brahme) 193–220 (Amsterdam: Elsevier, 2014).
116. Cloetens, P., Barrett, R., Baruchel, J., Guigay, J.-P. & Schlenker, M. Phase objects in synchrotron radiation hard x-ray imaging. *J. Phys. D. Appl. Phys.* **29**, 133–146 (1996).
117. Azaroff, L. V. Elements of X-ray crystallography by L. V. Azaroff. *McGraw-Hill* (1969) doi:10.1107/s0567739470001808.
118. Brombal, L. *X-Ray Phase-Contrast Tomography: Underlying Physics and Developments for Breast Imaging*. (Springer, Nature).
119. Snigirev, A., Snigireva, I., Kohn, V., Kuznetsov, S. & Schelokov, I. On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation. *Rev. Sci. Instrum.* **66**, 5486–5492 (1995).
120. Arhatari, B. D., Mancuso, A. P., Peele, A. G. & Nugent, K. A. Phase contrast radiography: Image modeling and optimization. *Rev. Sci. Instrum.* **75**, 5271–5276 (2004).
121. Gureyev, T. E. *et al.* Some simple rules for contrast, signal-to-noise and resolution in in-line x-ray phase-contrast imaging. *Opt. Express* **16**, 3223 (2008).
122. Gureyev, T. E. Composite techniques for phase retrieval in the Fresnel region. *Opt. Commun.* **220**, 49–58 (2003).
123. Nugent, K. A. X-ray noninterferometric phase imaging: a unified picture. *J. Opt. Soc. Am. A* **24**, 536 (2007).
124. Langer, M., Cloetens, P., Guigay, J. P. & Peyrin, F. Quantitative comparison of direct phase retrieval algorithms in in-line phase tomography. *Med. Phys.* **35**, 4556–4566 (2008).
125. De Caro, L. *et al.* Combined mixed approach algorithm for in-line phase-contrast x-ray imaging. *Med. Phys.* **37**, 3817–3827 (2010).

