Single-pixel, single-photon three-dimensional imaging

Gabriella Musarra

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science & Engineering
University of Glasgow

March 2020

Declaration of Authorship

With the exception of chapters 1, 2, which contain introductory material, all work in this thesis was carried out by the author unless otherwise explicitly stated.

I, Gabriella Musarra, declare that this thesis titled, 'Single-pixel, single-photon three-dimensional imaging' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Abstract

The 3D recovery of a scene is a crucial task with many real-life applications such as self-driving vehicles, X-ray tomography and virtual reality. The recent development of time-resolving detectors sensible to single photons allowed the recovery of the 3D information at high frame rate with unprecedented capabilities. Combined with a timing system, single-photon sensitive detectors allow the 3D image recovery by measuring the Time-of-Flight (ToF) of the photons scattered back by the scene with a millimetre depth resolution.

Current ToF 3D imaging techniques rely on scanning detection systems or multi-pixel sensor. Here, we discuss an approach to simplify the hardware complexity of the current 3D imaging ToF techniques using a single-pixel, single-photon sensitive detector and computational imaging algorithms. The 3D imaging approaches discussed in this thesis do not require mechanical moving parts as in standard Lidar systems. The single-pixel detector allows to reduce the pixel complexity to a single unit and offers several advantages in terms of size, flexibility, wavelength range and cost. The experimental results demonstrate the 3D image recovery of hidden scenes with a sub-second acquisition time, allowing also non-line-of-sight scenes 3D recovery in real-time. We also introduce the concept of intelligent Lidar, a 3D imaging paradigm based uniquely on the temporal trace of the return photons and a data-driven 3D retrieval algorithm.

List of Publications

Papers

1. A. Turpin, G. Musarra, F. Tonolini, A. Lyons, I. Starshynov, F. Villa, E. Conca, F. Fioranelli, R. Murray-Smith, and D. Faccio. *3D imaging with a single, time-resolving detector*. Accepted by Optica (2020)

2. G. Musarra, A. Lyons, E. Conca, Y. Altmann, F. Villa, F. Zappa, M. J. Padgett and D. Faccio. *Non-Line-of-Sight Three-Dimensional Imaging with a Single-Pixel Camera*. Physical Review Applied, 12 (1) 011002, 2019. doi:10.1103/PhysRevApplied.12.011002T

3. G. Musarra, P. Caramazza, A. Turpin, A. Lyons, C. F. Higham, R. Murray-Smith and D. Faccio. *Detection, identification, and tracking of objects hidden from view with neural networks*. Advanced Photon Counting Techniques XIII, 10978 (1097803) 063833, 2019. doi:/10.1117/12.2519721

4. G. Musarra, K. E. Wilson, D. Faccio, E. M. Wright. *Rotation-dependent nonlinear absorption of orbital angular momentum beams in ruby*. Optics letters, 43 (13) 3073–3075, 2018. doi:10.1364/OL.43.003073

5. D. Pierangeli, G. Musarra, F. Di Mei, G. Di Domenico, A. J. Agranat, C. Conti, E. DelRe. *Enhancing optical extreme events through input wave disorder*. Physical Review A, 94 (6) 063833, 2016. doi:10.1103/PhysRevA.94.063833

Conference contributions

1. G. Musarra, A. Turpin, I. Starshynov, A. Lyons, E. Conca, F. Villa, D. Faccio *Single-photon, single-pixel intelligent Lidar.* Oral presentation, Single-photon Workshop 2019 (October 2019).

2. A. Turpin, G. Musarra, F. Tonolini, R. Murray-Smith and D. Faccio*iLIDAR: intelligent, cross-modality training of intelligent LIDAR and single-point detectors.* Invited oral presentation, OSA Laser congress (September 2019).

3. G. Musarra, A. Lyons, E. Conca, F. Villa, F. Zappa, Y. Altmann, D. Faccio. *3D RGB Non-Line-Of-Sight single-pixel imaging*. Oral presentation, OSA Imaging and applied optics conference (June 2019).

4. A. Turpin, G. Musarra, I. Starshynov, A. Lyons, J. Brooks, and D. Faccio. *Single-point LIDAR: Full-3D single-frame LIDAR from a single-pixel.* Oral presentation, OSA Imaging and applied optics conference (June 2019).

5. G. Musarra, A. Lyons, M. P. Edgar, M. J. Padgett and D. Faccio. *Non line of sight single pixel camera*. Oral presentation, Photon18, (September 2018).

6. G. Satat, G. Musarra, A. Lyons, B. Heshmat, D. Faccio and R. Raskar. *Compressive Ultrafast Single Pixel Camera*. Oral presentation, OSA Imaging and applied optics conference, (June 2018).

7. P. Caramazza, A. Boccolini, G. Musarra, M. Hullin, R. Murray-Smith, and D. Faccio. *Machine Learning Assisted Identification of People Hidden Behind a Corner* . Oral presentation, Computational Optical Sensing and Imaging 2017 (June 2017).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

First used by Mait et al. [13], the term Computational Imaging (CI) refers to as the application of computational algorithms to optical measurements with the purpose of simplifying the hardware complexity of the physical measurement process. Establishing a strong coupling between post-processing computational algorithms and optical measurements, CI changed the concept of image formation and in most of the cases, the optical measurements may not even look like an image. CI allows to overcome the physical limits of the measurement process and the dimensionality mismatch between the 2D image and a 3D reality. The use of post-processing computational algorithms in imaging systems has been demonstrated in medical imaging [14, 15], quantum imaging [16], volumetric imaging [17] and Light Detection and Ranging (Lidar) [12]

The concurrent advances in CI algorithms and time-stamped, single-photon-sensitive detectors led to a new form of ultra-fast imaging capable of "freezing the light in motion". Time-stamped single-photon detectors are sensors able to measure the arrival-time of a single photon with picosecond temporal resolution, storing the information in a form of temporal histogram. An image can then be obtained by elaborating the temporal trace with CI algorithms. Operating in correlation with timing systems such as Time-Correlated Single-Photon Counting (TCSPC) modules, single-photon sensitive detectors are widely applied in many technologies including Light-in-Flight (LiF), i.e. the visual representation of the light in motion, and Time-of-Flight (ToF).

ToF technology infers the spatial information of a 3D scene by the ToF of the light, that is, the time the light takes to travel along a direct or indirect path. The combination of single-photon counting detectors and CI algorithms allows the 3D image recovery of Line-Of-Sight (LOS) and Non-Line-Of-Sight (NLOS) scenes from the ToF information of the back-scattered photons.

Current 3D imaging systems usually rely on scanning systems or pixelated detection. In scanning imaging systems such as Light-Detection-And-Ranging (LiDAR), a laser spot scans the target

scene and a single-pixel detector confocally acquires the return signal. On the contrary, pixelated detection technology flash-illuminates the whole scene and simultaneously acquires the return signal from different points with a spatially resolved (or multi-pixel) detector. Measuring the arrival-time of multiple scattered photons along indirect path, this technique can also be extended to the 3D imaging of NLOS scenes.
Current ToF techniques based on scanning systems have some limitations such as the number of measurements required to obtain the entire image with consequent longer acquisition times. Moreover they require the use of mechanical moving parts, as it happens in LiDAR or Velodyne Lidar systems. On the other hand, pixelated detectors or cameras can result in bulky devices with limited temporal resolution, low Photon Detection Efficiency (PDE) or limited operating wavelengths spectrum. An implementation of the current detection systems for the recovery of 3D scene is the single-pixel camera, a device requiring a Spatial-Light-Modulator (SLM) and only a single light sensor to create an image.

The work discussed in this thesis provides an implemented approach to simplify the hardware of current 3D imaging ToF techniques by using a single-pixel, single-photon sensitive detector. The single-pixel design offers several advantages in terms of size, flexibility, wavelength range, structure complexity and cost. Correlating the intensity of the return photons with a series of multiple spatially-resolved 2D patterns, single-pixel detectors may also provide faster timing response, higher detection sensitivity and lower dark counts.
In Chapter 2 we will describe the detection technologies we use in the experiments in order to measure the arrival-time of the return photons. We will discuss how Single-Photon-Avalanche-Diodes (SPADs) and Photomultiplier tubes (PMTs) sensors operate in TCSPC mode to generate the return temporal profile for applications requiring high PDE, high temporal resolution and low dark counts. Chapter 2 will also introduce the single-pixel camera concept.
Chapter 3 will experimentally investigate the 3D information retrieval of a LOS scene combining a single-pixel camera with a lock-in amplifier, a device used to extract signals with a given temporal dependence from a noisy background. As demonstrated with the simulations in Chapter 3.2.1, the lock-in amplifier provides a phase resolution one order of magnitude better than conventional continuous wave (CW) modulated ToF cameras. We experimentally demonstrate the 3D retrieval of a LOS scene with a depth resolution of 5 mm, providing an alternative CW modulation imaging system with no mechanical scanning parts or multi-pixel detectors. However, the proposed method is affected by some limitations such as the number of measurements and the time resources not compatible with real-time applications, as demonstrated by the experimental results in Section 3.4.
Chapter 4 discusses the 3D recovery of hidden scenes with a single-pixel camera. In the "look

around corner" experiment, a time-resolving single-pixel camera acquires the multi-bounced return signal and the back-propagation algorithm allows the 3D retrieval of the scene. Combining a high PDE single-pixel detector and Hadamard pattern detection, the experimental results in Chapter 4 demonstrate the 3D recovery of a NLOS scene with an improved sub-second acquisition time, paving the way to real-time 3D shape recovery of hidden scenes.

In the last three chapters of this thesis (Chapters 5-7) we introduce the concept of Intelligent Lidar (ILidar), a 3D imaging paradigm that uses a single temporal histogram and a neural network (NN) retrieval algorithm. Chapter 5 will introduce the main concepts of NNs and how supervised learning approach has been used in many fields of research. The aim of the supervised learning approach in our case is to infer the inverse light transport model transformation that maps the return photons temporal histogram into the corresponding 3D image.

In Chapter 6 we introduce the main current technologies used to retrieve the 3D information of LOS scenes, discussing their advantages, limitations and the pixelated sensor requirement. Chapter 6 will then theoretically introduce the concept of ILidar.

ILidar is a data-driven approach for 3D imaging that recovers dynamic LOS scenes using only a single ToF temporal histogram acquired by a single-pixel, time-resolving detector. This approach allows the 3D recovery potentially at a KHz or even MHz frame rate. Since the ILidar uses a single temporal histogram, no spatial structure is imprinted in the 3D image recovery.

The information contained in a single temporal histogram of the return photons is not sufficient to univocally determine the spatial information of the scene. The standard 3D imaging technologies provide the additional information of the scene either by scanning or by pixelated detection. The ILidar 3D imaging approach provides the additional information by statistical representation of the possible scenes, on which a neural network model can be trained by using pairs of arrival-time measurements and corresponding 3D images. A data-driven approach can then be used to retrieve the 3D image of a scene from a single arrival-time histogram by a single-pixel detector, which in this case is chosen as a SPAD sensor.

In conclusion, Chapter 7 will report the experimental results obtained by testing the proposed 3D imaging paradigm on dynamic scenes composed by targets of different shapes and size freely moving in a room. The experimental results demonstrate a compact 3D imaging paradigm changing the fundamental concepts of 3D imaging, requiring less amount of data transferring, storage and handling than other conventional ToF 3D imaging approaches. Since the proposed paradigm is based only on arrival-time measurements and a data-driven algorithm, this method can be extended and applied to completely different platforms, provided that an optical system based training is formerly performed. As an example of cross-modality imaging with pulsed source, we test our system on a radar platform by using an impulse radar chip transceiver, as described in the last chapter of this thesis.

Chapter 2

Time-stamped single-photon sensitive detectors for 3D imaging

In this chapter we describe how single-photon sensitive detectors such as SPAD or PMT sensors can be used to recover the 3D information of a Line-Of-Sight (LOS) and Non-Line-Of-Sight (NLOS) scene. We then discuss how time-gated single-photon sensitive detectors generate the temporal trace of the return photons by operating in correlation with a clock, referred to as Time-Correlated Single-Photon Counting (TCSPC). Finally we introduce the single-pixel camera concept.

2.1 Ultra-fast, single-photon sensitive cameras for 3D imaging.

Since the well known "The Horse in Motion" experiment of Eadweard Muybridge showing sequential pictures of the gallop of a horse [18], having cameras with high frame rate plays a crucial role in imaging dynamic objects or scenes moving at high speed. This requirement becomes prohibitive when we try to capture the motion of the fastest object in the universe travelling at a speed of 3x10^8m/s: light.

This problem is addressed with revolutionary cameras sensitive to single photons and capable of measuring the arrival-time of a photon with picosecond temporal resolution. These detectors such as Single Photon Avalanche Diodes (SPADs) or Photomultiplier Tubes (PMTs), can indeed capture the motion in time of the Light-In-Flight (LiF) by timing the photon detection with a clock. Recent advances in CI algorithms and single-photon counting cameras led to the development of a new form of high speed photography capable to capture images at $5{\cdot}10^{11}$ frames per second [19, 20]

with unprecedented capabilities for LOS and NLOS 3D imaging.
Single-photon detection techniques rely on intensity gating [18], holographic gating [21] or continuous capture [22]. The application of computational algorithms of transient imaging [23] allows then to retrieve an image from the optical measurements. Used for imaging events with picosecond dynamic evolution, time-stamped single-photon sensitive cameras allowed the visualization of light in flight in air [24] or murky water [25].

More related to the 3D imaging of direct and indirect scenes, single-photon sensitive cameras can also be used in measurements of Time-of-Flight (ToF). In a common ToF scenario, a light source illuminates the scene and a time-stamped, single-photon counting detector acquires the return signal. We can then retrieve the 3D information of the scene from the ToF of the return photons. Imaging based on single-photon ToF technologies have been demonstrated in LiDAR [26–28], real-time tracking of hidden objects [24,29], 3D imaging of hidden scenes, [30,31], imaging through diffusive media [32,33] and ghost imaging [34,35].
The techniques commonly used to achieve picosecond temporal resolution are based on ToF indirect measurement, temporal mapping or temporal gating. Techniques based on indirect measurement of the ToF recover the temporal information by the phase offset of the return of a continuous wave whose amplitude is sinusoidal modulated in time [22]. In serial time-encoded amplified imaging (STEAM), the temporal information is inferred by mapping the temporal-domain into a different one such as the wavelength domain [36]. Finally, temporal gating techniques recover the temporal information by using mechanical or optical implemented shutters such as Kerr cells [18].

2.2 Single-pixel, single-photon sensitive detectors

Here, we introduce the operating principles of single-photon sensitive detectors and how in their pixelated or array version they operate in time-resolved single photon counting applications.
We describe the operating principles of a SPAD, the detecting device that has been mainly used in the experiments reported in this thesis. Offering single-photon sensitivity with ultrafast time-resolving capability, a SPAD sensor allows to measure the ToF information with picosecond resolution operating in TCSPC mode.

2.2.1 Single Photon Avalanche Diodes

First invented in the 1990's, a SPAD detector is a semiconductor avalanche photodiode operating in a Geiger-mode above the breakdown voltage V_{BD} and therefore sensitive to single-photon detection. Recent results demonstrated the use of SPADs in photon counting applications such as LiDAR [37, 38], tracking of targets in NLOS scenes [24, 29], fluorescence lifetime measurements [39, 40] and quantum cryptography [41, 42]. SPAD detectors offer high single-photon sensitivity and picosecond temporal resolution, representing a leading technology among the time-resolving photon counting devices. Current SPAD sensors are available in a single-pixel or an array format. The main advantages of using SPADS for photon-counting are the high PDE, wider spectral range of operation and low power consumption.

A SPAD consists of a junction of a p-doped semiconductor and a n-doped semiconductor, as shown in the SPAD cross-section (Fig. 2.1 (a)). The p-doped section called anode, contains a concentration of holes, whereas the n-doped section called cathode, contains a concentration of electrons. On the contrary, there are no charges in the depletion region at the junction of the opposite-doped sections since the free carriers recombine. In order to attract the electrons and the holes respectively to the cathode and to the anode, a positive voltage is applied to the cathode and a negative voltage is applied to the anode. Operating the SPAD in reverse bias voltage beyond the breakdown voltage V_{BD} and applying a strong electric field across the junction, the device is sensitive to single photons.

The SPAD operating principle is depicted in Fig. 2.1(b) showing the current I through the device and the corresponding voltage V for a single-photon detection [43]. Before detecting a photon, a strong voltage $-\|V_{EB}\| + \|V_{BD}\|$ beyond the breakdown voltage V_{BD} is applied to the SPAD. Here, the quantity V_{EB} indicates the excess bias voltage applied in order to improve the Photon Detection Probability (PDP) and to increase the total bias voltage beyond the breakdown limit.

With reference to Fig. 2.1(b), the initial configuration of the SPAD is the state "1" and no current is circulating in the device. When a photon is detected in the active area, it creates a hole-electron pair, generating free carriers across the junction. In a self-sustainable state process, the resulting electron is accelerated by the strong electric field and it gains enough kinetic energy to generate an avalanche of free charges by impact ionisation. At the applied bias voltage the associated electric field is indeed so intense that a single carrier injected in the depletion region can trigger a self-sustained avalanche. As we increase the excess bias voltage, the Photon Detection Probability (PDP) of the detector increases. Indeed higher excess bias voltage induces higher electric fields and therefore higher probabilities of generating an electron-hole pair.

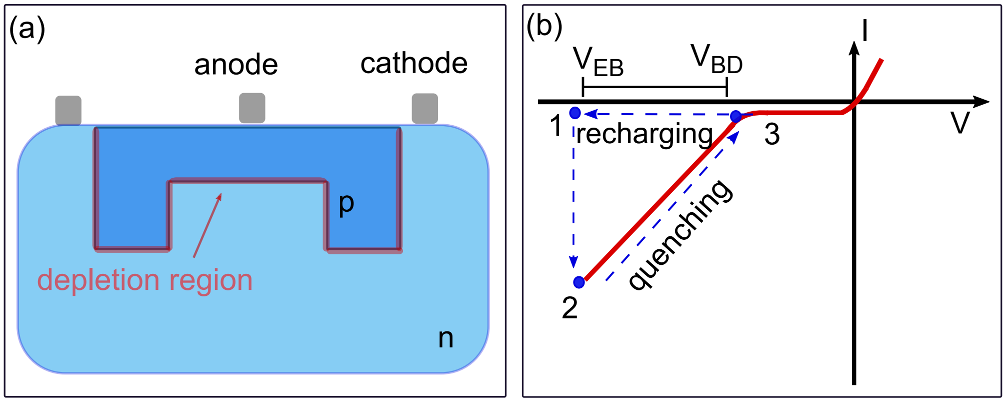

Figure 2.1: **Operating principle of a SPAD detector.** a) Cross-section of a SPAD detector composed by a junction of p-doped (holes) and n-doped (electron) semiconductor. By applying a reverse bias voltage, the holes are attracted to the anode, whereas the electrons are attracted to the cathode. In the depletion region at the junction of the two opposite doped parts there are no free charges. b) Current–voltage curve of a SPAD for a single photon detection. Before detecting a photon, the SPAD is in state "1" when a strong bias voltage $-\|V_{EB}\| + \|V_{BD}\|$ is applied but no current is passing through the SPAD. When a photon is detected, the generated photoelectron creates a self sustaining avalanche by impact ionisation and so a strong current (state "2"). In order to reset the SPAD to the initial state, the voltage is restored to the V_{BD} by the quenching process and then the SPAD is reset to the initial state "1" by the excess bias voltage V_{EB}. The SPAD is then ready for another photon detection. The SPAD repeats the same cycle (1-2-3) for every photon detection event.

The current generated by the avalanche denotes the detection of the photon and the SPAD occupies the state "2". In order to reset the SPAD to the initial state and proceed with the next photon detection, the current in the SPAD is reset to the initial value with the quenching process. During this quenching the bias voltage of the SPAD is restored to the breakdown value until the depletion area is free of charges (state "3"). The SPAD is restored to the initial state applying again a V_{EB} in the "recharging" process. However, during the quenching the SPAD is not able to detect any photon for an amount of time called "dead time" varying from hundreds of nanoseconds to microseconds range.