126. Wu, X., Yan, A. & Liu, H. X-ray phase-shifts-based method of volumetric breast density measurement. *Med. Phys.* **39**, 4239–4244 (2012).
127. Barrett, H. H. III The Radon Transform and Its Applications. *Prog. Opt.* **21**, 217–286 (1984).
128. Buzug, T. M. *Computed Tomography: from photon statistics to modern cone-beam CT*. (2008).
129. Natter, F. The Mathematics of Computerized Tomography. (2001).
130. Nielsen, J. & McMorrow, D. *Elements of modern X-ray physics*. (John Wiley & Sons, 2011).
131. Rivers, M. Tutorial introduction to X-ray computed microtomography data processing. *University of Chicago* (1998).
132. Tromba, G. *et al.* The SYRMEP beamline of elettra: Clinical mammography and bio-medical applications. *AIP Conf. Proc.* **1266**, 18–23 (2010).
133. T. E. Gureyev, Y. I. Nesterets, A. Kozlov, D. M. Paganin, and H. M. Q. On the "unreasonable" effectiveness of transport of intensity imaging and optical deconvolution. *J. Opt. Soc. Am. A* **34**, 2251–2260 (2017).
134. Brombal, L. *et al.* Phase-contrast breast CT: The effect of propagation distance. *Phys. Med. Biol.* **63**, (2018).
135. A. Piai, A. Contillo, F. Arfelli, D. Bonazza, L. Brombal, M. A. Cova, P. Delogu, V. Di Trapani, S. Donato, B. Golosio, G. Mettivier, P. Oliva, L. Rigon, A. Taibi, M. Tonutti, G. Tromba, F. Zanconati, and R. L. Quantitative characterization of breast tissues with dedicated CT imaging. *Phys. Med. Biol.* **64**, 155011 (2019).
136. Lewis, R. R. Making Shaders More Physically Plausible. *Comput. Graph. Forum* **13**, 109–120 (1994).
137. Cook, R. L. & Torrance, K. E. Reflectance Model for Computer Graphics. *Comput Graph.* **15**, 307–316 (1981).
138. Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W. & Limperis, T. Geometrical considerations and nomenclature for reflectance. *NBS Monogr. 160* (1977).
139. Nicodemus, F. E. Directional reflectance and emissivity of an opaque surface. *Appl. Opt.* **4**, 767–775 (1965).
140. Koenderink, J. J. & van Doorn, A. J. Phenomenological description of bidirectional surface reflection. *J. Opt. Soc. Am. A* **15**, 2903 (1998).
141. Kajiya, J. T. The rendering equation. *Proc. 13th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1986* **20**, 143–150 (1986).
142. Clarke, F. J. J. & Parry, D. J. Helmholtz reciprocity: Its validity and application to reflectometry. *Light. Res. Technol.* **17**, 1–11 (1985).
143. Ng, K. H., Tameh, E. K. & Nix, A. R. A new hybrid geometrical optics and radiance based scattering model for ray tracing applications. *IEEE Int. Conf. Commun.* **4**, 2168–2172 (2005).
144. Potemin, I. S. *et al.* Realistic image synthesis in presence of birefrigent media by backward ray tracing technique. 12 (2018) doi:10.1117/12.2312675.
145. Glassner, A. S. *Introduction to Ray Tracing*. (Academic Press Inc., 1989).
146. Leloup, F. B., Forment, S., Dutré, P., Pointer, M. R. & Hanselaer, P. Design of an instrument for measuring the spectral bidirectional scatter distribution function. *Appl. Opt.* **47**, 5454–5467 (2008).
147. International, ASTM West Conshohocken, P. ASTM E2387-19, Standard Practice for Goniometric Optical Scatter Measurements. (2019).
148. Ferrero, A. Theoretical evaluation of the impact of finite intervals in the measurement of the bidirectional reflectance distribution function. *J. Coatings Technol. Res.* **17**, 81–90 (2020).
149. Sahoo, P. K., Soltani, S. & Wong, A. K. C. A survey of thresholding techniques. *Comput. Vision, Graph. Image Process.* **41**, 233–260 (1988).
150. Pikaz, A. & Averbuch, A. Digital image thresholding, based on topological stable-state. *Pattern Recognit.* **29**, 829–843 (1996).
151. McCormick, N. & Lord, J. Digital image correlation. *Mater. Today* **13**, 52–54 (2010).
152. Chu, T. C., Ranson, W. F. & Sutton, M. A. Applications of digital-image-correlation techniques to experimental mechanics. *Exp. Mech.* **25**, 232–244 (1985).
153. Spont, H. & Cardelino, J. A Review of Classic Edge Detectors Algorithms Based on the First Derivative. *Ipol* **5**, 90–123 (2015).
154. Gupta, S., Gupta, C. & Chakarvarti, S. K. Image Edge Detection : A Review. *Int. J. Adv. Res. Comput. Eng. Technol.* **2**, 2246–2251 (2013).
155. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 142–158 (2016).
156. Rajab, M. I., Woolfson, M. S. & Morgan, S. P. Application of region-based segmentation and neural network edge detection to skin lesions. *Comput. Med. Imaging Graph.* **28**, 61–68 (2004).
157. Gonzalez, R. C. & Woods, R. E. *Digital image processing*. (Prentice Hall, 2002).
158. Vikram Mutneja, D. Methods of Image Edge Detection: A Review. *J. Electr. Electron. Syst.* **04**, (2015).
159. Mukherjee, S. & Acton, S. T. Region based segmentation in presence of intensity inhomogeneity using