The SPAD used in the experiments reported in this thesis has been manufactured by the SPAD lab group of Dipartimento di Elettronica, Informazione e Bioingegneria at Politecnico di Milano [2]. The SPAD detector is manufactured in a 0.16 μm BCD technology with monolithically integrated sensing circuit. It has an active surface of 57x57 μm^2 with a temporal resolution of 32 ps, as shown in Fig. 4.7. SPAD devices tend to have small active area in order to have faster timing and to reduce the thermally generated dark counts.

2.2.2 Photomultiplier tube

First invented in the 1930’s, current photo-counting PMT technologies offer a picosecond temporal resolution and high sensitivity in a wide range of wavelengths [44, 45].

Figure 2.2 shows the basic principles of a PMT. When a photon hits the cathode of the PMT, an electron is generated through the photoelectric effect. The electron is accelerated towards a dynode chain of multiple electrodes in series. While the electron cascades down the chain, it creates an avalanche of electrons exciting secondary electrons that are accelerated to the next dynode. The millions of electrons generated at the end of the dynode chain are then absorbed by the anode, generating an electrical pulse. Measuring the electric pulse with an electronic counter is then possible to count the absorbed photon.

To ensure that the electrons cascade down the tube, a high voltage is distributed along the dynode chain and the entire system is assembled in a vacuum glass tube.

PMTs have some limitations such as high voltage or the high sensibility to mechanical vibrations and electromagnetic disturbances. Additionally, PMTs in the IR range usually require a cooling system, demanding more mechanics to stabilize the photon detection. PMTs usually offer fast timing response and high temporal resolution.

The PMT used in the experiment reported in Section 4 is an HPM-100-07 module manufactured by the Becker&Hickl [45] fully controlled by the Becker&Hickl DCC-100 card [46] and it contains a Hamamatsu R10467 GaAsP hybrid PMT tube. It has a Quantum Efficiency (QE) up to a 23% in the visible range, an active area of 3 mm of diameter and a temporal resolution of 27 ps, as shown in Fig. 4.6.

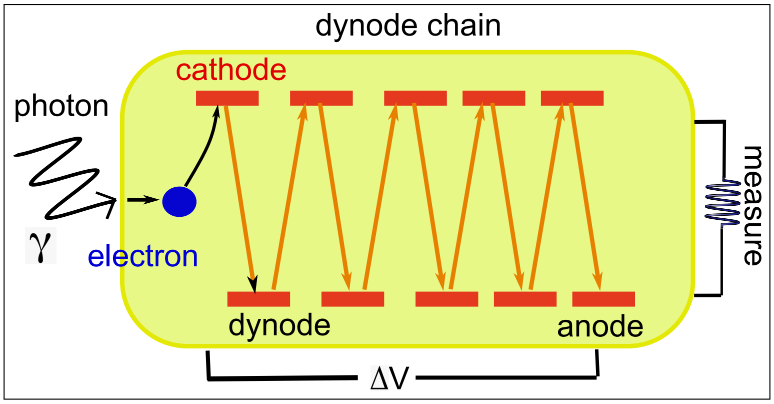

Figure 2.2: **Operating principle of a PMT detector.** When a photon hits the cathode, an electron is generated by photoelectric effect. The electron is then accelerated towards a dynode chain creating an avalanche of electrons through secondary emission. The millions of electrons are then absorbed by the anode generating an electrical pulse measured by an electronic counter.

2.2.3 Silicon photomultiplier

Another detecting technology used in the experiments discussed in this thesis is the Silicon Photomultiplier (SiPM) or Multi-Pixel Photon Counter (MPPC). Suitable for photon detection in the visible and in the near infrared region, it consists of a solid state photomultiplier of microcell SPADs all connected to a common current summing node. Suitable for single-photon counting in low light levels, SiPM detectors are characterized by high timing resolution, resilience to magnetic field, low voltage and visible and near infrared range wavelengths. Used in applications such as LiDAR [47], dark matter detection [48] and chemiluminescence [49], current SiPM detectors offer a fill factor of 30%-80%, a gain factor of 10^5-10^6.

The SiPM detector we used in the experiment in Chapter 3 is a compact C14455-3050GA module manufactured by Hamamatsu Photonics. It consists of thermo-electrically cooled Multi-Pixel Photon Counters (MPPC, S14422 series), a temperature controller, a voltage power supplier circuit and an amplifier. The SiPM detector (or single channel) is composed by 2836 pixels and it has an effective photosensitive area of 3 mm of diameter and a 34% PDE at 532 nm. It works in analog mode where the measured output is either a current or a voltage and the modules can be operated just by an external voltage.

2.2.4 Time-Correlated Single-Photon Counting

Usually built as a separate unit from the single-photon sensitive detector, advanced TCSPC electronics [50] is becoming a crucial technology for applications such as short and long depth range [29, 51] and fluorescence life-time measurements [52, 53]. By operating in correlation with the temporal signal created by the absorption of a photon across the SPAD, it provides the clock used to measure the photon arrival-time with picosecond resolution.

Usually provided as a card, the role of the TCSPC module is to measure the difference in time between a trigger and the detection of a photon by time-to-digital converters (TDC). Many experiments requiring time-resolved photon counting employ the train of pulses emitted by a pulsed laser as a trigger. When a photon is absorbed, the TDC computes the time difference between the trigger and the detection of the photon measuring the difference in the transit time of electric signals through a series of logic gates. The time can be measured also in a reverse start-stop mode where the detection of the photon starts the clock and the trigger terminates the clock.

The number of detected photons and the corresponding arrival-time are stored in a temporal histogram containing the temporal profile of the photons detection. The x-axis of the temporal histogram represents the measured arrival-time of the photons and the y-axis represents the number of photons (or photon counts) detected at that corresponding time. The temporal axis of the arrival-time histogram is arranged in time bins with a given duration. Once a photon is detected, a count is added to the time bin of the corresponding arrival-time measured by the TDC and so on for the following detected photons. More details about the number and the duration of the time bins of the temporal histogram are provided in the following chapters.

In the pixelated version of a time-resolving single-photon sensitive detector, each pixel of the camera has a single-photon sensitive detector and a temporal histogram which contains the arrival-time of the photons detected in the pixel-correlated portion of the field of view.
The SPAD operates in a photon starved-regime in order to collect a photons statistics uniformly distributed in time. A SPAD operates in the photon starved-regime when the number of detected photons is up to the 10% of the laser pulses. When the SPAD collects a photon for each laser pulse, the SPAD detects only the earliest photons, i.e. only the photons located at the first time-bins of the temporal histogram. When the SPAD operates in a photon starved-regime, it detects also the photons arriving later in time, obtaining an uniform statistics of the photons. By considering that only the 10% of the laser pulses contribute to the detection of a photon, the acquisition time required to collect a representative statistics is still suitable for real-time applications by using lasers with high repetition rate. In order to time the photon counting of the single-photon detectors used in the reported experiments, we used the SPC-150NX TCSPC card manufactured by Becker&Hickl [54].
We now introduce the single-pixel camera, a device that requires just one single-photon sensitive detector to create an image, representing a competitive and cheaper alternative to the pixelated cameras.

2.3 Single-pixel camera

Current 3D imaging techniques for LOS and NLOS scenes are based either on raster scanning systems or on acquiring the return signal by spatially resolved sensors. Scanning 3D imaging techniques are characterized by moving parts such as galvo-mirrors or rotating systems [55], resulting in bulky devices with limited frame rate whose proficiency scales inversely with the image resolution. On the other hand, multi-pixel detectors contain a separate sensor for each pixel of the image, resulting in expensive devices especially at wavelengths outside the visible range. Therefore, single-pixel cameras can offer a competitive alternative to the conventional cameras in terms of wavelength range, dark counts, temporal resolution and PDE.
Widely proposed in LiDAR [27, 56], 3D imaging [57, 58] and fluorescence microscopy [59], a

single-pixel camera is an imaging system composed by a programmable Spatial-Light-Modulator (SLM) and only a single-pixel light sensor. In a single-pixel camera device, the SLM maps the transverse spatial information of the scene by 2D structured masks, while a sensor with no spatial resolution collects the signal scattered back by the scene.
Multiplying a series of different applied patterns with the corresponding photon counts detected by the single-pixel detector, it is then possible to reproduce the 2D information of the entire scene using only a single-pixel sensor. Using a time-resolved single-pixel sensor, we can then retrieve the 3D information of the scene inferring the depth from the ToF information. The most common examples of SLMs are liquid-crystal devices (LCDs) or DMDs offering a modulation rate up to 20 KHz and mostly used in CI applications.

Fig. 2.3 (a) shows a schematic representation of a single-pixel camera device where a light source illuminates the SLM which in this case is chosen as a DMD. The DMD is typically made by 1024x768 square micromirrors of 16x16 μm^2 area. The individual orientation of each mirror can be electrostatically chosen according to user defined mask. With respect to the normal incidence, each mirror of the DMD can be tilted either by +24° or by -24° according to whether the mirror state is 1 or -1 respectively. By setting the state of each mirror, it is then possible to selectively redirect the light.
The scene is then structured illuminated projecting structure light patterns through the projection lens PL. According to the structured pattern applied on the DMD, only selected portions of the transverse plane of the scene are then illuminated. The single-pixel detector collects the light scattered back by the selected portion of the scene. Using selective patterns such as Hadamard masks, this method allows to reconstruct the entire image using only a one pixel sensor as opposed to multi-pixel detectors. Further details about the image reconstruction can be found in Section 4.2. The configuration using the DMD to project patterns of light illuminating the scene is referred to as "structured illumination".

Additionally, the single-pixel camera approach can also be used in "structured detection" configuration, as shown in Fig. 2.3 (b). In this case the light source flood illuminates the entire scene and the DMD images the scene through the imaging lens IL. In order to have the scene on focus on the DMD surface, the focal length of the IL is chosen according to the distance DMD-scene. By applying a user defined pattern, the DMD reflects the light scattered back by the selected transverse portions of the FoV onto the single-pixel detector. The suggested single-pixel approach can also be extended to the 3D retrieval of hidden scenes, as demonstrated in Chapter 4.

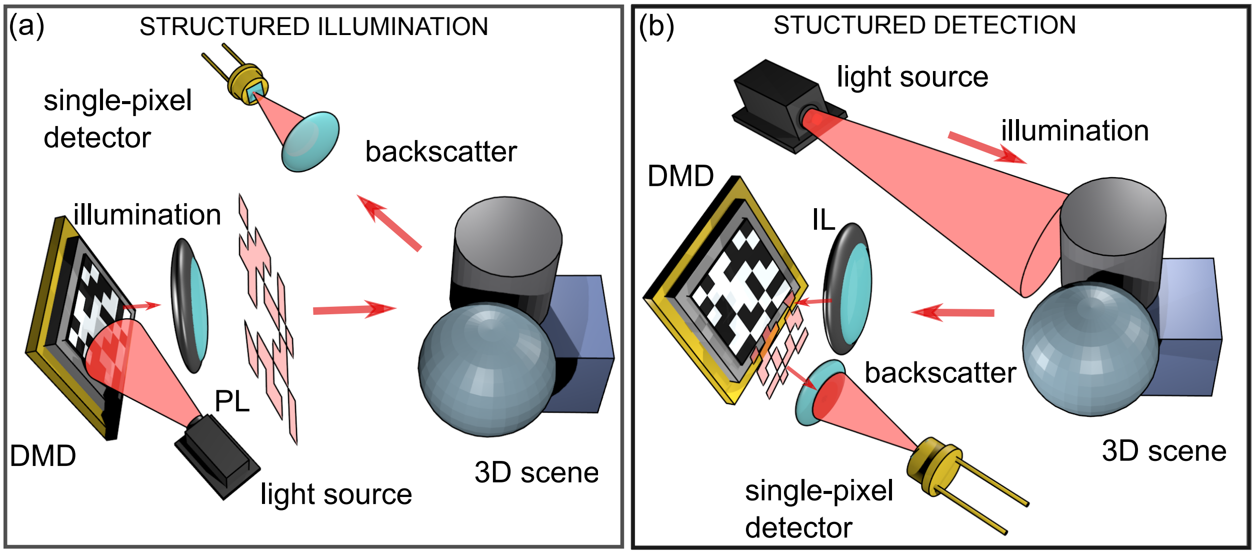

Figure 2.3: **Single-pixel camera.** The single-pixel camera consists of two main components: a SLM chosen in this case as a DMD, and a single-pixel detector. The transverse spatial information can be recovered correlating the user defined patterns applied on the DMD and the corresponding photon counts detected by the single-pixel sensor. a) Structured illumination configuration. The light source illuminates the DMD surface that in turns, projects structured illumination patterns on the scene by the projection lens PL. A single-pixel detector collects the light scattered back by the selected portions. b) Structured detection configuration. The light source flood-illuminates the scene and the DMD reflects the light scattered back by the scene selected portions onto the single-pixel detector.

Reducing the pixel complexity to a single unit, the single-pixel design offers several advantages in terms of size, flexibility, wavelength range and cost. Single-pixel detectors are indeed available for hyperspectral imaging [60] at a wide wavelength range outside the visible region where the pixelated counterpart is expensive. Single-pixel detector applications have been already demonstrated in the X-ray [61], infrared [58] and Terahertz [62] domain.

In addiction, single-pixel cameras offer an improved timing performance compared to the focal plane array counterpart, as demonstrated by the 17 picoseconds temporal resolution of the single-pixel PMT [45] described in Section 4.1.2. Applying patterns such as Hadamard masks where half of the pixels are always collecting signal, the average number of photons detected in a single-pixel camera measurement is N/2 times greater than in a pixelated sensor measurement [63].

Single-pixel detectors may provide enhanced performances, such as faster timing response, higher detection sensitivity and lower dark counts, providing a valid alternative to the pixelated counterpart. Single-pixel cameras offer a further advantage in terms of fill factor defined as the ratio between the sensitive area and the illumination area on the detector. The fill factor offered by single-pixel detectors is indeed close to 100%, against the 3% of single-photon sensitive cameras [24, 29].

The long acquisition time required for high spatial resolution images can be reduced applying CI algorithm such as compressive sensing (CS). CS considers the sparsity of the image in the frequency domain and allows the full reconstruction of the image acquiring only a selected subset of the full imaging reconstruction bases and reducing the number of measurements down to $log_2(N)$ for an N pixels image [64].

We now employ the single-photon sensitive, single-pixel technology discussed so far to simplify the hardware complexity of the current ToF 3D imaging techniques for LOS and NLOS scenes.

Chapter 3

Lock-in single-pixel camera for 3D imaging of line-of-sight scenes

In this chapter we investigate the 3D retrieval of a LOS scene combining a single-pixel camera with a lock-in amplifier. According to the simulations reported in this chapter, the lock-in amplifier theoretically provides a phase resolution of one order of magnitude better than conventional continuous wave (CW) modulated ToF cameras. However, the proposed method is affected by some limitations such as the number of measurements and the time resources not compatible with real-time applications.

3.1 Lock-in Time-of-Flight camera

Thanks to advances in microelectronics, microtechnologies and sensing [65], the 3D imaging of direct scenes at real-time frame rates attracted a great interest within the research community in several research fields. The recovery of depth resolved images indeed plays a crucial role in everyday-life situations with commercial applications in robot navigation [66, 67], machine vision [68, 69], autonomous vehicles [70, 71], 3D remote ranging [72] and modern photography systems [73].
Current ToF imaging technologies use either a pulsed light source or a CW intensity modulated beam to infer the depth information. Pulsed ToF systems recover the depth information by measuring the round-trip time a short pulsed signal takes to reach the targets and return to the detector [74]. Pulsed ToF cameras usually employ p-i-n structure sensors, avalanche photodiode (APD) detectors or SPADs operating in TCSPC mode. However, pulsed light 3D imaging technologies require short pulses and picosecond temporal resolution sensors to retrieve sub-centimetre depth resolved 3D images.

On the contrary, CW intensity modulated ToF cameras are eye-safe, compact and robust devices able to simultaneously provide the depth and the intensity images of direct scenes at real-time frame rate [75, 76]. CW modulation ToF cameras recover the 3D images by emitting a sinusoid amplitude modulated signal and measuring the amplitude of the return with Charge-coupled device/CMOS (CCD/CMOS) sensors [77–79]. One of the most common and miniaturized example is the Swissranger SR3000 camera [76], a CW modulated ToF camera based on high sensitive solid-state CCD/CMOS lock-in pixelated sensors [78, 80] and low-cost NIR light-emitting diodes. In CW intensity modulated 3D camera systems, the transverse information of the investigated scene is encoded in the pixelated structure of the camera's sensor in which each pixel detects the photons scattered from a specific (x-y) portion of the transverse plane. The depth information z is encoded in the phase difference $\phi(x,y)$ between the emitted and the received signal. Known as four bucket sampling [77], the phase difference $\phi(x,y)$ is measured by sampling the intensity of return beam four times at equally distributed intervals within the oscillation period. The depth information is then inferred by the formula

$$z(x,y) = \frac{c\phi(x,y)}{4\pi f_m} \tag{3.1}$$

where c is the speed of light and f_m is the modulation frequency of the emitted beam.
Compared to other 3D imaging technologies, conventional CW modulation ToF cameras offer high frame-rate, low power consumption, quick and easy data extraction, compactness and portability. Moreover they simultaneously provide the depth and the intensity image without requiring any moving or scanning parts.
Despite the robustness, the portability and the easy data-extraction of CW modulated 3D cameras, some limitations affect their applicability. Since the phase is inferred by sampling over four equally distributed points within the entire oscillation period, they offer a limited depth resolution in the 0.5-1 cm range. Moreover, the ambiguity-free distance limits the depth range to 6 metres. Due to the methodology used to measure the depth, the targets whose phases differ 360° are indeed undistinguishable. For that reason, standard 3D cameras may have limited practicality in long-range scenarios or in 3D imaging with sub-centimetre depth resolution.

An alternative device capable to isolate the signal of interest from a noisy background is the lock-in amplifier, a device which exploits the information on the signal's time dependence [81]. A lock-in amplifier is an electronic device capable of isolating from a noisy background a signal oscillating in time within a user-defined bandwidth f_{BW} around a reference frequency. Known as phase demodulation, lock-in amplifiers multiply the acquired signal with an in-phase and a 90° out-of-phase copy of the reference signal. A low-pass filter is then applied to extract the phase and the amplitude of the signal. Mostly used for sensing applications [82], atomic force microscopy [83] and Hall effect measurements [84], lock-in amplifiers accurately operate in noise level up to a million times higher than the signal [1].

Since the phase is measured by continuously sampling over the entire oscillation period, the use of lock-in amplifiers offers a considerable advantages in the depth resolution. As demonstrated in the simulation present in this chapter, the phase resolution obtained by continuously sampling over the entire oscillation period is indeed one order of magnitude better than the one obtained by the four points sampling of a standard CW ToF approach. Additionally, no fixed ambiguity-free distance limits the depth range of a lock-in amplifier, representing a valid alternative to conventional 3D cameras. The ambiguity-free range of a lock-in amplifier can indeed be tuned according to the modulation frequency f_m of the reference signal.

Here, we demonstrate the depth-image recovery of a LOS 3D scene by combining a lock-in amplifier and a single-pixel camera. The suggested method provides a full 3D retrieval of a LOS scene with a depth resolution of 5 mm without requiring short pulses light source, picosecond temporal resolution detectors or time-correlated single-photon counting electronics. Moreover the suggested method provides a tunable ambiguity-free distance according to the reference modulation frequency, representing an alternative to conventional 3D cameras for long-range scenarios. However, some limitations affect the suggested approach, such as the prohibitive acquisition time required for high resolution images. Additionally, the proposed method results in a bulky device composed by a DMD, a lock-in amplifier and a single-pixel detector, reducing the versatility and the flexibility of the proposed technology.

3.2 Principles of lock-in detection

Here, we describe how the phase demodulation process isolates the amplitude and the phase of the signal oscillating within a given bandwidth f_{BW} around the reference modulation frequency f_m. In the phase demodulation process, the lock-in amplifier selects the signals oscillating within a user-defined bandwidth around a reference frequency ω_r defined by an external reference beam. The demodulation method of a lock-in amplifier is depicted in Fig. 3.1 where $V_r(t)$ is the reference beam, and $V_s(t)$ is the input noisy signal. The lock-in amplifier selects the signal oscillating within a given frequency bandwidth by multiplying the input signal with an in-phase and 90° out-of-phase copy of the reference signal. It then applies an adjustable low-pass filter to efficiency reject the undesired frequencies. Figure 3.2 shows the block diagram of the phase demodulation.

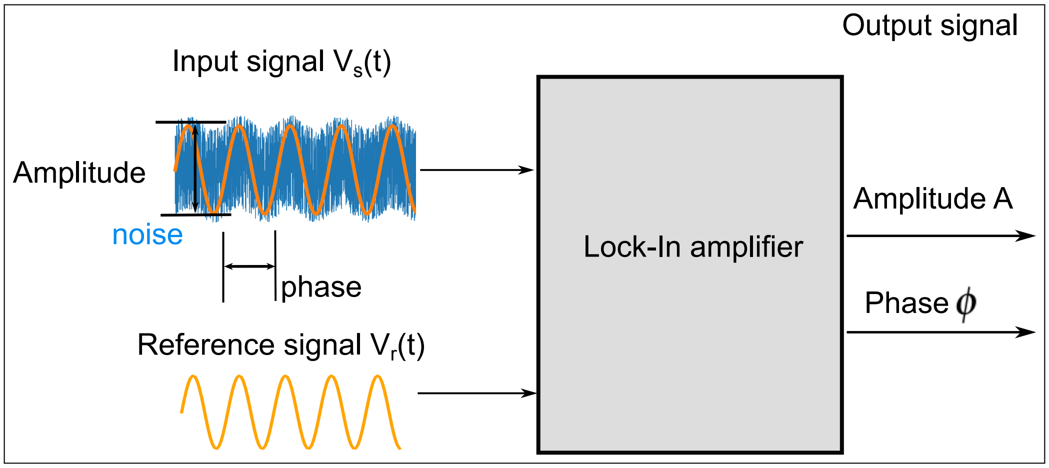

Figure 3.1: **Demodulation method of a lock-in amplifier.** The reference beam $V_r(t)$ is a sinusoid signal oscillating at the reference frequency ω_r, whereas $V_s(t)$ is the noisy input signal. The lock-in amplifier selects only the signals oscillating within a user defined bandwidth around a given frequency ω_r defined by the reference signal. All the remaining signals oscillating at a frequency outside the defined bandwidth are rejected. It then extracts the amplitude A and the phase ϕ of the desired signal.

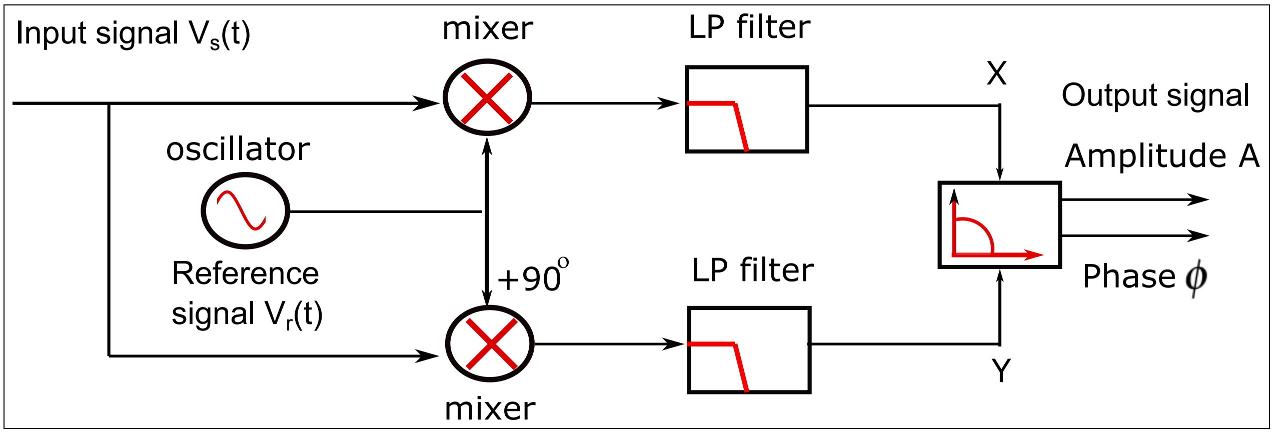

Figure 3.2: **Block diagram of a lock-in detection system.** The lock-in amplifier multiplies the input signal with a in-phase reference signal and a 90° out-of-phase reference to extract the amplitude A and the phase ϕ of the input signal $V_s(t)$. A low-pass filter is then applied to separate the constant and the periodic component. The required amplitude and phase are obtained by combining the two perpendicular components X and Y obtained at the output of the lock-in amplifier.

We then mathematically describe the demodulation process. In order to demodulate the input signal with two different phase signals, an in-phase and a 90° phase-shifted copy of the reference signal are created. We therefore consider a typical periodic reference signal $V_r(t)$ with defined amplitude R

$$V_r(t) = Rsin(\omega_r t) \tag{3.2}$$

and its 90° out-of-phase copy

$$V_{90^\circ r}(t) = Rcos(\omega_r t) \tag{3.3}$$

where ω_r is the reference oscillation frequency.

In order to extract the phase and the amplitude of the desired signal we consider the following periodic input signal:

$$V_s(t) = Asin(\omega_r t + \phi) \tag{3.4}$$

oscillating at the same frequency ω_r of the reference beam. According to the demodulation method, we then multiply the input signal $V_s(t)$ by the in-phase and the 90° out-of-phase reference contributions, obtaining the following two signals $Y_0(t)$ and $Y_{90^\circ}(t)$:

$$Y_0(t) = Rsin(\omega_r t) \cdot Asin(\omega_r t + \phi) = \frac{RA}{2}[cos(\phi) - cos(2\omega_r t)] \tag{3.5}$$

$$Y_{90^\circ}(t) = Rcos(\omega_r t) \cdot Asin(\omega_r t + \phi) = \frac{RA}{2}[sin(\phi) + sin(2\omega_r t)] \tag{3.6}$$

Two distinct contributions compose the signals in Eqs. (3.5)-(3.6). The first contributions $\frac{RA}{2}cos(\phi)$ and $\frac{RA}{2}sin(\phi)$ are constant in time, whereas the second contributions oscillate at a frequency two times faster than the reference signal. These last two terms $\frac{RA}{2}cos(2\omega_r t)$ and $\frac{RA}{2}sin(2\omega_r t)$ respectively the $Y_0(t)$ and $Y_{90^\circ}(t)$ signal. Figure 3.3 shows the frequency spectrum of the two signals in Eqs. (3.5)-(3.6).

We then separate the constant from the oscillating component by applying a low pass filter (blue colour dashed line in Fig. 3.3) to both signals. We obtain the following filtered components Y_{0F} and $Y_{90^\circ F}$ at the output of the amplifier:

$$Y_{0F} = \frac{RA}{2}cos(\phi) \qquad Y_{90^\circ F} = \frac{RA}{2}sin(\phi) \tag{3.7}$$

After reformulating the previous formulae as

$$\sqrt{{Y_{0F}}^2 + {Y_{90^\circ F}}^2} = \frac{RA}{2} \qquad \frac{Y_{90^\circ F}}{Y_{0F}} = \frac{sin(\phi)}{cos(\phi)} \tag{3.8}$$

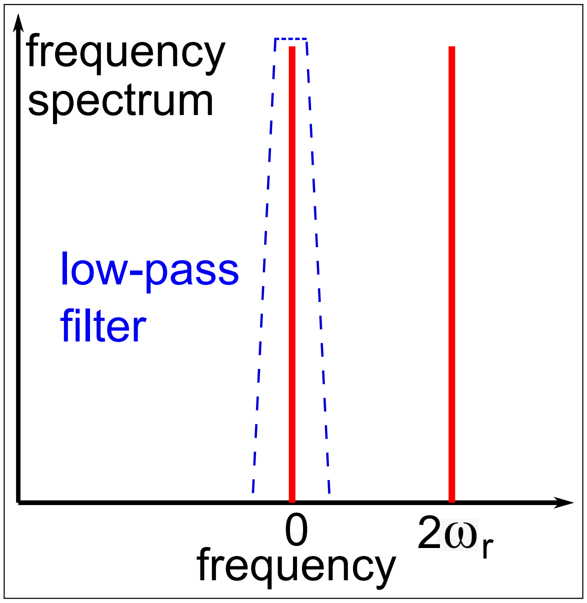

Figure 3.3: **Frequency spectrum of the phase demodulation of a periodic signal.** The demodulated signal is composed by a constant component and a periodical component oscillating at double the frequency ω_r of the input signal. We then applying a low-pass filter (blue colour dashed line) to extract the signal.

we obtain the amplitude A and the phase ϕ of the signal of interest V_s in Cartesian components:

$$A = \frac{2}{R}\sqrt{{Y_{0F}}^2 + {Y_{90^\circ F}}^2} \qquad \phi = tan^{-1}\frac{Y_{90^\circ F}}{Y_{0F}} \tag{3.9}$$

We now consider an input signal $V_s(t)$ composed by the signal to be extracted $Asin(\omega_r t + \phi)$ and a random noise component $Zsin(\omega_z t + \phi_z)$ described as follows:

$$V_s(t) = Asin(\omega_r t + \phi) + Zsin(\omega_z t + \phi_z) \tag{3.10}$$

Here, $\omega_z \neq \omega_r$ is the oscillating frequency of the noise component.
We apply the same demodulation method to the input signal described in Eq.(3.10). Since all the mathematical operations performed in the demodulation process are linear, the resulting signal is the sum of the two contributions $Asin(\omega_r t + \phi)$ and $Zsin(\omega_z t + \phi_z)$. In this case, the first component remains the same of the constant component described in Eq. (3.7), whereas the second component produced by the noisy background is:

$$Y_{0Z}(t) = Rsin(\omega_r t) \cdot Zsin(\omega_z t + \phi_z) = \frac{RZ}{2}[cos((\omega_r - \omega_z)t - \phi_z) - cos((\omega_r + \omega_z)t + \phi_z)] \tag{3.11}$$

$$Y_{90^\circ Z}(t) = Rcos(\omega_r t) \cdot Zsin(\omega_z t + \phi_z) = \frac{RZ}{2}[sin((\omega_r + \omega_z)t + \phi_z) - sin((\omega_r - \omega_z)t - \phi_z)] \quad (3.12)$$

The obtained signal is composed by a constant component (oscillating in time at a zero frequency), and multiple periodical components oscillating at $(\omega - \omega_z)$, $(\omega + \omega_z)$ and 2ω.
Both the signals in Eqs. 3.11-3.12 are periodic and oscillate at a non zero frequency. We then apply a low pass filter to reject the oscillating components without effecting the constant contribution of Eq. (3.7). Figure 3.4 shows the corresponding frequency spectrum for a noisy input signal.

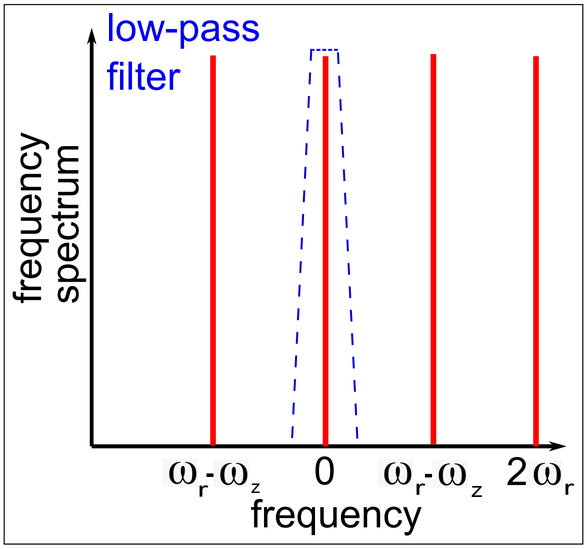

Figure 3.4: **Frequency spectrum of the phase demodulation of a noisy periodic signal.** The phase demodulated signal is composed by a constant component (oscillating in time at a zero frequency) and multiple periodical components oscillating at $(\omega_r - \omega_z)$, $(\omega_r + \omega_z)$ and $2\omega_r$. We then apply a low-pass filter centred at the reference frequency ω_r (blue colour dashed line) to extract the signal of interest.

After applying the low-pass filter, we then obtain the same filtered signal described in Eq. (3.9). We can then select the frequency band by setting the bandwidth of the low-pass filter. In this way we reject the other frequency signals improving the signal to noise ration of the detection system. As an example of the synchronous detection, Fig. 3.5 shows the temporal profile of the amplitude of the signal passing through the lock-in amplifier before and after the wave-mixing for $\omega_s = \omega_r$ (a-b) and $\omega_s \neq \omega_r$ (c-d).We then mathematically describe the demodulation process. In order to demodulate the input signal with two different phase signals, an in-phase and a 90° phase-shifted copy of the reference signal are created.

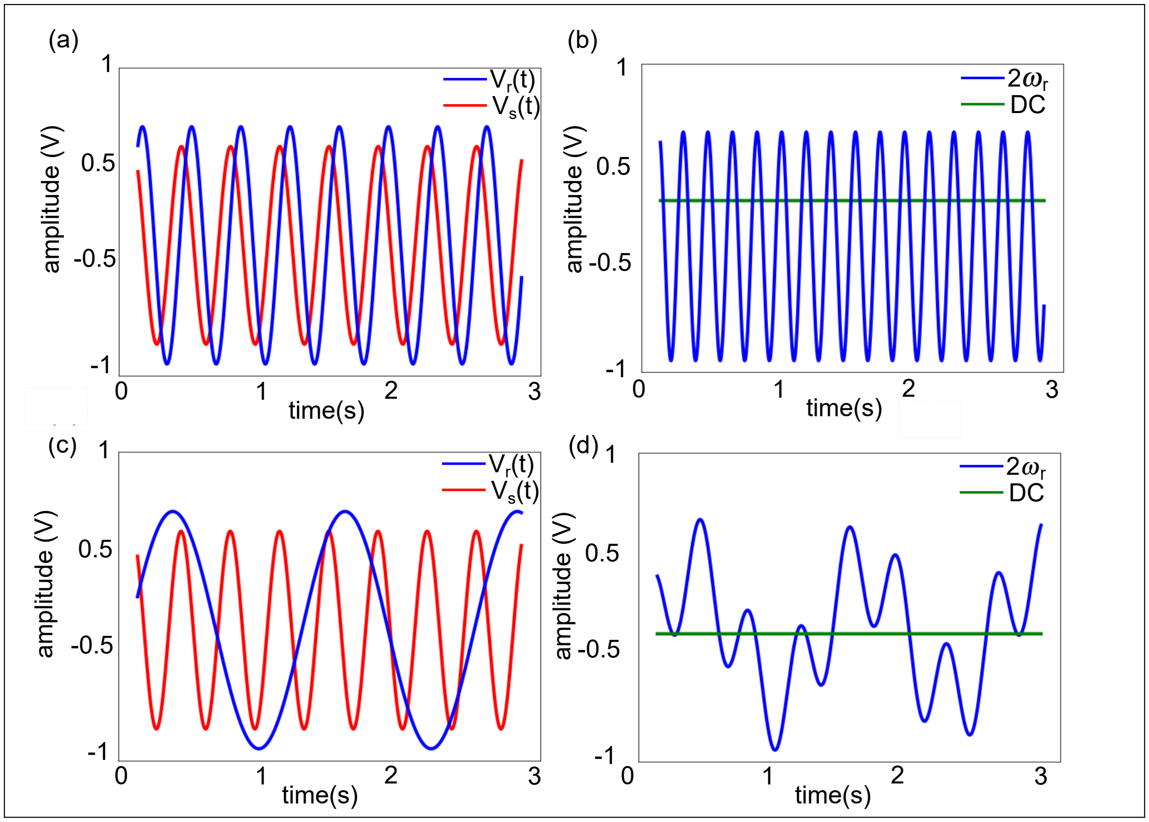

Figure 3.5: **Amplitude profile in time of periodic signals through synchronous detection.** (a) Temporal profile of the amplitude of a reference $V_r(t)$ and of an input $V_s(t)$ signal oscillating at the same frequency $\omega_s = \omega_r$. (b) After mixing the two signals, the signal is composed by a constant component (green line) and a periodic component oscillating twice times faster than the input frequency ω_r (blue component). After applying the low-pass filter, the DC component is isolated and the periodic component is rejected. (c) Temporal profile of the amplitude of the reference $V_r(t)$ and of the input $V_s(t)$ signal oscillating at two different frequencies $\omega_s \neq \omega_r$. (d) After mixing the two signals, the signal is composed by two periodic components (blue line) oscillating at $\omega_r \pm \omega_s$. After applying the low-pass filter, the average signal is zero and no signal is detected at the output of the amplifier.

When the input $V_s(t)$ and the reference signal $V_r(t)$ oscillate at the same frequency (Fig. 3.5 (a)), the output signal is composed by a constant component (green line) and a periodic component (blue line) oscillating twice times faster than the reference component ω_r (Fig. 3.5(b)). After applying the low-pass filter, only the DC component is preserved, corresponding to the 90° phase-shift of the output signal.

When the input $V_s(t)$ and the reference signal $V_r(t)$ oscillate at different frequencies (Fig. 3.5 (c)), the signal obtained after the mixing is composed by two periodic components (blue line) oscillating at $\omega_r \pm \omega_s$ (Fig. 3.5(d)). After applying the low-pass filter, the resulting average signal is zero and no signal is detected at the output. The lock-in detection extracts only the periodic signal oscillating at the same frequency of the reference signal and it thus represents a perfect example of synchronous detection.

A brick-wall filter should be applied to select all the frequencies within a given bandwidth f_{BW} around the reference frequency. Indeed, a brick wall transmits all the frequencies below the bandwidth f_{BW} and rejects all the remaining frequencies. Since ideal brick-wall filters perfectly selecting only the desired frequencies are impossible to be realized, we approximate a low-pass filter by a RC filtering model (Fig. 3.6).

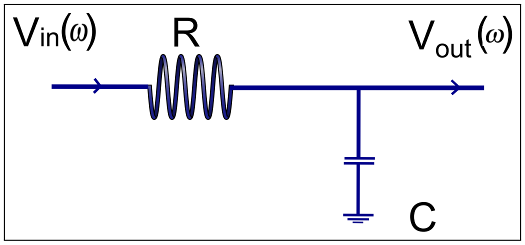

Figure 3.6: **Schematic of the RF filter model.** Since an ideal brick-wall filter is impossible to be realized, we consider a RC filter model to select the frequencies within a given range around the reference signal. The RC circuit transmits periodic signals whose frequency is within the bandwidth $f_{BW} = 1/2\pi\tau$ and rejects all the remains frequencies.

The RC filter model is characterized by the following transfer function defined as the ratio between the transmitted and the input signal power:

$$H(\omega) = \frac{1}{1 + i\omega\tau} \tag{3.13}$$

where ω is the angular frequency and $\tau = RC$ is the time constant with resistance R and capacitance C. We define the cut-off frequency bandwidth f_{BW} as the frequency at which the transmitted signal power is attenuated by -3dB (or by half). According to Eq. (3.13), the f_{BW} of an RC is inversely proportional to the time constant τ as follows:

$$f_{BW} = \frac{1}{2\pi\tau} \tag{3.14}$$

However, the RC filter model defined by Eq. (3.13) has poor performance compared to ideal brick-wall low-pass filters. Therefore, we add multiple RC filter models in series to improve the performance of a single RC filter model in order to select the desired frequency signals more efficiently.

Since the multiple RC filters are connected in series, the resulting transfer function of a n order RC filter is:

$$H_n(\omega) = \big(H_1(\omega)\big)^n = \big(\frac{1}{1+i\omega\tau}\big)^n \tag{3.15}$$

The cut-off bandwidth of multiple RC filters is

$$f_{BW} = \frac{OF}{2\pi\tau} \tag{3.16}$$

where the order factor OF depends on n, that is the number of the RC filters. We select the periodic signals oscillating at a frequency within a desired bandwidth f_{BW} setting the value of the time constant τ and of the filter order n.