legendre polynomials. *IEEE Signal Process. Lett.* **22**, 298–302 (2015).
160. Butchiraju, K. & Saikiran, B. Region-based segmentation and object detection. *Int. J. Innov. Technol. Explor. Eng.* **8**, 366–368 (2019).
161. Matas, J., Chum, O., Urban, M. & Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image Vis. Comput.* **22**, 761–767 (2004).
162. Sheng, M., Li, J. & Shi, Y. Relative degree adaptive flooding broadcast algorithm for Ad hoc networks. *IEEE Trans. Broadcast.* **51**, 216–222 (2005).
163. Arango, J., Degermark, M., Efrat, A. & Pink, S. An efficient flooding algorithm for ad-hoc networks. *Proc. Second Work. Model. Optim. Mob. Ad Hoc Wirel. Networks* (2004).
164. Forssén, P. E. & Lowe, D. G. Shape descriptors for maximally stable extremal regions. *Proc. IEEE Int. Conf. Comput. Vis.* (2007) doi:10.1109/ICCV.2007.4409025.
165. Lorensen, W. E. & Cline, H. E. A High Resolution 3D Surface Construction Algorithm.PDF. *ACM Transactions on Graphics* vol. 21 163–169 (1987).
166. Wenger, R. *Isosurfaces: geometry, topology, and algorithms*. (CRC Press, 2013).
167. Chen, H. *et al.* Edge-Enhanced Maximally Stable Extremal Regions. *Ieee Int. Conf. Image Process.* 2657–2660 (2011).
168. Wu, Z. & Sullivan, J. M. Multiple material marching cubes algorithm. *Int. J. Numer. Methods Eng.* **58**, 189–207 (2003).
169. Noble, W. S. What is a support vector machine? *Nat. Biotechnol.* **24**, 1565–1567 (2006).
170. Moh'd A Mensleh, A. Chi square feature extraction based SVMS arabic language text categorization system. *J. Comput. Sci.* **3**, 430–435 (2007).
171. Joachims, T. Text Categorization with SVM: Learning with Many Relevant Features. *Eur. Conf. Mach. Learn.* 137–142 (2005).
172. Cauwenberghs, G. & Poggio, T. Incremental and decrementai support vector machine learning. *Adv. Neural Inf. Process. Syst.* (2001).
173. Tang, Y. Deep Learning using Linear Support Vector Machines. (2013).
174. Cho, B. H. *et al.* Nonlinear support vector machine visualization for risk factor analysis using nomograms and localized radial basis function kernels. *IEEE Trans. Inf. Technol. Biomed.* **12**, 247–256 (2008).
175. Onel, M., Kieslich, C. A. & Pistikopoulos, E. N. A nonlinear support vector machine-based feature selection approach for fault detection and diagnosis: Application to the Tennessee Eastman process. *AIChE J.* **65**, 992–1005 (2019).
176. Amari, S. & Wu, S. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks* **12**, 783–789 (1999).
177. Howley, T. & Madden, M. G. The genetic kernel support vector machine: Description and evaluation. *Artif. Intell. Rev.* **24**, 379–395 (2005).
178. Heylen, R., Parente, M. & Gader, P. A review of nonlinear hyperspectral unmixing methods. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 1844–1868 (2014).
179. Kuo, B. C., Ho, H. H., Li, C. H., Hung, C. C. & Taur, J. S. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**, 317–326 (2014).
180. Cao, H., Naito, T. & Ninomiya, Y. Approximate RBF Kernel SVM and Its Applications in Pedestrian Classification. *Mach. Learn.* 1–9 (2008).
181. Ravale, U., Marathe, N. & Padiya, P. Feature selection based hybrid anomaly intrusion detection system using K Means and RBF kernel function. *Procedia Comput. Sci.* **45**, 428–435 (2015).
182. Chang, Q., Chen, Q. & Wang, X. Scaling Gaussian RBF kernel width to improve SVM classification. *Proc. 2005 Int. Conf. Neural Networks Brain Proceedings, ICNNB'05* **1**, 19–22 (2005).
183. Fernandes, A. O., Moreira, L. F. E. & Mata, J. M. Machine vision applications and development aspects. *IEEE Int. Conf. Control Autom. ICCA* 1274–1278 (2011) doi:10.1109/ICCA.2011.6138014.
184. Wallace, A. M. Industrial Applications of Computer Vision Since 1982. *IEE Proc. E Comput. Digit. Tech.* **135**, 117–136 (1988).
185. Kruger, R. P. & Thompson, W. B. A Technical and Economic Assessment of Computer Vision for Industrial Inspection and Robotic Assembly. *Proc. IEEE* **69**, 1524–1538 (1981).
186. Shafi, A. Machine vision in automotive manufacturing. *Sens. Rev.* **24**, 337–342 (2004).
187. Lindeberg, T. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. J. Comput. Vis.* **11**, 283–318 (1993).
188. Logothetis, N. K., Pauls, J. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5**, 552–563 (1995).
189. Mikolajczyk, K. & Schmid, C. An affine invariant interest point detector. *Proc. ECCV* 128–142 (2002).
190. Low, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 91–110 (2004).