The choice of the order factor n plays a fundamental role in the lock-in detection, determining the bandwidth f_{BW} and the shape of the cut-off frequency. A wide filter bandwidth f_{BW} around the reference frequency could lead to noisier measurements and lower signal to noise ratio. In this case, the undesired frequencies such as the $2\omega_r$ component could leak within the selected threshold, inducing a systematic error on the output signal. On the other hand, a narrow bandwidth guarantees the selection of the desired frequencies only and a higher signal to noise ratio at the cost of lower time resolution. Indeed, the low-pass filtering induces a phase delay increasing with the filter order and equal to the argument of the transfer function of Eq. (3.13). Therefore, higher phase delays require longer settling time to obtain accurate measurements.

Figure 3.7(a-b) shows the Bode plots of the frequency responses of the lock-in amplifier as a function of the filter order [1]. Figure 3.7 shows the attenuation of $n = 1,2,4,5,8$ filter orders lock-in systems for a fixed time constant τ. Higher filter orders produce a narrower bandwidth and a transfer function more similar to a brick-wall filter.

Figure 3.7(b) shows the Bode plots of the frequency response for a fixed bandwidth f_{BW} but different time constant τ for $n = 1,2,4,8$ filter orders. In this case, low-pass filters with the same frequency bandwidth but higher filter orders are characterized by a steeper roll-off at high frequencies. Figure 3.7(c) reports the settling time in time constant units required to achieve phase measurements with 99% of accuracy for filter order $n = 1,2,4,8$.

Tab. 3.1 reports the value of the order factor OF, the bandwidth f_{BW} and the settling times with a phase precision of 99.9 % for filters order $n = 1,2,4,8$ [1].

We then evaluate the advantages of using a phase demodulation single-pixel camera approach in depth resolving measurements.

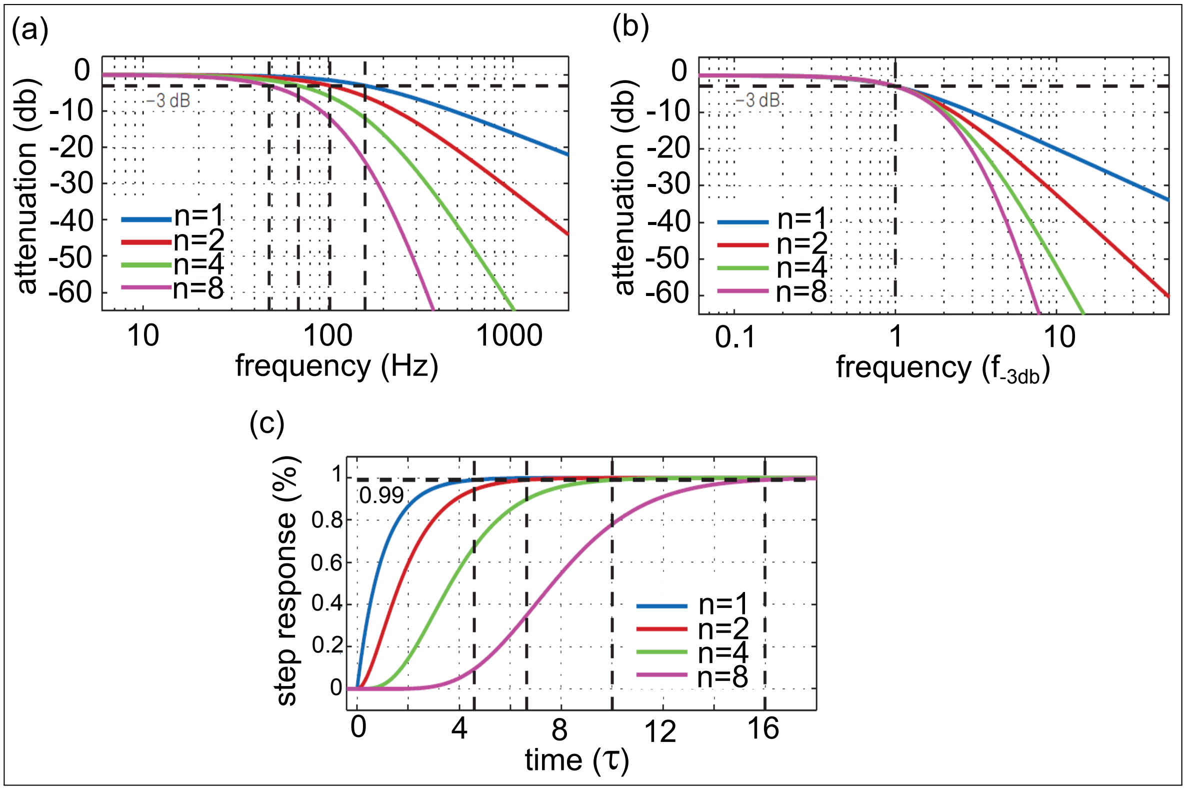

Figure 3.7: **Bode plots of a lock-in detection system for different filter orders.** (a) Frequency responses of a lock-in amplifier for filter order $n = 1, 2, 4, 8$ with fixed time constant and varying f_{BW}. The blue trace represents the frequency response of a first order filter model. By increasing the filter order (red, green and purple traces), the frequency bandwidth of the accepted frequencies is narrower. (b) Frequency responses of a lock-in amplifier for filter order $n = 1, 2, 4, 8$ with fixed frequency bandwidth f_{BW} and varying time constant τ. The blue trace represents the frequency response of a first order filter model. By increasing the filter order, the low-pass filter bandwidth has a steepest roll-off at high frequencies. (c) Percentage step responses of a lock in amplifier as a function of the time constant τ for filter orders $n = 1, 2, 4, 8$. In order to overcome the increasing phase delay induced by higher filter orders, the lock-in detection requires longer settling time, limiting the temporal resolution and slowing down the measurements.

Order	OF $(1/\tau)$	Settling times (τ)	$f_{BW}(1/\tau)$
1	1	6.91	0.159
2	0.64	9.23	0.102
4	0.43	13.06	0.069
5	0.38	14.79	0.060
8	0.30	19.62	0.048

Table 3.1: Order factors, settling times and cut-off bandwidth of a lock-in amplifier for different order filters [1].

3.2.1 Phase error comparison.

In order to compare the standard ToF camera with the lock-in amplifier, we compare the error over the phase of the two 3D imaging technologies. We thus numerically simulate a sinusoid amplitude modulated signal with a Gaussian distributed noise with average μ=5 and standard deviation σ. Figure 3.8 (a) shows the temporal profile of the sinusoid amplitude modulated signal oscillating at 1 kHz frequency, whereas Fig. 3.8 (b) shows the sinusoid amplitude modulated signal with a Gaussian distributed random noise with average $\mu = 5$ and standard deviation σ=0.1. We then compute the standard deviation of the phase obtained by the standard ToF and by the lock-in detection approach as a function of the standard deviation σ of the noise component.

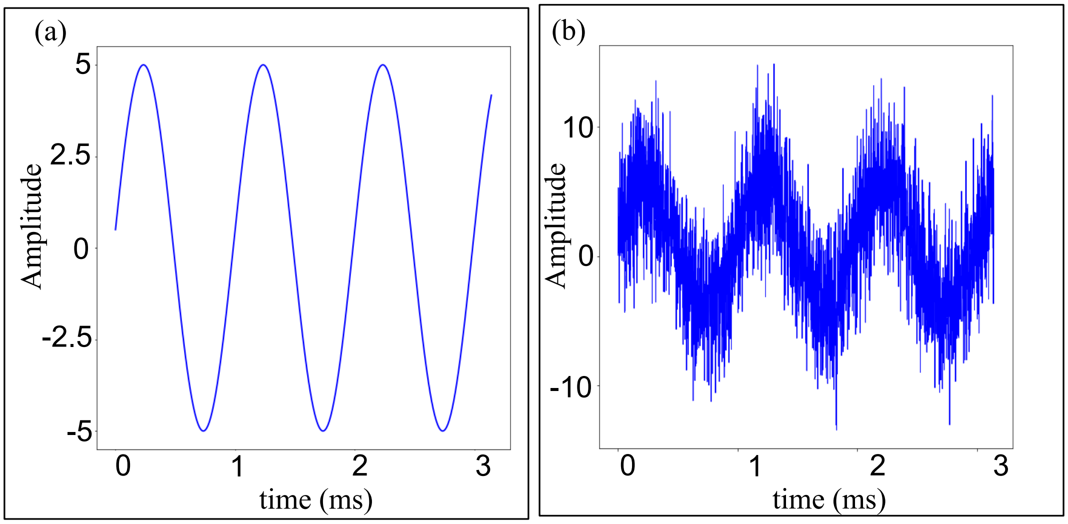

Figure 3.8: **Temporal profile of a sinusoid amplitude modulated signal.** (a) Sinusoid amplitude modulated signal oscillating at 1 kHz frequency. (b) Sinusoid amplitude modulated signal with a Gaussian distributed random noise with average $\mu = 5$ and standard deviation σ=0.1.

The phase ϕ of the standard ToF camera approach is computed sampling the signal amplitude A_i over four points $0, 1, 2, 3$ equally distributed along the oscillation period:

$$\phi = \frac{A_3 - A_1}{A_0 - A2} \tag{3.17}$$

In contrast, the phase of the lock-in amplifier approach is computed sampling over the entire oscillation period according to Eq. (3.9). In more details, we multiply the noisy periodical signal with the in-phase and the 90° out-of-phase reference beam, obtaining the two components $Y_0(t)$ and $Y_{90^\circ}(t)$. We then sum over the oscillation period to select only the constant component of the two signals as happens in a low-pass filtering. The phase is then computed by the inverse tangent of the ratio of the perpendicular components according to Eq. 3.9. The phase error $\Delta\phi(\sigma)$ is obtained by the standard deviation of the phase of 50 Gaussian distributed noise signals for each value of σ. Figure 3.7 shows the sinusoid amplitude modulated signal oscillating at 1 kHz frequency without (a) and with (b) the Gaussian distributed noise component.

Figure 3.9 shows the phase error $\Delta\phi(\sigma)$ as a function of the standard deviation σ of the Gaussian distributed random noise by a conventional ToF camera (red trace) and by a lock-in amplifier (blue trace). As reported in the figure, the error on the phase retrieved by the four points sampling approach increases 27 times faster than the phase retrieved by the entire oscillating period sampling. The suggested lock-in detection then provides a phase resolution 27 times more accurate than the four points sampling approach of a standard ToF camera.

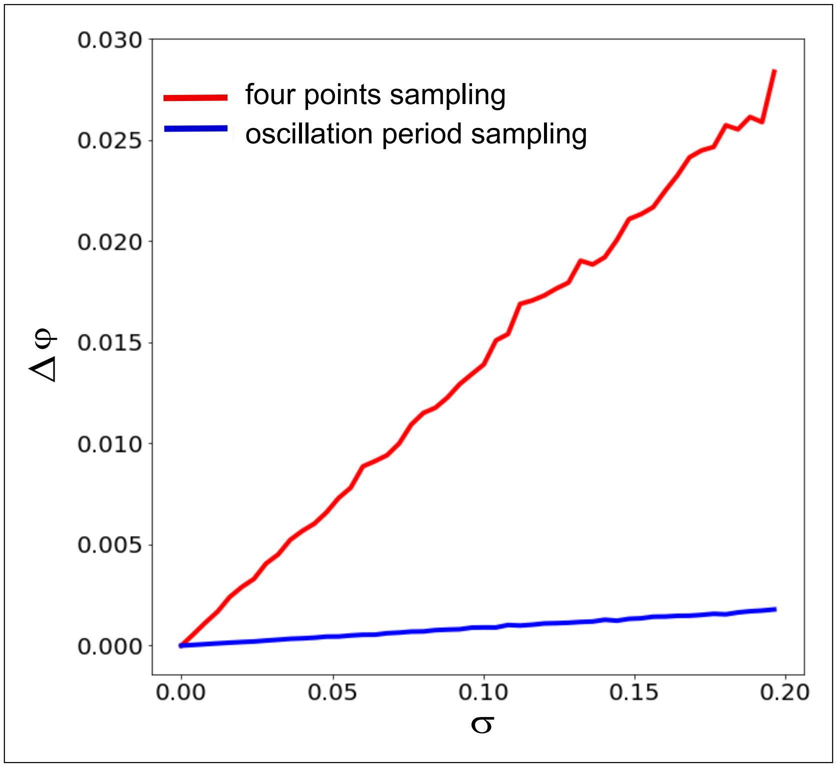

Figure 3.9: **Theoretical comparison of the phase error $\Delta\phi$ between a four points and an entire oscillation period sampling.** Trend of phase error as a function of the standard deviation σ of the Gaussian random noise signal. The red trace represents the error of the phase retrieved with the four points sampling approach. The blue trace represents the error of the phase retrieved with sampling over the entire oscillation period. The lock-in approach provides a phase error 27 times better than the four point sampling. The standard deviation of the phase is computed averaging 50 Gaussian random noise signals for each value of σ.

3.3 Lock-in single-pixel camera

The scene to be recovered is flood illuminated with an intensity modulated beam oscillating at a reference modulation frequency f_m and a single-pixel camera combined with a lock in amplifier collects the return signal. Figure 3.10 shows the experimental setup of the lock-in single-pixel camera. Here, the sinusoid modulated amplitude beam is created from a continuous wave (CW) beam by an electro-optic modulator (EOM). In more details, a CW linearly polarized Gaussian beam of wavelength λ= 532 nm and power P=150 mW propagates through the optical system shown in Fig. 3.10 (a). The beam outgoing from the laser is collimated by a set of lenses not shown in the figure. The half-waveplate and the polarizing beam splitter (PBS) set the beam power at 130 mW. The beam passes through an electro-optic amplitude modulator (EOM) in order to produce a laser beam whose amplitude is sinusoid modulated in time at a reference frequency. A converging lens L4 of focal $f = 100$ mm focuses the beam into the EOM. Acting as a variable waveplate, the EOM changes the polarization state of the laser according to an external driving voltage. We then change the temporal modulation of the beam amplitude by placing at the EOM output a linear polarized (LP) whose axis is perpendicular to the input beam polarization. We then apply a sinusoid amplitude modulated external voltage by a programmable function generator. Since the EOM requires a voltage of 170 V to induce a 90° change in the beam polarization, we amplify the 10 V amplitude sinusoid signal. The bi-convex lens L5 of focal length f=100 mm collimates the diverging beam outgoing from the EOM. The sinusoidal beam at the output of the EOM oscillates in time at a modulation frequency of 5 MHz.

Figure 3.10 (b) shows the lock-in single-pixel camera setup. A sinusoid amplitude modulated beam is flood-illuminating a 45x45 cm^2 3D scene by the bi-concave lens L1 of focal length f=-50 mm. The return signal is collected from a single pixel camera by the lens L2 (focal length f=50 mm). A DMD and a Silicon photomultiplier detector (SiPM, C14455-3050GA) compose the single-pixel camera. The SiPM detector (or single channel Multi-Pixel Photon Counters, module C14455-3050GA) is manufactured by Hamamatsu Photonics and it has an effective photosensitive area of 3 mm of diameter and a PDE of 34% at 532 nm. All the 2836 the pixels of the SiPM are connected to the unique analog output of the single channel SiPM. The DMD is manufactured by Texas Instruments and it has 1024x768 micromirrors of 16x16 μm^2.
A 50 mm focal length lens (L3) collimates the beam after the DMD and a long working distance microscope objective (magnification factor 50X, Mitutoyo Plan Apo) focuses the transmitted beam on a free-running SiPM detector. In order to isolate the desired signal from the background light of external sources present in the scene, we use a 532 nm band-pass filter with a 40 nm bandwidth

before the detector. A lock-in amplifier extracts the desired signal from the analog output of the SiPM sensor according to the reference signal. The reference signal is provided by a copy of the electric signal generated with the programmable function generator. We acquire the return signal by 64 x 64 pixels raster scan masks on the DMD and collect the reflected light onto the detector. Each raster scan mask collects the return signal scattered back from a 0.7x0.7 cm^2 transverse portion of the field of view for an overall imaging area of 45x45 cm^2. The masks projection on the DMD and the lock-in amplifier acquisition are synchronized by a Matlab software code.

In order to evaluate the depth resolution of the proposed method, the scene is composed by a 24x24 cm^2 square target shown in Fig. 3.11(b). The target consists of a series of 6x6 or 12x12 cm^2 squares, 6x12 cm^2 rectangles and a "T" letter. The targets are placed at varying depth within a range of 0-22 mm in order to evaluate the depth resolution of the suggested method. In details, the targets are located at an increasing relative depth of 0, 5, 6, 7, 10, 11, 12, 14, 16, 17, 21, 22 mm respect to the farthest target.

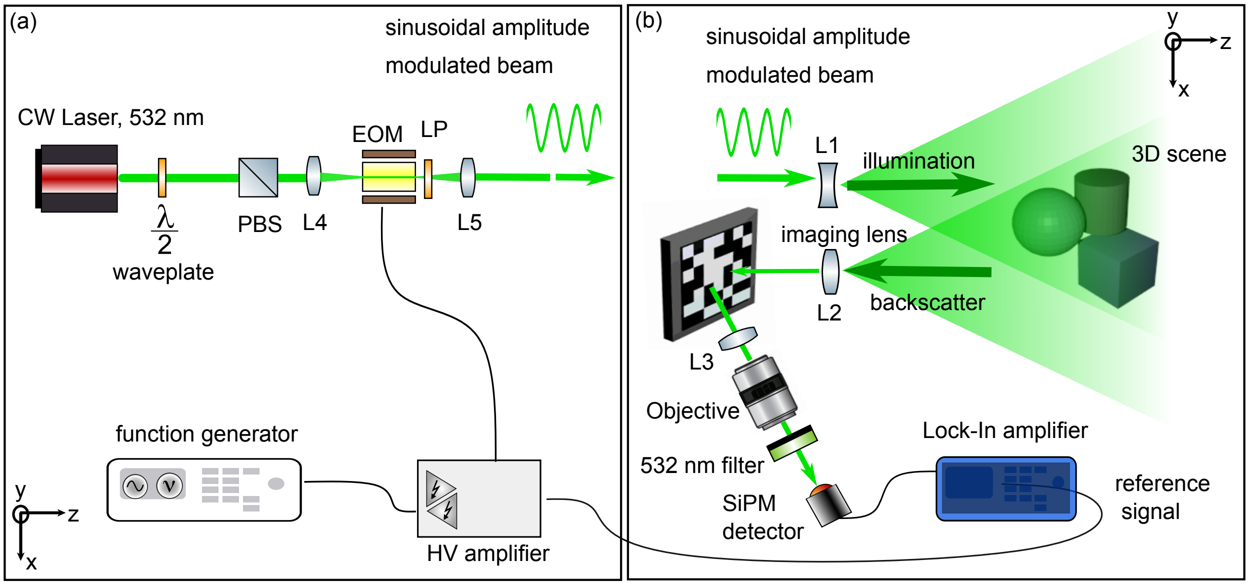

Figure 3.10: **3D imaging by a lock-in single-pixel camera.** a) A CW beam passes through an electro-optic modulator (EOM) to obtain a sinusoid modulated amplitude beam. The EOM acts as a Pockels cell modulator that varies the polarization of the beam linearly with the applied voltage. A copy of the electric signal produced by the function generator is provided in input to the lock-in amplifier as a reference. b) A sinusoid amplitude modulated beam flood-illuminates the 3D scene and a single-pixel camera composed by a SiPM detector and a DMD collects the signal scattered back from the 45x45 cm^2 field of view. The lock-in amplifier then filters the detected signal according to the reference signal and provides the in-phase and the 90° out-of-phase component of the return at each acquisition.

In order to set the lock-in cut-off frequency and efficiently isolate the return signal, we use a filter order $n = 5$ and a time constant $\tau = 100$ ms for a 5 MHz modulation frequency. These values correspond to a settling time of 1.48 s and a cut-off frequency bandwidth $f_{BW} = 0.60$ Hz, as reported in Tab. 3.1. In this case the acquisition time of the lock-in poll duration is 0.1 seconds per mask at a sampling rate of 132. The lock-in amplifier provides in output the in-phase and the 90° out-of-phase component (Eqs. 3.9) of the return at each acquisition. We then proceed with the 3D retrieval of the scene following the procedure described in Section 3.2.