191. Donoser, M., Riemenschneider, H. & Bischof, H. Shape guided Maximally Stable Extremal Region (MSER) tracking. *Proc. - Int. Conf. Pattern Recognit.* 1800–1803 (2010) doi:10.1109/ICPR.2010.444.
192. Shah, R. & Ward, P. T. Lean manufacturing: Context, practice bundles, and performance. *J. Oper. Manag.* **21**, 129–149 (2003).
193. Yang, M. G., Hong, P. & Modi, S. B. Impact of lean manufacturing and environmental management on business performance: An empirical study of manufacturing firms. *Int. J. Prod. Econ.* **129**, 251–261 (2011).
194. Wu, Y. C. Lean manufacturing: A perspective of lean suppliers. *Int. J. Oper. Prod. Manag.* **23**, 1349–1376 (2003).
195. Ben Naylor, J., Naim, M. M. & Berry, D. Leagility: integrating the lean and agile manufacturing paradigms in the total supply chain. *Int. J. Prod. Econ.* **62**, 107–118 (1999).
196. Melton, T. The benefits of lean manufacturing: What lean thinking has to offer the process industries. *Chem. Eng. Res. Des.* **83**, 662–673 (2005).
197. Sodkomkham, T. & Chutima, P. Lean Six Sigma Application in Rear Combination Automotive Lighting Process. *IOP Conf. Ser. Mater. Sci. Eng.* **131**, (2016).
198. Giraud, D., Borit, F., Guipont, V., Jeandin, M. & Malhaire, J. M. Metallization of a polymer using cold spray: Application to aluminum coating of polyamide 66. *Proc. Int. Therm. Spray Conf.* **6**, 265–270 (2012).
199. Kong, L. B., Cheung, C. F., Jiang, J. B., To, S. & Lee, W. B. Characterization of freeform optics in automotive lighting systems using an Optical-Geometrical Feature Based Method. *Optik (Stuttg).* **122**, 358–363 (2011).
200. Lachmayer, R., Kloppenburg, G. & Wolf, A. Rapid prototyping of reflectors for vehicle lighting using laser activated remote phosphor. *Light. Diodes Mater. Devices, Appl. Solid State Light. XIX* **9383**, 938305 (2015).
201. Sarmadi, A. M., Ying, T. H. & Denes, F. HMDSO-plasma modification of polypropylene fabrics. *Eur. Polym. J.* **31**, 847–857 (1995).
202. Vautrin-Ul, C. *et al.* Plasma-polymerized coatings using HMDSO precursor for iron protection. *Prog. Org. Coatings* **38**, 9–15 (2000).
203. Vincens, A., El-Khatib, F. & Chang, M. Influence of aluminum coating thickness on automotive lamps. *SAE Tech. Pap.* **2008**, (2008).
204. Mota, R. P. *et al.* HMDSO plasma polymerization and thin film optical properties. *Thin Solid Films* **270**, 109–113 (1995).
205. Jones, R. Synchrotron radiation microcomputed tomography for assessing internal cracks in cold spray repairs. *Fatigue Fract. Eng. Mater. Struct.* **43**, 431–432 (2020).
206. Wu, S. C., Xiao, T. Q. & Withers, P. J. The imaging of failure in structural materials by synchrotron radiation X-ray microtomography. *Eng. Fract. Mech.* **182**, 127–156 (2017).
207. Robinson, S. & Robinson, A. H. Chemical composition of sweat. *Physiol. Rev.* **34**, 202–220 (1954).
208. Fukumoto, T. *et al.* Differences in composition of sweat induced by thermal exposure and by running exercise. *Clin. Cardiol.* **11**, 707–709 (1988).
209. Granier, A., Vervloet, M., Aumaille, K. & Vallée, C. Optical emission spectra of TEOS and HMDSO derived plasmas used for thin film deposition. *Plasma Sources Sci. Technol.* **12**, 89–96 (2003).
210. Gosar, Ž. *et al.* Characterization of Gaseous Plasma Sustained in Mixtures of HMDSO and O2 in an Industrial-Scale Reactor. *Plasma Chem. Plasma Process.* **40**, 25–42 (2020).
211. Bruckman, L. S. *et al.* Degradation of back surface acrylic mirrors: Implications for low concentration and mirror augmented photovoltaics. *2012 IEEE Energytech, Energytech 2012* 1–4 (2012) doi:10.1109/EnergyTech.2012.6304678.
212. Bernacki, B. E., Johnson, T. J., Myers, T. L. & Blake, T. A. Modeling liquid organic thin films on substrates. 37 (2018) doi:10.1117/12.2299873.
213. Freniere, E. R., Gregory, G. G. & Chase, R. C. Interactive Software for Optomechanical Modeling. **3130**, 128–133.
214. Harvey, J. E. Light Scattering Characteristics of Optical Surfaces. *Stray Light Probl. Opt. Syst. Int. Soc. Opt. Photonics* **107**, 41–47 (1977).
215. Audenaert, J., Leloup, F. B., Durinck, G., Deconinck, G. & Hanselaer, P. Bayesian Deconvolution Method Applied to Experimental Bidirectional Transmittance Distribution Functions. *Meas. Sci. Technol.* **24**, 035202 (2013).
216. Harvey, J. E. Total integrated scatter from surfaces with arbitrary roughness, correlation widths, and incident angles. *Opt. Eng.* **51**, 013402 (2012).
217. Hass, G., Hunter, W. R. & Tousey, R. Reflectance of Evaporated Aluminum in the Vacuum Ultraviolet. *J. Opt. Soc. Am.* **46**, 1009 (1956).
218. Karoui, A. Aluminum Ultra Thin Film Grown by Physical Vapor Deposition for Solar Cell Electric Nanocontacts. *ECS Trans.* **41**, 21–28 (2019).