3.4 Experimental results

The in-phase and the 90° out-of-phase component of the return provide the 3D information of the scene depicted in Fig. 3.11 (b). The square root of the quadratic sum of the in-phase and the 90° out-of-phase component provide the x-y information of the scene, as described in 3.9 (a). The depth information z is then computed by considering the 2π change of the phase ϕ of a modulated beam over a modulation wavelength distance. The z information is then expressed as

$$z(mm) = \lambda \frac{\phi}{2\pi} \tag{3.18}$$

where λ is the wavelength of the modulation frequency and ϕ is the measured phase described in Eq. (3.9) (b). Since the modulation frequency is 5 MHz, the previous depth can be rewritten as follows:

$$depth(mm) = \frac{3 * 10^5}{5} \frac{\phi}{2\pi} \tag{3.19}$$

In order to take into account the offset over the measured phase, we consider the relative phase difference between each (x-y) pixel and the farthest measured pixel centred at coordinates (x=35 cm, y=38 cm, z=0).

Figure 3.11 shows the experimental results. Figure 3.11 (a) shows the amplitude of the return signal on the x-y plane. To facilitate the visualization a threshold of 0.83 is applied. The blue dotted line in Fig. 3.11 indicates the actual position of the target on the transverse plane. The inverse tangent of the ratio of the Y_{90° and Y_0 component provide the phase of the return. The depth is then computed by Eq. 3.19.

Figure 3.11 (c) shows the experimental depth. In order to evaluate the depth accuracy of the retrieval, Fig. 3.11(d) shows the actual depth of the target as a colour-encoded depth image in a 0-20 mm depth-range. Comparing the experimental retrieval with the ground truth, the proposed method provides a 5 mm depth accuracy 3D retrieval.

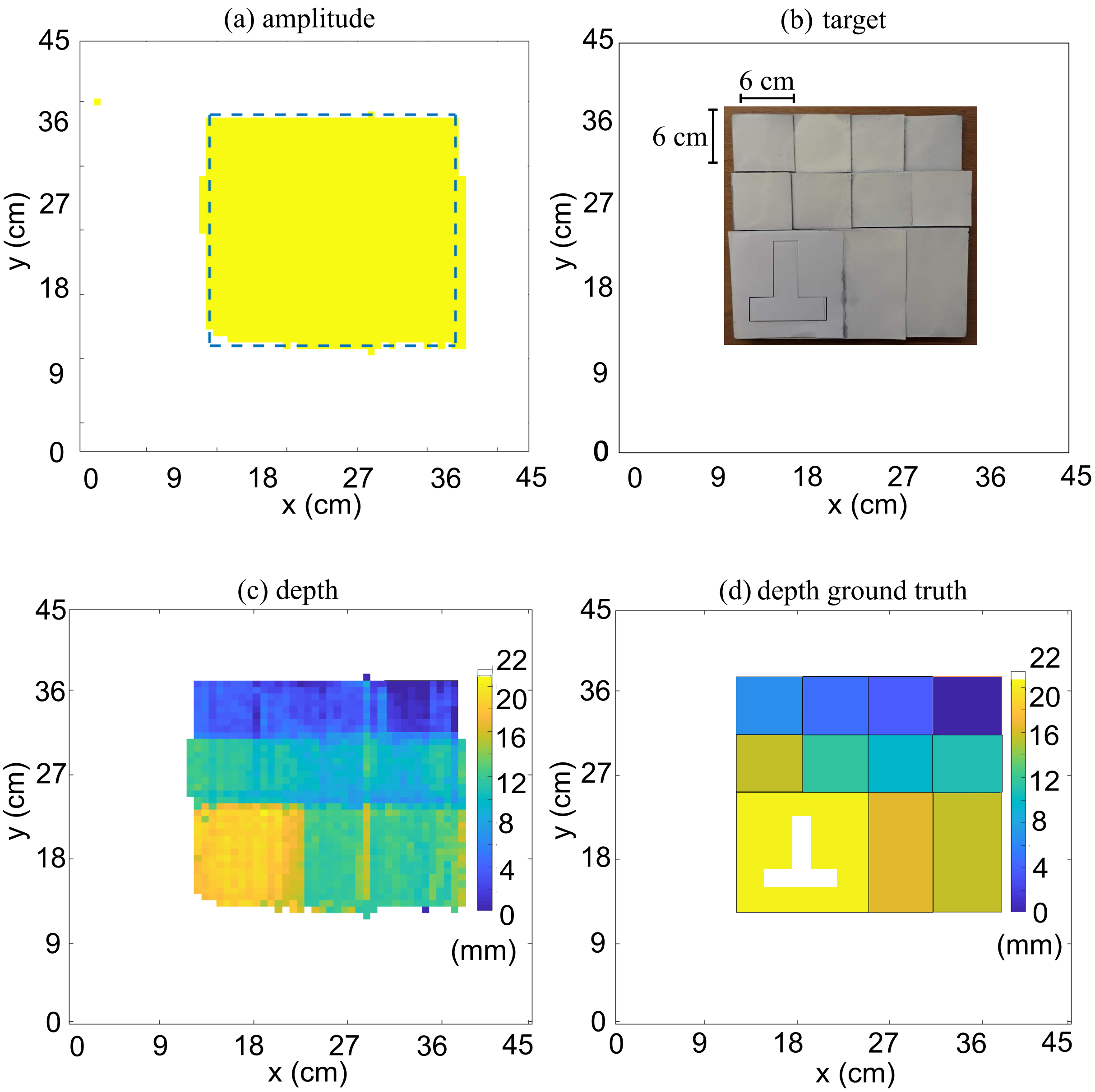

Figure 3.11: **3D imaging with lock-in single-pixel camera results.** a) Retrieval of the 3D shape of the investigated scene on the (x-y) plane. In order to facilitate the visualization, a threshold of 0.83 has been applied to suppress the lower amplitude pixels. The dotted blue line indicates the actual position of the target. b) 24x24 cm^2 square target to be retrieved. The target consists of a series of 6x6 or 12x12 cm^2 squares, 6x12 cm^2 rectangles and a "T" letter. The targets are located at varying depth within a range of 0-22 mm in order to evaluate the depth resolution of the method. In detail, the targets are placed at an increasing relative depth of 0, 5, 6, 7, 10, 11, 12, 14, 16, 17, 21, 22 mm respect to the farthest target. c) Retrieval of the depth information of the investigated scene by a colour-encoded depth image within a depth range 0-22 mm. d) Colour-encoded depth ground truth. The colourmap encodes the relative phase difference between each (x-y) pixel and the farthest measured pixel centred at coordinates (x=35 cm, y=38 cm, z=0).

By sampling over the entire oscillation period of the return signal, the suggested method based on lock-in demodulation detection provides an accurate 3D retrieval of LOS scenes with a depth resolution of 5 mm. The experimental results demonstrate that the proposed approach and the standard ToF cameras offer a comparable depth resolution.

3.5 Conclusions

Preferred over ultra-sonic and RADAR 3D imaging systems for lateral resolution and fastness, optical non-scanning ToF systems recover the 3D image of the scene by illuminating the scene and measuring the intensity and the return time of the back-scattered signal. In order to infer the 3D information, current ToF systems use either pulsed light sources or CW intensity-modulated optical beams. Since this approach combines the intensity and the ToF information of the return signal, it allows the recovery of the transverse plane and of the depth information simultaneously without requiring any additional data processing as happens for holography or stereovision systems.

The pulsed light and the CW intensity modulated light source technologies employ two distinct approaches to measure the time-of-flight information. The pulsed light technology infers the depth information d by the temporal difference Δt between the arrival-time of the return photons and a time-zero reference signal. On the other hand, intensity modulated ToF cameras infer the depth information by sampling the phase offset between the return (received) signal and a reference (emitted) signal over the oscillation period. As opposed to pulsed ToF cameras requiring time-resolving detectors, intensity modulated ToF cameras require only a CW light beam whose amplitude is modulated in time at a given modulation frequency f_m.

Despite the versatility, the high frame-rate and the compactness of the commercial intensity modulated ToF camera, standard CW modulated ToF cameras are affected by some limitations. Since the depth information is inferred sampling the return signal four times at equal intervals in a oscillation period, the depth resolution of the investigated scene is limited to 0.5-1 cm [85]. Moreover the ambiguity free-range distance limits the versatility and the performance in long-range 3D imaging scenarios according to the modulation frequency of the reference signal.

Here, we demonstrated the 3D retrieval of a LOS scene by a lock-in single-pixel camera with a depth resolution of 5 mm. Our approach uses an intensity modulated flood-illumination oscillating at a given reference frequency and a lock-in amplifier to extract the return signal acquired by raster scan. The suggested approach simultaneously provides the depth and the intensity image of the investigated scene requiring minimal additional processing. The combination of the structured illumination and the intensity of the return provides the transverse information of the scene. The depth information is then retrieved by the phase offset between the emitted and the return signal. This proposed method provides an alternative CW modulation 3D imaging system with comparable depth resolution.

However, some limitations affect the suggested approach. Due to the phase delay induced by cascading multiple RC filter, the lock-in amplifier indeed requires settling time of the order of tens time constant in order to achieve a 99% phase accuracy. With a time constant of 100 ms and a spatial resolution of 64x64 pixels, the lock-in demodulation detection requires an overall settling time of 100 minutes. The overall settling time is drastically impractical for real-time application and not compatible with the 160 fps of the compact ToF cameras on the market [86]. Providing the depth and the intensity information for each pixel, commercial ToF cameras represent a 3D imaging device more compact and faster than the proposed lock-in single-pixel camera with comparable depth accuracy. Another limitation is the compactness and the portability of the system. Indeed the combination of a single-pixel detector, a lock in amplifier and a digital mirror device may results in a bulky device not compatible with portable conventional ToF 3D cameras.
Although the acquisition time required to retrieved a full 3D reconstruction is not comparable with conventional real-time 3D cameras [87], the suggested method has ambiguity-free distance range tunable according to the reference signal. Since the reference signal is usually provided by an external source, the lock-in single-pixel camera therefore provides flexibility in choosing the optimal modulation frequency according to the application depth range. Additionally, the proposed approach provides the 3D image of direct scenes without employing any short pulses illumination or picosecond temporal resolution sensors with comparable depth resolution.

Chapter 4

Non-line-of-sight 3D imaging with a single-pixel camera

The 3D recovery of scenes hidden from the direct line of sight is a crucial task with applications in remote sensing, surveillance, defence and self-driving vehicles. Typical real-life scenarios are represented by objects hidden behind an obstacle such as a wall or a blind corner, making the retrieval of hidden scenes a field of research under constant investigation. The 3D recovery of hidden scenes by the detection of the third echo of a multiple scattered signal, has been demonstrated in optics, radar systems and acoustics, with the possibility to investigate new imaging modalities with a wide range of applications. Most of the promising approaches in optics are based on using ultra-fast, time-gated, single-photon sensitive cameras and pulsed light sources [30, 88–90] or continuous wave illumination [91–93]. In order to simplify the data acquisition and the data processing, alternative approaches for NLOS imaging are based on deep learning algorithms [31, 94] or light-cone transform [95].

Recently results also explored the physical properties of the so called Phasor-Field (PF) to look around corners and to retrieve 3D informations about the hidden scene [96–99] by investigating the physical behaviour of an amplitude modulated continuous wave at frequencies in the MHz range. These results demonstrated an analogy between conventional LOS and NLOS imaging. Indeed, when the aperture roughness is significantly larger than the optical wavelength and much smaller than the modulation wavelength, the scattering surface such as the relay wall can be considered as a mirror for the MHz modulated beam. Since the proposed approach relies on the physical property of the modulated light, the retrieval is therefore just the measurement of an optical beam with almost no calculation required. This approach allows the 3D imaging of a NLOS scene without

requiring any complex computational imaging algorithms.
The 3D recovery of NLOS scenes has also been demonstrated using radar frequencies where the recovery of the 3D shape from inverse problem solution is remarkably simpler than optical systems [100]. Inspired by radar systems [101], recent results demonstrated the 3D recovery of hidden scenes in acoustics by exploiting the specular reflection properties of walls [102]. Suggested by Dokmanic et al. [103], this system relies on the emission of a chirped sound by an emitter array and on the measurement of the returning signal by an array of microphones [104]. However, wave effects need to be considered in the modelling of the sound propagation for acoustic NLOS imaging.

As discussed in this chapter, current Light-Detection-And-Ranging (LiDAR) and NLOS techniques in optics rely on illuminating the scene under investigation with a pulsed light source and collecting the light scattered back by the objects in the scene by time-resolved detectors [105,106]. Measuring the time the light takes to reach the detector allows to locate the object and to retrieve the 3D scene following the well-known formula $s = c \times t$ where c is the speed of light and s is the distance cover by the light in a time t. In particular, this technique typically requires a pulsed laser beam pointed on a scattering surface such as a wall, producing a sphere of light propagating into the hidden scene. When the first scattering hits the hidden object, the sphere of light is scattered back towards the scattering surface. Collecting the third bounce echo scattered from the hidden target permits the tracking and the 3D retrieval of the hidden scene by applying imaging reconstruction algorithms [30,107,108].
Imaging reconstruction algorithms are typically based on the overlapping of the back-projected ellipsoids [30, 89]. Recent results demonstrated the 3D recovery of a hidden scene by using a frequency-wavenumber (or $f-k$) migration method to solve the inverse NLOS problem [109]. Inspired by the seismic imaging process of recovering the complex subsurface by detection of the waves at the surface [110, 111], the f-k migration approach exploits the similarities between the seismic problem and NLOS imaging, providing a robust method to retrieve the 3D information of complex surfaces in confocal and non confocal imaging system.
Since the multiple back-scattered signal is typically weak, visible and near-infrared methods require single-photon sensitive detectors, many acquisitions and complex reconstruction imaging algorithms to infer the information about the hidden scene [30, 112]. Recently results demonstrated only the tracking of moving targets at short and long-range distances [24,92,113]. Since the 3D recovery of hidden scenes requires more information and longer acquisition times, the real-time 3D retrieval of hidden scenes is still impractical, even if recent progresses have been made

to reduce the acquisition and reconstruction time for retroreflective scenes [95]. Moreover the spatial resolution of the retrieval is strictly dependant on the temporal resolution of the detector: single-photon detectors in the market have a temporal resolution of the order of tens picoseconds corresponding to a centimetre spatial resolution. Since the 3D retrieval of hidden scenes requires high speed and high temporal resolution single-photon sensitive cameras, the NLOS imaging still represents a challenge and is currently an interesting topic of research under constant investigation.

The purpose of this chapter is the investigation of the full 3D retrieval of four scenes hidden from the direct line-of-sight by using a time-resolved single-pixel camera, providing more flexibility in choosing the optimal detector for the imaging purpose. The single-pixel camera approach allows to reduce the acquisition times with good reconstruction fidelity by combining high sensitivity single photon detectors of sub 30 ps temporal resolution with a digital micromirror device (DMD) with up to 20 kHz refresh rate without requiring any scanning parts.

By using a high sensitivity detector (80% of QE), our results demonstrate the full 3D retrieval of hidden targets with an improved acquisition time down to sub-second, paving the way to real-time 3D imaging of non-line-of-sight scenes.

In this chapter we describe the experimental setup used to collect the return signal in a form of temporal histograms. We then apply the back-propagation imaging algorithm [30, 107] to infer the 3D information of the hidden scene by using a single-pixel camera. A further improvement in this method would be to reduce the acquisition times by applying imaging algorithms such as those based on compressive sensing. Compressive sensing allows to fully exploit the capabilities of a single-pixel camera and retrieve comparable quality images with less measurements [63, 114, 115]. Moreover, our results demonstrate the full-colour retrieval of an Red-Green-Blue (RGB) coloured scene by using a white-light source with a broad continuous spectrum of emission within the visible frequencies. Additionally we demonstrate the 3D retrieval of a hidden scene in a "time-reversal" scenario by structured illuminating the field of view (FoV) and collecting the light from a single observation point. The "time-reversal" provides more flexibility in choosing the optimal setup configuration for the imaging challenge being addressed.

4.1 Looking around corners experimental setup

Here, we describe the experimental setup used to investigate the hidden scene. The experimental setup (Fig. 4.1) is composed by a pulsed light source and a time-resolved single-pixel camera to image the signal scattered back by the hidden objects on the observation area. The single-

pixel camera is composed by a DMD (Vialux SuperSpeed V-7001 Module) combined with a time-resolving single-pixel detector, in this case either a SPAD detector or a PMT.
In this case the imaging system has an observation area of 50 x 50 cm^2 on the scattering surface and each DMD mirror reflects the light from a specific spatial portion of the observation area by using a camera lens objective (Samyang, 8 mm focal length, f/3.5). The collected light is then projected on the time-resolved single-pixel detector. Since the tilt of each mirror can be chosen according to the mask applied on the DMD, the use of a DMD allows to spatially map the light passing on the entire observation area on a single-pixel detector, recovering the spatial information lost using a single pixel detector. In this case each DMD mask has a spatial resolution of 20x20 pixels corresponding to 2.6x2.6 cm^2 pixel area on the FoV.

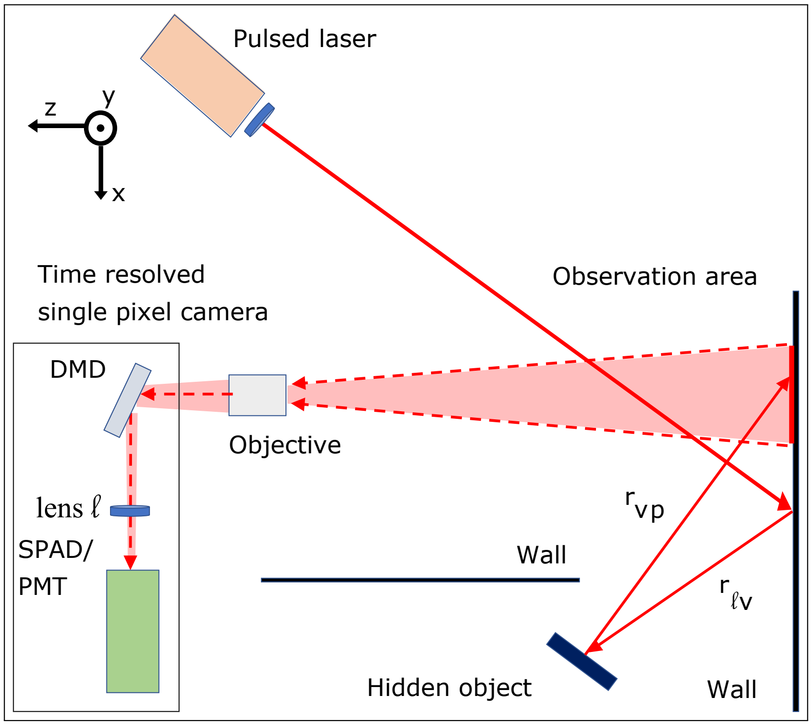

Figure 4.1: **Schematics of the experimental set-up for non line of sight imaging.** A pulsed laser source scatters on a scattering surface producing a spherical wave propagating in the surrounding area. The imaging system is composed by a time-resolved single-pixel camera. The DMD is imaging a 50x50 cm^2 area of the scattering surface and the time resolved single-pixel camera collects the signal back-scattered from the hidden targets in temporal histograms. We sequentially collecting the return photons in each of the 20x20 pixels either by raster scan patterns or by Hadamard patterns. The DMD (placed 1.16 m far from the wall) then projects only selected portions (20x20 pixels masks) of the image onto a single-pixel, single-photon detector by using a converging lens ℓ of 10 cm focal length. The detected signal is then stored in time histograms of 4096 time bins of 6.1 ps duration each.

The time-resolving single-pixel detector operates in time-correlated single photon counting (TCSPC) mode and records the arrival-time of the collected light in a temporal histogram. When a photon is collected, the detector stores the arrival-time measured as the difference in time between the arrival of the photon and a TTL trigger. The trigger is provided by the pulsed laser source. Each temporal histogram is composed by 4096 time bins of 6.1 ps duration each.