219. Schmauder, T., Sauer, P. & Ickes, G. New Reflectors and Reflector Coaters. in *In 57th Annual Technical Conference Proceeding of the Society of Vacuum Coaters* (2014).
220. Schmauder, T., Kuper, S. & Wohlfahrt, P. High Reflective Silver Coatings on 3D Plastic Parts for Solar Concentrators. in *52nd Annual Technical Conference Proceedings of SVC* (2009).
221. Nielsen, J. B., Jensen, H. W. & Ramamoorthi, R. On Optimal, Minimal BRDF Sampling for Reflectance Acquisition. **34**, (2015).
222. Marschner, S. R., Westin, S. H., Lafortune, E. P. F. & Torrance, K. E. Image-based bidirectional reflectance distribution function measurement. *Appl. Opt.* **39**, 2592 (2000).
223. Amma, S. I., Luo, J., Pantano, C. G. & Kim, S. H. Specular reflectance (SR) and attenuated total reflectance (ATR) infrared (IR) spectroscopy of transparent flat glass surfaces: A case study for soda lime float glass. *J. Non. Cryst. Solids* **428**, 189–196 (2015).
224. Bhavsar, H. & Panchal, M. H. A Review on Support Vector Machine for Data Classification. *Int. J. Adv. Res. Comput. Eng. Technol.* **1**, 2278–1323 (2012).
225. Mountrakis, G., Im, J. & Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **66**, 247–259 (2011).
226. Thorsteinsson, S. *et al.* Accurate microfour-point probe sheet resistance measurements on small samples. *Rev. Sci. Instrum.* **80**, (2009).
227. Panta, G. & Subedi, D. Electrical characterization of aluminum (Al) thin films measured by using four-point probe method. *Kathmandu Univ. J. Sci. Eng. Technol.* **8**, 31–36 (2013).
228. Julesz, B. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biol. Cybern.* **41**, 131–138 (1981).
229. Tugnait, J. K. & Luo, W. On channel estimation using superimposed training and first-order statistics. *IEEE Commun. Lett.* **7**, 413–415 (2003).
230. Ajdadi, F. R., Gilandeh, Y. A., Mollazade, K. & Hasanzadeh, R. P. R. Application of machine vision for classification of soil aggregate size. *Soil Tillage Res.* **162**, 8–17 (2016).
231. Alvarenga, A. V., Pereira, W. C. A., Infantosi, A. F. C. & Azevedo, C. M. Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images. *Med. Phys.* **34**, 379–387 (2007).
232. Pare, S., Bhandari, A. K., Kumar, A. & Singh, G. K. An optimal color image multilevel thresholding technique using grey-level co-occurrence matrix. *Expert Syst. Appl.* **87**, 335–362 (2017).
233. Gebejes, a, Master, E. M. & Samples, a. Texture Characterization based on Grey-Level Co-occurrence Matrix. *Conf. Informatics Manag. Sci.* 375–378 (2013).
234. Li, C. H. & Lee, C. K. Minimum cross entropy thresholding. *Pattern Recognit.* **26**, 617–625 (1993).
235. Vazquez, P.-P., Feixas, M., Sbert, M. & Heidrich, W. Viewpoint Selection using Viewpoint Entropy. *VMV* **1**, (2001).
236. Tou, J. Y., Lau, P. Y. & Tay, Y. H. Computer Vision-based Wood Recognition System. *Proc. Int'l Work. Adv. Image Technol.* (2007).
237. Wang, Y. Research on the computer vision cracked eggs detecting method. *Int. J. Comput. Appl. Technol.* **50**, 215–219 (2014).