A pulsed light source is then sent 10 cm to the right of the FoV of the single-pixel camera, producing a spherical wave propagating in the surrounding area. A part of the spherical wave hits the objects hidden behind the occluder, scattering back the light on the observation area. Considering the arrival-time of the collected photons it is possible to locate the hidden object and retrieve the 3D image. We then acquire the signal scattered back on the FoV by applying consecutive patterns and collecting the corresponding signal. The number of the patterns we apply depends on the type of patterns used in the acquisition, in this case Hadamard or raster scan masks. Considering the laser spot as the origin $O(x=0, y=0, z=0)$ of our frame of reference, the following formula describes the intensity $I(x', y', z'=0)$ collected at given pixel $(x', y', z'=0)$ of the observation area:

$$I(x', y'z'=0, t) = \int_{x,y,z} \frac{I_0 \delta(tc - r_{\ell v} - r_{vp})}{r_{\ell v}^2 r_{vp}^2} dxdydz \tag{4.1}$$

where I_0 is the initial intensity of the spot on the scattering surface. Here, the delta function describes the propagation of the light as a spherical wave from the laser spot to the observed pixel. Indeed, the quantity tc is the distance covered by the light travelling at a speed $c = 300000\ km/s$ in a time t, $r_{\ell v}(x,y,z) = \sqrt{x^2+y^2+z^2}$ is the distance between a generic portion of the object (x,y,z) and the laser spot, and $r_{vp}(x,y,z,x',y',z'=0) = \sqrt{(x-x')^2+(y-y')^2+z^2}$ is the distance between the portion of the object and the observed pixel. The δ describes the 3D shape of the object through the time-of-flight signal scattered back from each point of the object. Finally, the $1/r^2$ term describes the decay of the intensity with the distance due to diffusive reflection from the wall and the object. We then integrate over the entire object.

We investigate various imaging scenarios using different combinations of laser sources and detectors and different $20x20$ pixels masks chosen accordingly to the imaging challenge being addressed. We now report the spectrum of emission of the laser sources we used. In this case we used three lasers sources at different wavelength.

4.1.1 Laser sources

The first laser source we used in the experiment is a pulsed Ti:Sapphire femtosecond oscillator (Chameleon Ultra II, Coherent) emitting 120 fs pulses of 10 nJ with a repetition rate of 80 MHz and an average power of 800 mW. Figure 4.2 (a) shows the spectrum of emission measured by a spectrometer in the visible-near infrared range. The laser source emits pulses of light at 809 nm with a 4 nm bandwidth.

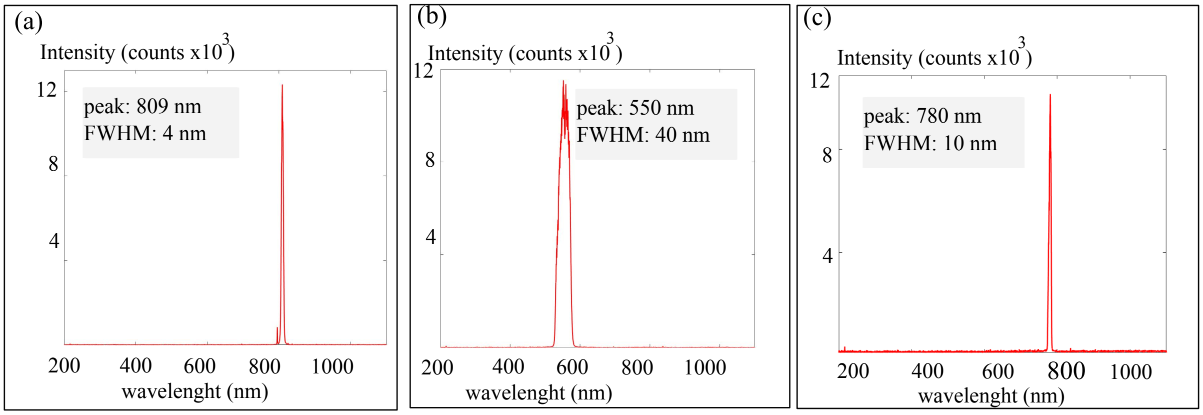

Figure 4.2: **Spectrum of emission of the three laser sources used in the experiment.** (a) The near-infrared laser source emits pulses at 809 nm wavelength with a 4 nm bandwidth. (b) The white-light laser emits light at 550 nm wavelength with a 40 nm bandwidth by using a corresponding spectral filter. (c) The Toptica FemtoFErb laser emits light at central wavelength of 780 nm with a bandwidth of 10 nm.

Another light source we used in the experiment is a supercontinuum white-light laser ((SuperK EXTREME/FIANIUM, NKT Photonics) producing $\sim$ 10 ps pulses of 1.5 nJ with a repetition rate of 67 MHz and an average power of 100 mW. Since the laser emits in the all visible spectrum, we select the desired wavelength by a band-pass spectral filter centred at 550 nm with 40 nm bandwidth after the laser source. Figure 4.2 (b) shows the spectrum of emission. The last laser source we used in the experiment is a near-infrared 100 fs pulsed laser (Toptica FemtoFErb) of 1.4 nJ pulses with a 100 MHz of repetition rate and 140 mW average power. As shown in Fig. 4.2 (c) the laser source emits at a central wavelength of 780 nm with 10 nm bandwidth.

4.1.2 Characterization of detectors

We now characterize the impulse response function (IRF) of the single-pixel detectors in order to determine the ability of the detectors to distinguish peaks adjacent in time. In this case we used a single-pixel PMT and a SPAD detector shown in Fig. 4.3, operating in TCSPC mode. The detec-

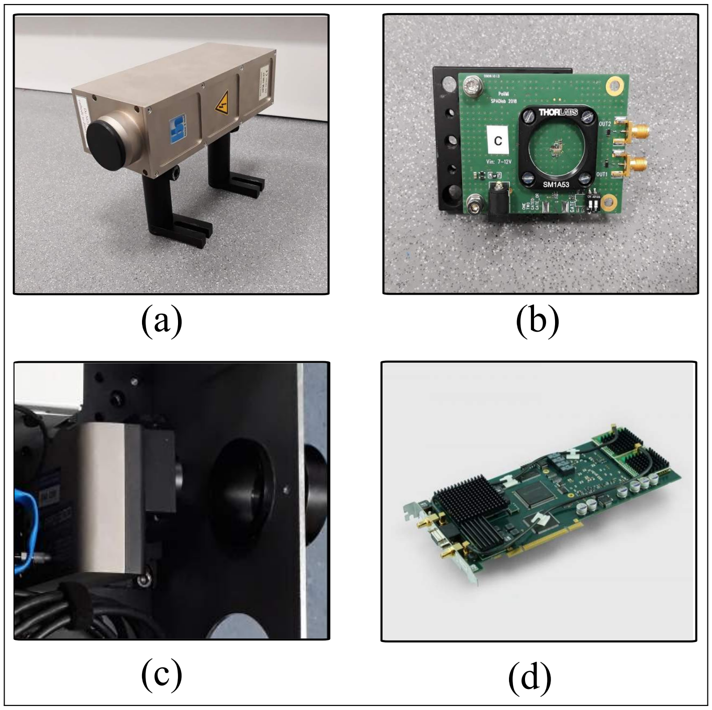

Figure 4.3: **Pictures of the time-resolving single-pixel detector used to collect the signal scattered back by the hidden object.** (a) HPM-100-07 hybrid Photo-multiplier-tube (PMT, Becker & Hickl) with a multi-alkali cathode. The complete module is operated by the Becker & Hickl DCC-100 detector controller card that provides for overload shutdown, control of the avalanche-diode reverse voltage, and power supply. (b) Silicon single-photon avalanche-diode (SPAD) detector manufactured in a 0.16 μm BCD technology with monolithically integrated sensing circuit [2]. (c) Picosecond photon counting module (PPD-900, Horiba) PMT detector used to investigate the non-line-of-sight scene in a time-reversal scenario. (d) The single-photon data is recorded in time-correlated-single-photon-counting (TCSPC) mode integrating the detector with a TCSPC card (Figure (d)). The TCSPC card has two input signals, one for the trigger from the laser and one for the photon counting from the detector.

tors are connected to the TCSPC card via an SMA cable. In particular, the PMT (Fig. 4.3 (a)) is a hybrid photo detector HPM-100-07 (Becker&Hickl), with 1% of QE at 808 nm and an active-area diameter of 3 mm. The HPM-100 module integrates an Hamamatsu R10467 hybrid detector tube with a pre-amplifier in one compact housing with high timing resolution. The SPAD detector (Fig. 4.3 (b)) is manufactured in a 0.16 μm BCD technology with monolithically integrated sensing circuit [2], square active area of 57x57 μm^2 and high PDE (70 % peak PDE at 550 nm). In the time-reversal experiment the compact single-pixel PMT is a picosecond photon detection PPD-900 (Horiba) (Fig. 4.3 (c)). The detector has an active area of 77 x 107 mm^2 and a QE of 8% at 780 nm wavelength. Figure 4.4 shows the efficiency of the single-pixel detectors.

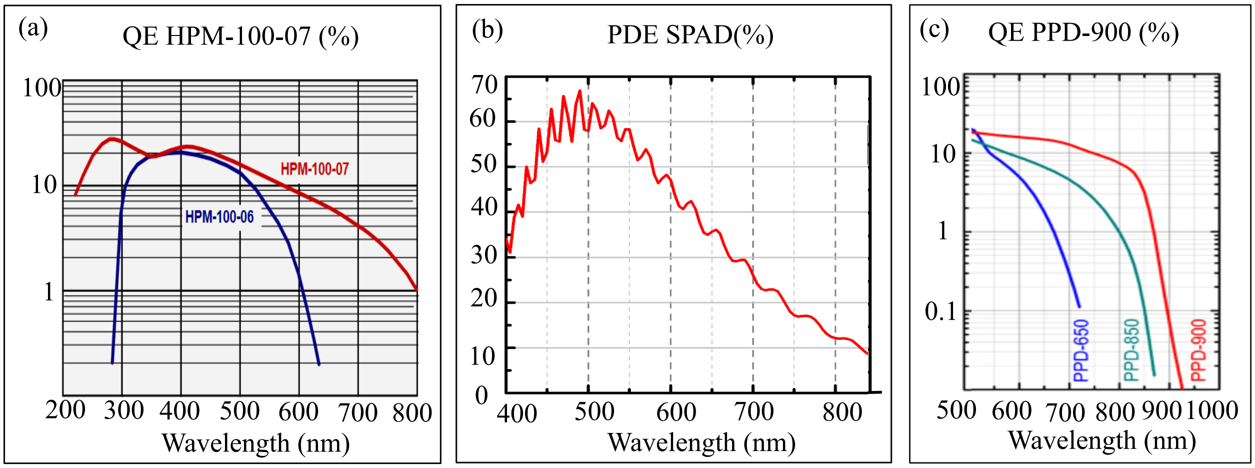

Figure 4.4: **Efficiency of the single-pixel detectors.** (a) The PMT has a hybrid photo detector HPM-100-07 (Becker&Hickl), with 1% of QE at 808 nm wavelength [3]. (b) The SPAD has a PDE of 70 % at 550 nm wavelength [2,4] (c) The Picosecond photon detection PPD-900 PMT has a QE of 8 % at 780 nm wavelength [5].

We then measure the IRF of each detector using the experimental setup in Fig. 4.5. In order to measure the temporal resolution we use a femtosecond oscillator (Chameleon ultra II) sending pulses of 120 fs duration at 808 nm wavelength with a repetition rate of 80 MHz. The laser passes through the optical system of Fig. 4.5 composed by two mirrors and a converging lens (focal length= 100 mm) in order to focus the signal onto the detector. A neutral density filter of optical density OD = 8 reduces the number of photons to 2×10^5 photons. To synchronize the acquisition between the laser and the TCSPC card, an electronic trigger signal is sent from the laser to the detector. Once the laser reaches the detector, a peak pulse corresponding to the arrival of the 120 fs pulses is detected and the collected light is stored in a temporal histogram of 4096 time bins of 410 fs each.

Since the pulse temporal duration is much shorter than the temporal width of the time bins, the signal to be measured can be considered as a Dirac function. However, the measured signal has a definite width due to the finite temporal resolution of the detectors. Figures 4.6-4.8 report the measured IRF for the PMTs and for the SPAD detector in the linear and in the logarithmic scale for the y axis. The temporal resolution of the detectors has been measured by the full-width-half-maximum (FWHM) of the temporal histogram. As reported in Figs. 4.6-4.7, the PMT has a temporal resolution of 27 ps corresponding to 66 time bins, whereas the SPAD has a temporal resolution of 32 ps corresponding to 78 time bins. The PDD PMT detector has a temporal resolution of 180 ps corresponding to 7 time bins (Fig. 4.8).

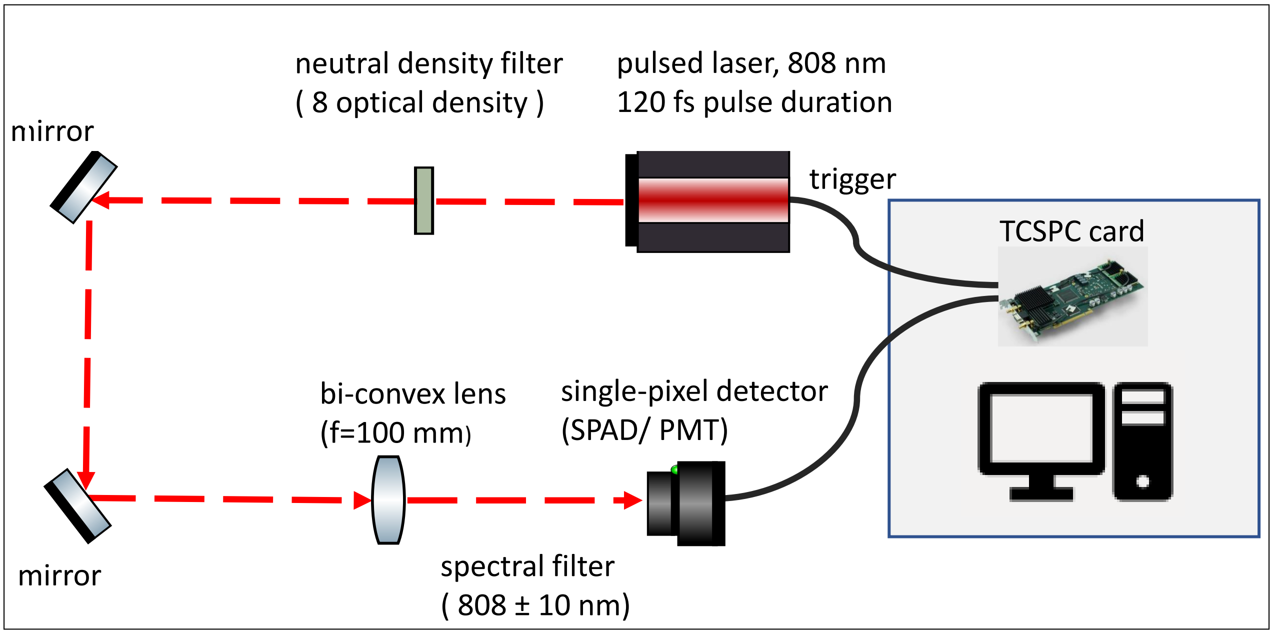

Figure 4.5: **Experimental setup used to measure the temporal resolution of the detectors.** A ultrashort pulses of 120 fs temporal duration is sent onto the detector to measure a single sharp temporal peak on the single-pixel detector. The time-of-arrival of the photons are stored in a temporal histogram form.

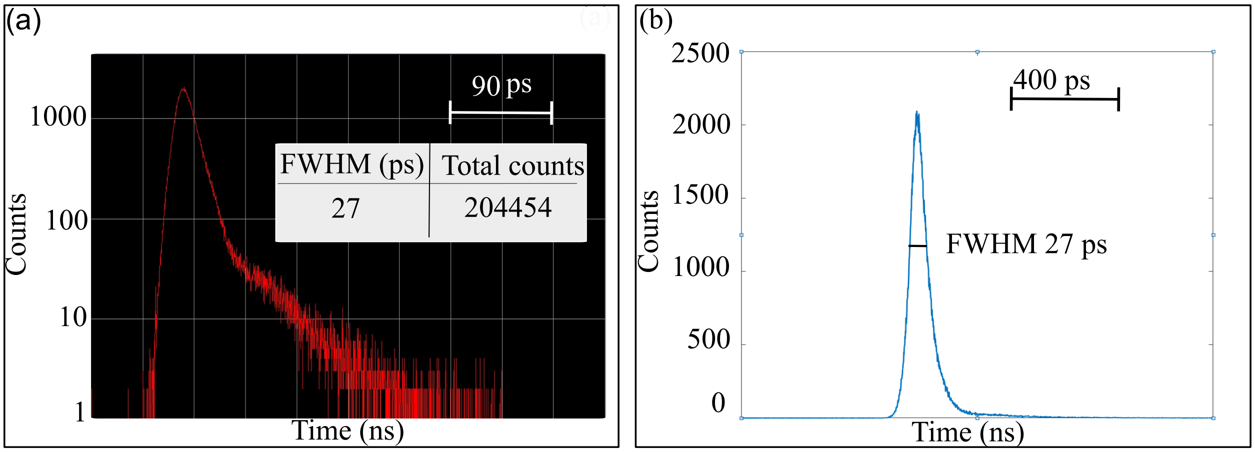

Figure 4.6: **IRF of the single-pixel PMT detector in logarithmic (Fig. (a)) and linear (Fig. (b)) scale for the y axis.** The IRF has been measured storing the peak signal in a temporal histogram of 4096 time bins of each 410 fs time duration. The PMT has a temporal resolution of 27 ps FWHM.

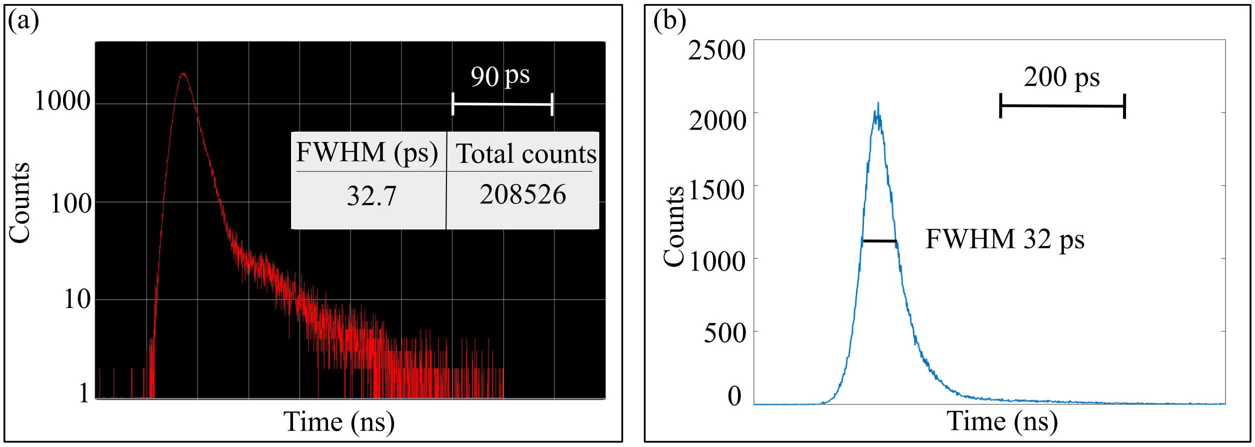

Figure 4.7: **IRF of the single-pixel SPAD detector in logarithmic (Fig. (a)) and linear (Fig. (b)) scale for the y axis.** The IRF has been measured storing the peak signal in a temporal histogram of 4096 time bins of each 410 fs time duration. The SPAD detector has a temporal resolution of 32 ps FWHM.

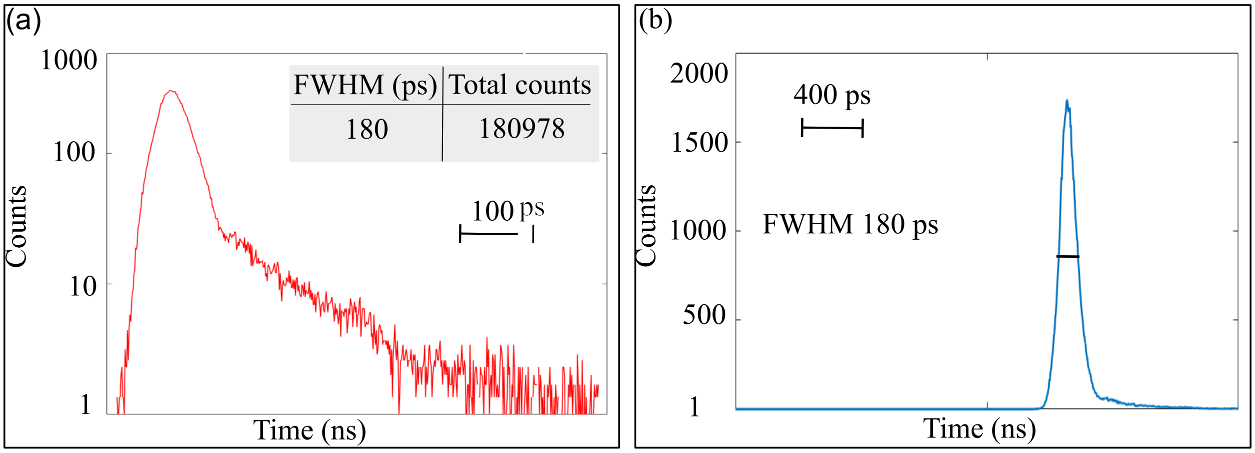

Figure 4.8: **IRF of the PMT PPD-900 detector in logarithmic (Fig. (a)) and linear (Fig. (b)) scale for the y axis.** The IRF has been measured storing the peak signal in a temporal histogram of 512 time bins of each 25 ps time duration. The PMT detector has a temporal resolution of 180 ps FWHM.

4.2 Measurement of the back-scattered signal

We then acquire the return signal using the experimental setup in Fig. 4.1 sequentially applying 20x20 pixels masks on the DMD and projecting the light onto the single-pixel detector. We collect the signal in time across a certain portion of the FoV accordingly to the applied mask. The return signal is stored in a temporal histogram of 4096 time bins of 6 ps each. We acquire the back-scattered signal sequentially detecting the light from each of the 20x20 pixels either by raster scan patterns or by Hadamard patterns. For the raster scan acquisition we apply a total of 400 masks. For the Hadamard patterns acquisition we combine one binary mask and its negative for each Hadamard pattern, for a total of 800 masks applied. After acquiring the complete set of masks, we obtain an histogram for each applied pattern.

We distinguish the target signal from all the other background signals from the environment. Since the background and the target signals are well separated in time, we isolate the signal of interest selecting its temporal range from the background signal, as reported in the temporal histogram of Fig. 4.9 for a two objects scenario. Here, the higher peak at the beginning of the histogram corresponds to the first scattering of the laser spot on the relay wall.

Figure 4.10 shows the temporal histogram of the back-scattered signal for two distinct pixels of the FoV for a two objects scenario collected with the PMT detector. In this case we used a 120 fs pulsed laser at 809 nm wavelength with 80 MHz repetition rate and an average power of 800 mW. As reported in Fig. 4.10, the temporal width of the signals is larger than the temporal resolution of the single-pixel detector. Indeed, the width is not limited by the resolution of the detector, but it's affected by the scattering surface of the hidden targets.

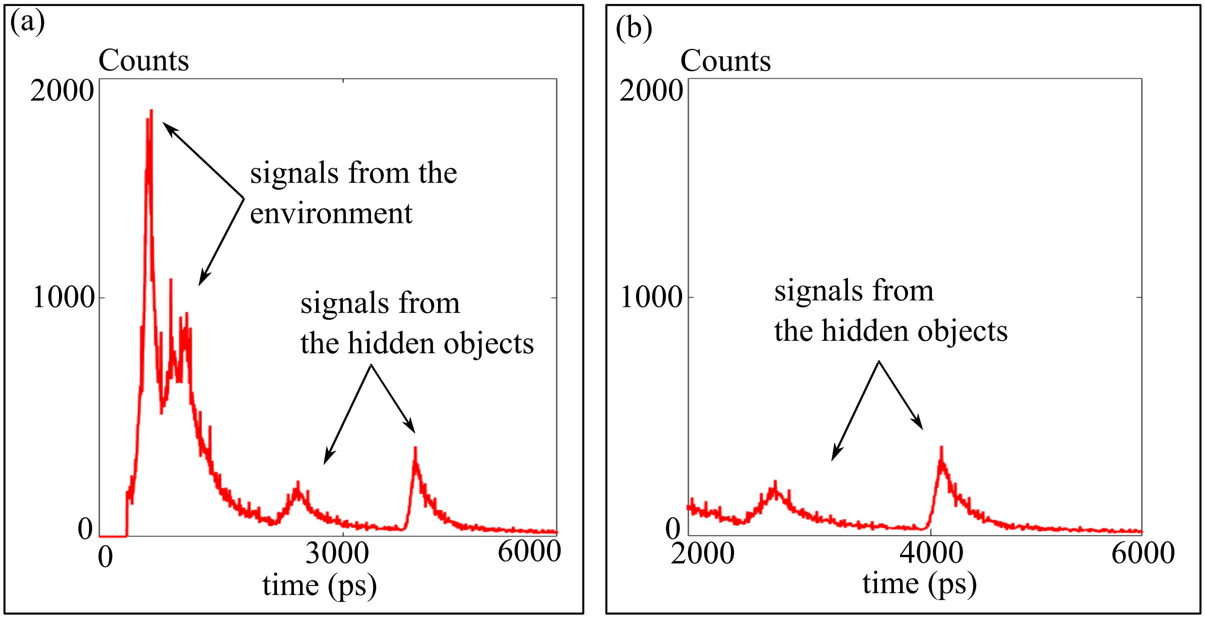

Figure 4.9: **Acquired temporal histograms of the back-scattering.** (a) The collected histogram contains either the background signal coming from the environment and the return signal from the hidden targets. (b) The return signal of the hidden targets are well separated in time from the background signal and we can isolate the return signal from the hidden targets without any background subtraction.

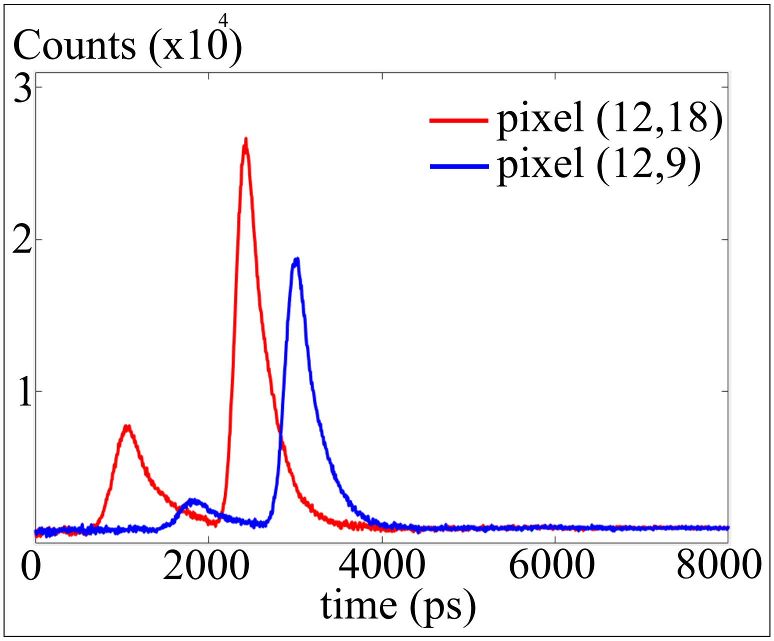

Figure 4.10: **Acquired temporal histograms of the back-scattered signal for two pixels of the field of view.** The two peaks correspond to the two objects in the hidden scene.

As reported in Fig. 4.11, we rearrange the acquired data in a 3D matrix to obtain the back-scattered signal passing in time across the FoV during the acquisition time. Fig. 4.11 shows the 3D (x;y;t) data matrix of the back-scattered signal. Looking at the FoV data at a particular time t, we obtain the back-scattered signal across the FoV at that time. (4.3).

Figure 4.12 shows the temporal evolution of the back-scattered signal across the FoV for a two objects scenario in a 6300 ps time interval. The time frames are separated by 700 ps each and the return is acquired by the PMT detector in Section 4.3 (a). The objects to be retrieved are two round targets of 2.54 cm and 7.62 cm of diameter placed at different positions. In this case the total acquisition time is 66 minutes corresponding to an acquisition time of 10 seconds per mask (i.e. pixel). Starting from Fig. 4.12 (a), the return signal of the smaller object is approaching the 50x50 cm^2 FoV and then it propagates from right to the left (Figs. (a-c)). After that, the return signal of the second object starts to propagate across the FoV (Figs. (c-i)). From Fig. (e) the return signal of the smaller target becomes less visible since the signal is affected by low signal-to-noise ratio. We then proceed with the retrieval of the 3D shape of the objects from the temporal evolution of the return signal applying the back-projection imaging algorithm.

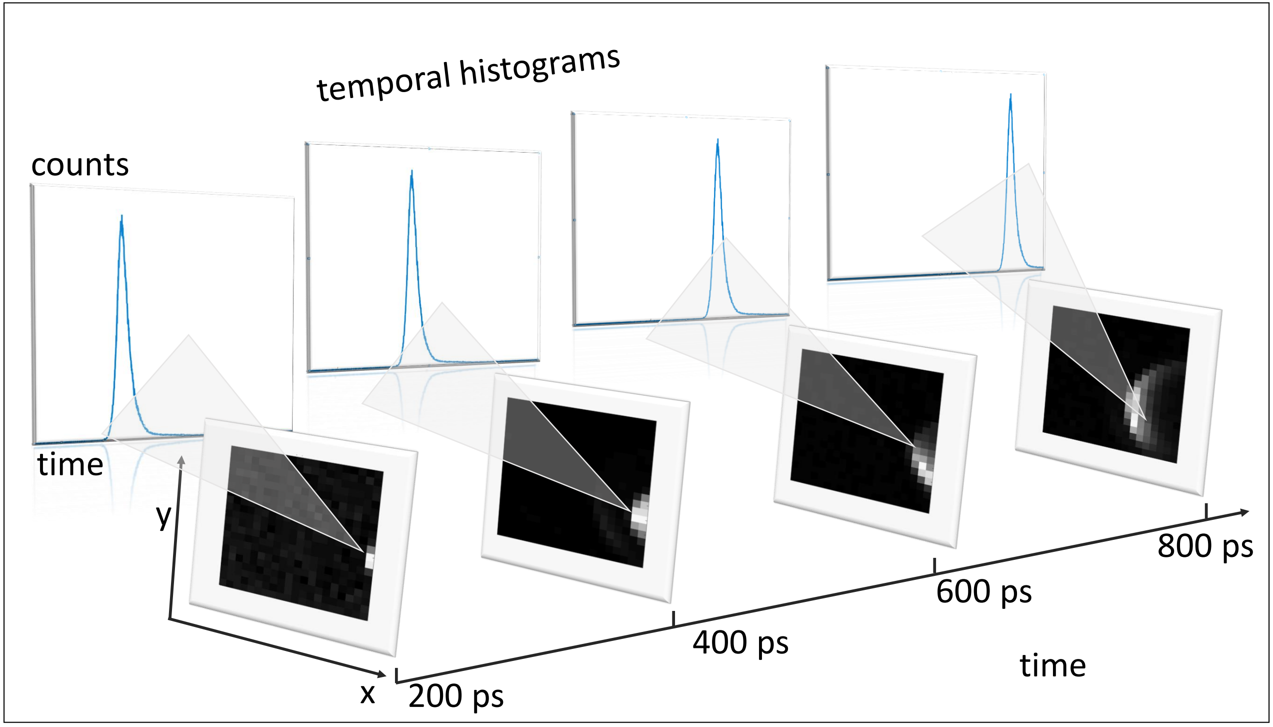

Figure 4.11: **3D data matrix of the back-scattering.** After each acquisition, we obtain the temporal histogram of the back-scattered signal across a specific portion of the field of view corresponding to the applied pattern. The data are rearranged in a 3D matrix where the x-y dimensions represent the x-y pixel of the field of view and the z dimension represents the time. Sequentially Looking at the x-y plane sections of the 3D matrix, we obtain the temporal evolution of the back-scattered signal across the filed of view (here for 200, 400, 600 and 800 ps). Looking at a particular (x; y) pixel along the z dimension, we obtain the temporal histogram of the back-scattered signal collected at that pixel.

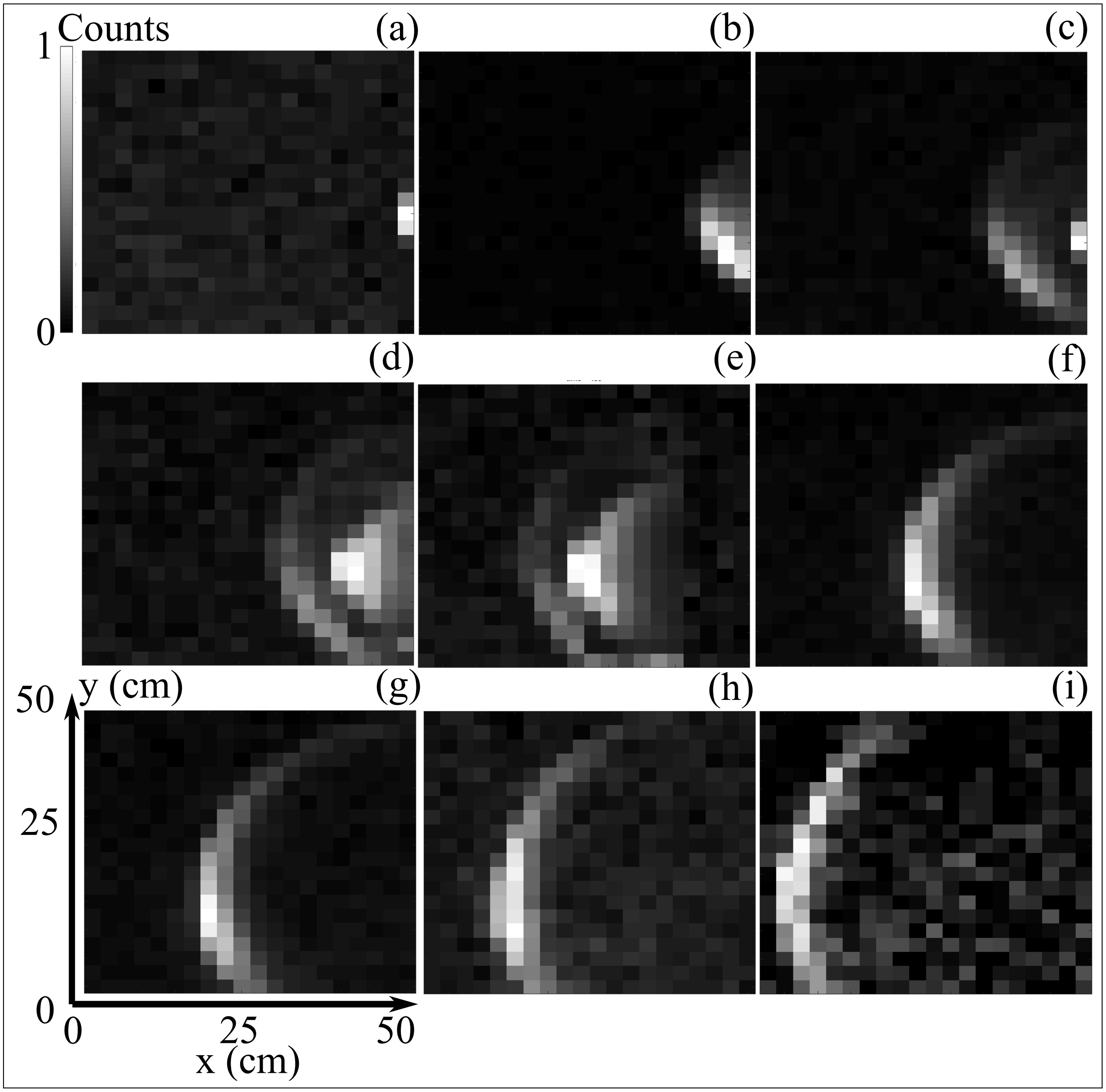

Figure 4.12: **Temporal evolution of the return signal across the 20x20 pixels FoV for a two objects scenario.** The colour bar indicates the number of photons detected at each pixel. Each time frame is separated by 700 ps for an overall time interval of 6300 ps. To facilitate the visualization the number of counts is normalized to the maximum value.

4.2.1 Hadamard patterns acquisition

We acquire the return signal across the FoV by applying either raster scan masks or Hadamard masks onto the DMD. The former set of masks is composed by 400 different masks which have one pixel "on" and 399 pixels "off". The latter set of mask is composed by 800 Hadamard matrix masks of 20x20 pixels.

The Hadamard matrix is a square, invertible matrix with mutually orthogonal rows (and columns) whose elements are either +1 or -1 [116]. An Hadamard matrix H of order n verifies the following condition:

$$HH^{'} = H^{'}H = HH^{T} = nI \tag{4.2}$$

where $H^{'}$ and H^{T} are the inverse and the transpose matrix of H and I is the identity matrix. Here, n is the order of the Hadamard matrix and it could be 1, 2 or a positive integer of 4. An example of Hadamard matrix with order $n = 4$ is given by:

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

In order to the achieve maximum efficiency, patterns with no overlapping information about the scene have to be applied during the back-scattering detection. Since Hadamard matrices are composed by mutually orthogonal rows, "Hadamard matrix derived" patterns represent a convenient choice to maximize the efficiency of the back-scattering acquisition. We then consider the 400x400 pixels Hadamard matrix and we built a set of 400 "Hadamard derived" patterns reshaping each of the 400 rows of the Hadamard matrix into a 20x20 pixels binary matrix. The complementary 400 Hadamard derived matrices are then created by multiplying each matrix entries by -1. Since the derived patterns are orthogonal to each other, there is no overlap between each measurement and so we achieve a higher reconstruction efficiency. In this way, we guarantee the orthogonality and the efficiency of the acquisition patterns. For simplicity, the "Hadamard derived" patterns are referred to as "Hadamard" pattern in the present thesis. Appendix B reports the code used for generating the Hadamard patterns generations.

We acquire the return signal across the FoV applying Hadamard patterns onto the DMD by programmable binary masks. Since half of the 20x20 pixels are always collecting signal, the use of Hadamard basis guarantees the 50% transmission of the image intensities. Therefore, the average number of the detected photons in the Hadamard mask measurements is N/2 times greater than in a raster scan mask measurements during each acquisition. The sequence of the intensity measurements and of the corresponding Hadamard patterns allow the reconstruction of the back-scattered signal.

In order to obtain the 3D data matrix in Fig. 4.12, we sum over the sampling Hadamard patterns weighted according to the corresponding measured intensities, that is, the return photons histograms. Since we combine one binary mask and its negative, the number of counts Counts(x;y;t) for a (x;y) pixel at a time t are obtained as follows:

$$Counts(x,y,t) = \sum_{i=1}^{\frac{N_{masks}}{2}} Mask_i(x,y) * (Counts_{i,+}(t) - Counts_{i,-}(t)) \tag{4.3}$$

where the index i indicates the mask from 1 to 400. Here, $Counts_{i,+}(t)$ and $Counts_{i,-}(t)$ are the counts of the histogram of the i-th positive mask and its negative at a time t. The quantity $Mask_i(x,y)$ represents the (x;y) pixel of the i-th Hadamard mask whose value is $+1$ or -1 according to whether the pixel is collecting or not the light. We then obtain the back-scattering across the FoV by applying Eq. 4.3.

4.3 Back-projection imaging algorithm for NLOS 3D retrieval

The collection of the third bounce echo scattered back by the hidden object allows to retrieve the 3D information of the hidden scene applying reconstruction imaging algorithms. These algorithms typically rely on solving the inverse light transport problem [117] in order to infer the information about a scene b from the available set of data s described as:

$$b = F^{-1}(s) \tag{4.4}$$

where F is the light transport function which models the direct propagation of light. As opposed to the well-known model for the direct light transport, the inverse light transport is still a topic of interest aimed to exploit the information embedded in multiple scattering of light to infer scene informations [118, 119]. Recent advances in inverse light transport and NLOS scenes demonstrated the 3D retrieval of hidden scenes by applying back-projection imaging algorithm with improvements in remote sensing, imaging systems and autonomous vehicles [30, 88, 90, 120].
The resolution of the 3D retrieval can be further improved applying iterative back-projection algorithms [107] or ellipsoid mode decomposition for multiple objects scenes [112].