Acknowledgements

During the 3-years-long project, I had the possibility to work with a lot of expert scientists from all over the world. From my point of view, this was the most stimulating aspect of my PhD: living in challenging environments which strive for excellence and perfection gave me the chance to learn something new every day and thus increase my knowledge and my skills. I will be always grateful for this opportunity, which helped me growing both from the scientific and human sides.

I would like to thank in particular prof. Erik Vesselli and dr. Sara Paroni: without them all this project would have been impossible. Thanks to their example I understood how to better behave in both the research and industrial fields, knowing the specific aims of each sector and how to reach them.

Mr. Mirco Del Linz helped me understanding the physics behind the metallization process and how to correctly metallize the samples: I will always be grateful to him for his never-ending explanations and demonstrations.

I would like to thank all the people I have met during this long itinerary. Unluckily, they are too many to write each of their names, therefore I will write only the institution/department they belong too. Please, do not be offended.

- Automotive Lighting R&D (particularly the ELT),
- Dipartimento di Ingegneria ed Architettura (Units),
- Light&Lighting laboratory,
- CNR-IOM,
- SYRMEP beamline staff,
- Datamind.

The key to keep pursuing my dreams is represented by my girlfriend and all my family: their never-ending support was (and still is) unbelievable.

Lastly, I would like to thank all my friends: you have been a trustworthy source of entertainment to set my mind free in the busiest times of this itinerary.