The back-projection imaging algorithm is based on modelling the inverse light propagation to infer the 3D information from the ToF of the return photons. The hidden space is then divided in patches or voxels space V (x; y; z) representing a (x, y, z) portion of the object. We then calculate the likelihood of each voxel to contribute to the detected signal by the ToF information of the collected return signal, as discussed in the forward and back-projection algorithms.

4.3.1 Forward projection algorithm

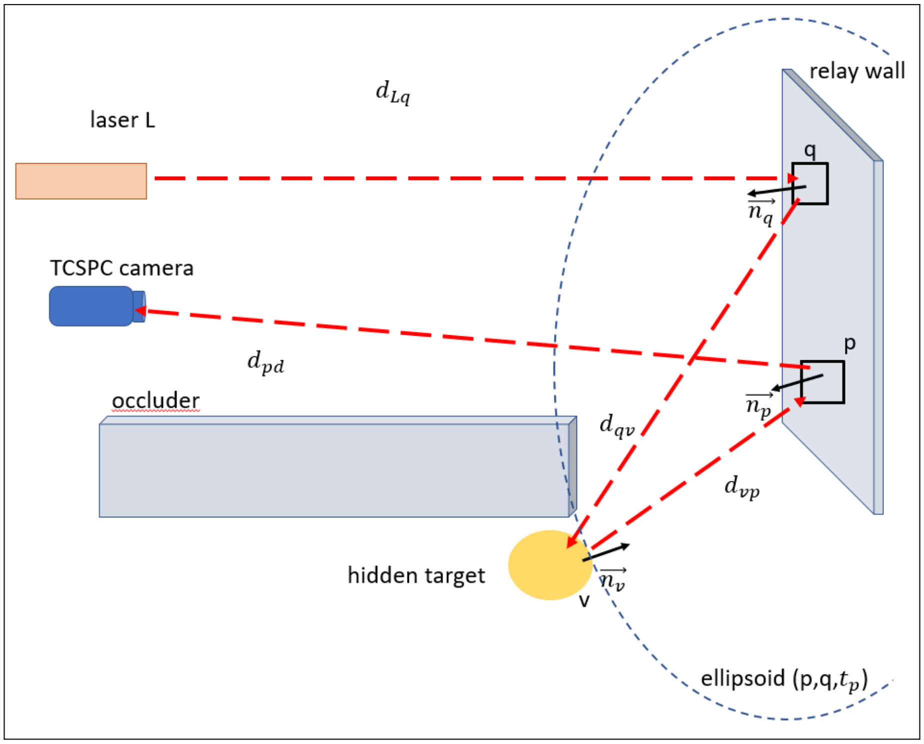

Figure 4.13: **Example of a typical NLOS scenario.** The target is out of the direct line of sight of either the camera and the laser.

We then discuss the forward and the back-projection imaging algorithms to retrieve the 3D information of the hidden scene [107, 119, 121].

We consider the scene shown in Fig. 4.13 where a pulsed laser source L is sending pulses of light on a unit area patch q of the scattering surface, producing a spherical wave propagating in the surrounding area. The hidden object scatters back a portion of the spherical wave and we collect the temporal return signal across the FoV by a camera operating in TCSPC mode. The quantity q and v represent a generic pixel of the FoV and a portion of the hidden target whose normals are n_p and n_v respectively. The radiance $F(v)$ of the laser beam in the target patch v is described by the formula

$$F(v) = F_L(q)b(v)\chi(q,v) \tag{4.5}$$

where $F_L(q)$ is the radiance at the patch q on the relay wall and $b(v)$ quantifies the diffuse reflection of the target patch v. Here, the quantity

$$\chi(q,v) = \frac{cos\angle(v-q,n_q)cos\angle(q-v,n_v)}{|v-q|^2} \tag{4.6}$$

reflects the orientation of the patch v. Taking into account the volume V of the target, the total radiance $F(p)$ at a patch p on the relay wall is described as:

$$F(p) = \int_V F_L(q) * \Big(b(v)\chi(q,v)\Big) * \Big(b(p)\chi(v,p)\Big)dv \tag{4.7}$$

Defining the quantity

$$g(v) = \frac{cos\angle(v-q,n_q)cos\angle(q-v,n_p)}{|v-q|^2|v-p|^2} \tag{4.8}$$

we can rewrite the expression in Eq. (4.7) as follows:

$$F(p) = F_L(q)b(p)\int_V g(v)(b(v)\chi(v)dv \tag{4.9}$$

We then consider the ToF information of the collected signal as the time t the light travelling at a c speed takes to cover a distance s where $s = ct$. For a given signal collected at a pixel p at a time t_p, all the possible location (x,y,z) of the target patches that contribute to the detected signal, lie on the surface of an ellipsoid whose foci are the laser spot q and the collecting pixel p. (Fig. 4.13). Indeed, considering the time the laser hits the scattering surface as the zero time reference, the possible contributions $v(x,y,z)$ are then described by the equation

$$t_p = \frac{d_{vq}(x,y,z) + d_{pv}(x,y,z) + d_{Dp}}{c} \tag{4.10}$$

where $d_{vq}(x,y,z)$ is the distance target voxel-laser, $d_{pv}(x,y,z)$ is the distance pixel-voxel and d_{Dp} is the distance between the pixel and the detector. The dotted line in Fig. 4.13 represents the ellipsoid of the possible contributions for a given signal at a time t_p collected at a pixel p for a laser position q. We then obtain the 3D retrieval of the hidden scene considering the intersection of all the individual ellipsoids $E(p_m,t_p)$ of each pixel of the FoV at any time t_p.

We then modify the Eq. (4.9) including the temporal information by a δ function:

$$F(t_p,p) = \int_V F_L(q)b(p)\delta\Big(t_pc - d_{vq}(x,y,z) - d_{pv}(x,y,z) - d_{Dp}\Big) \tag{4.11}$$

Considering a constant laser position, the previous formula can be described as a direct light transport model

$$s(p,t_p) = F(b) \tag{4.12}$$

where $s(p,t_p)$ is the data acquired by the detector, b is the reflectivity of the target patch and F is the light transport tensor describing the light propagation. In this case the dataset is made by as many temporal histograms as the number of masks. Each histogram has 4096 time bins.

In order to rewrite the Eq. (4.11) as the Eq. (4.12) we discretize the space under investigation in K_{voxels} voxels $\vec{V} = [v_1, v_2, ..., v_K]$ and the pixels position in $\vec{P} = [p_1, p_2, ..p_{Npixels}]$. We then calculate the signal $s(p_m,t_p)$ collected at the pixel p_m at a time t_p considering the contribution of each voxel of the target and applying the forward projection function [107]:

$$s(p_m,t_p) = \sum_{i=1}^{K_{voxels}} \alpha_{im}\delta\Big(t_pc - d_{v_iq} - d_{p_mv_i} - d_{Dp_m}\Big) \tag{4.13}$$

Here, the index i indicates the voxel v_i and goes from 1 to the total number of voxels and the index m indicates the pixel. The quantity α_{im} is defined as follows:

$$\alpha_{im} = \alpha_d(v_i,p_m)\alpha_{ls}(v_i,p_m) \tag{4.14}$$

The first term in Eq. (4.14)

$$\alpha_d(v_i,p_m) = \frac{1}{d^2_{qv_i}d^2_{v_ip_m}d^2_{p_mD}} \tag{4.15}$$

considers the intensity decay of a spherical wave with the distance due to scattering process from the wall and object, while the term

$$\alpha_{ls}(v_i,p_m) = b(v_i)cos\angle(q - v_i, n_{v_i})cos\angle(p_m - v_i, n_{v_i}) \tag{4.16}$$

encodes the Lambertian reflectance of the target where $b(v_i)$ is the reflectivity of the v_i voxel. Thus, Eq. (4.13) describes the light transport model $s = F(b)$ to compute the dataset s from a known scene $b(v_k)$ using the light propagation model.

4.3.2 Back-projection algorithm

We then proceed with the inverse light transport model to recover the 3D information of an unknown hidden scene b from the corresponding dataset s by solving the inverse rendering problem [30, 107, 121, 122].

Considering Eq. (4.10), the aim of the back-projection algorithm is to project each data $s(p_m,t_p)$ on the corresponding ellipsoid $E(P_m,t_p)$. The confidence map of the 3D retrieval is then obtained by considering the intersection of the set of ellipsoids of each pixel at each time t_p of the temporal histogram. Figure 4.15 shows an example of the intersection of all the ellipsoids on the (z-x) and (x-y) plane for a 20x20 pixels FoV.

In order to retrieve the 3D information of the hidden scene we discretize the space under investigation in 10^6 voxel v_i and calculate the likelihood of each voxel to contribute to the collected signal of dataset $s(\vec{p},\vec{t})$. We quantify the likelihood of a voxel v_i in terms of its reflectivity $b(v_i)$. The reflectivity $b(v_i)$ quantifies the likelihood of having the hidden object in the v_i patch.

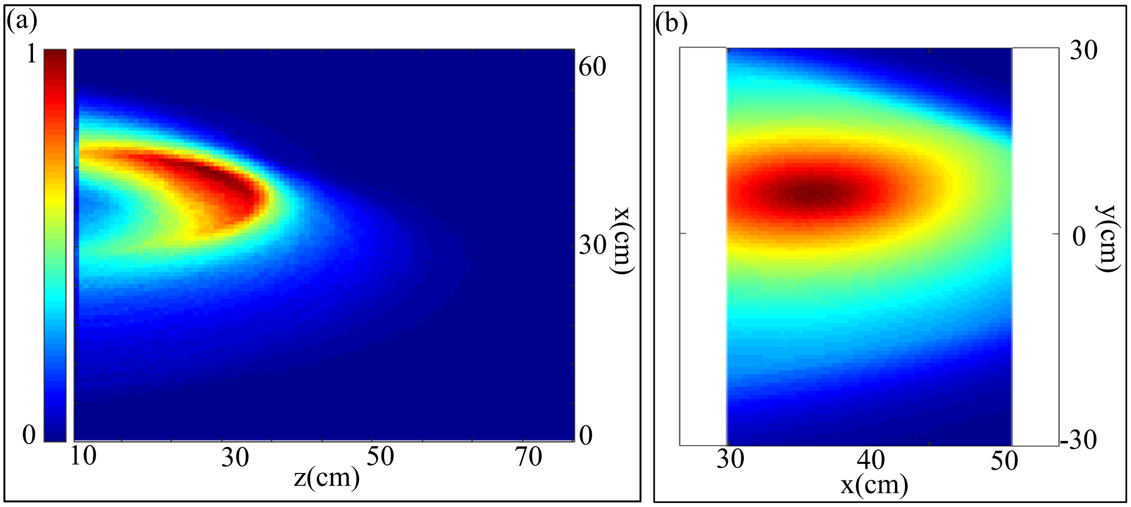

Figure 4.14: **Example of the map of confidence of a hidden scene on the (z-x) and (x-y) plane.** The intersection of ellipsoids are retrieved by applying the back-projection algorithm. In this case the scene under investigation was a hidden round object of 2.54 cm of diameter. The ellipsoid are retrieved by considering the 4096 time bins temporal histograms of the return signal acquired in a 20x20 pixels FoV.

In order to compute the likelihood of the voxel v_i, we calculate the time of arrival $t(v_i, p_m)$ of the contribution associated to that voxel and to a fixed pixel p_m of the FoV. We assign the signal $s(p_m,t(v_i, p_m))$ collected at a pixel p_m at the time $t(v_i, p_m)$ to the voxel likelihood and repeat the process for all the pixels in the FoV. Formally, the reflectivity $b(v_i)$ of the voxel v_i is described as follows:

$$b(v_i) = \sum_{m=1}^{N_{TotPixels}} \beta_{im} s(p_m,t) \delta(tc - d_{v_i q} - d_{p_m v_i} - d_{Dp_m}) \tag{4.17}$$

where $s(p_m,t)$ is the signal collected at the pixel p_m at a time t, $d_{v_i q}$ is the distance between the laser spot q and the voxel v_i, $d_{p_m v_i}$ is the distance between the pixel p_m and the voxel v_i and d_{Dp_m} is the distance between the detector and the pixel p_m. Here, the quantity β_{im} is described as

$$\beta_{im} = \beta_d(v_i, p_m, q)\beta_{ls}(v_i, q) \tag{4.18}$$

where the term $\beta_d(v_i, p_m, q)$ includes the distances factor

$$\beta_d(v_i, p_m, q) = \frac{1}{\alpha_d(v_i, p_m, q)} \tag{4.19}$$

and the term $\beta_{ls}(v_i, q)$ considers the Lambertian reflectance

$$\beta_{ls}(v_i, q) = \frac{1}{cos\angle(q - v_i, n_{v_k})} \tag{4.20}$$

The back-projection Eq. (4.17) is then describing the inverse transport model to infer the 3D information of the hidden surface $\vec{b}$ from a given dataset $s(\vec{p},\vec{t})$.

We then apply the back-projection algorithm to the experimental dataset s. In this case the dataset s is made by the 400 temporal histograms of the return signal collected at each of the 20x20 pixels of the FoV.

4.4 3D retrieval of the hidden scenes

Here, we report the 3D retrieval of the hidden scenes applying the back-projection algorithm to the experimental data with a time-resolving single-pixel camera (Fig. 4.1). The single-pixel camera technique permits to choose an optimised single-photon detector of 27 ps temporal resolution and high PDE to reduce the acquisition times down to sub-second with good reconstruction quality. Combining the single-pixel detector with the DMD allows to remove the need for any scanning components such as galvo-mirrors [30, 95].

In this case we investigate four different scenarios under different experimental conditions by acquiring the return signal across the FoV either by Hadamard or raster scan masks. Figure 4.1 shows the experimental setup used to investigate the hidden scene in each investigated scenario. Using a white-light laser emitting light in the all visible spectrum, we further extend our method to retrieve the full RGB colour retrieval of a non-retroreflective objects scene by applying corresponding spectral filters.

After obtaining the return signal by 400 temporal histograms of 4096 time bins, we divide the space under investigation in 10^6 voxels and we calculate the likelihood of each voxel in terms of reflectivity applying the back-projection imaging algorithm. We then improve the 3D information of the scene applying a thresholding and a Laplacian filter along the $\hat{z}$ direction of the voxel grid as in [30].

The confined prospective and the limited number of angles of projection of the imaging system can create artefacts such as the blurring in the 3D retrieval. Since those artefacts contribute to a low probability distribution in the voxel reflectivity [123], we can filter them out by the proposed thresholding approach. However, more complex filtering and error back-projection algorithms [107,112] might be needed to retrieve more complex scenes.

4.4.1 Two retroreflective objects scene

The first scene under investigation is composed by two retroreflective round targets of 2.54 cm and 7.62 cm of diameter placed at a varying depth and tilted (Fig. 4.15(b)). In this case we used the Ti:Sapphire laser source described in Section 4.1.1, emitting 120 fs pulses of 10 nJ at a repetition rate of 80 MHz and 800 mW average power. The return signal has been acquired raster scanning the FoV in 20x20 pixels. In this case the detector has been chosen as the photo multiplier tube characterized in Section 4.1.2 with temporal resolution of 27 ps and a 4% single-photon sensitivity at 809 nm. The acquisition time for each histogram (i.e. pixel) is 10 seconds with an overall acquisition time of 66 minutes.

Figure 4.15 (a) shows the third bounce echo signal of the two hidden objects across the 50x50 cm^2 FoV at a time frame of 7.1 ns. We then retrieve the 3D information of the scene by applying the back-projection algorithm to the return signal. The origin $O(x=0, y=0, z=0)$ of the voxels' space has been chosen as the laser spot on the scattering surface. Figures 4.15 (c-d) shows the retrieved reflectivity of the hidden scene on the (x-y) plane and on the (x-z) plane. The reflectivity is normalized to the local maxima to facilitate the visualization. The retrieval has been obtain by applying a threshold of 0.89 over the obtained 3D reflectivity. In this case the dimension of each voxel is 0.2x0.2x1 cm^3 for an overall volume to investigate of 20x20x100 cm^3. As reported by the dotted black line in Figs. 4.15(c-d) showing the actual position of the targets on the (x-y) and (x-z) plane, the proposed method provides an accurate 3D retrieval of the hidden scene.

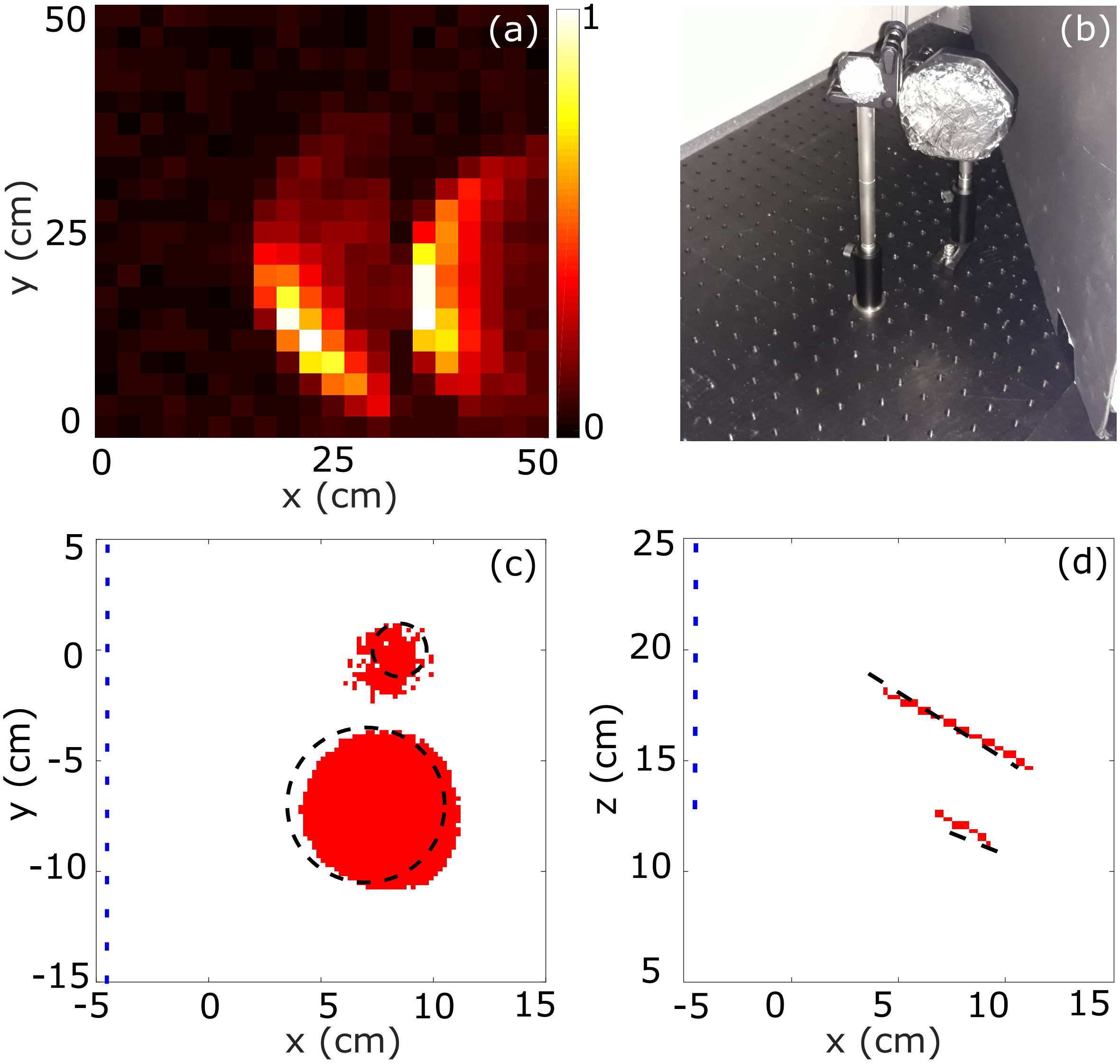

Figure 4.15: **Two retroreflective objects scenario for NLOS imaging by using the single-pixel PMT detector.** (a) Return signal scattered back by the hidden targets on the 20x20 pixels FoV at a time frame of 7.1 ns acquired by raster scanning. (b) Picture of the two retroreflective targets. The picture (b) is for illustrative purpose only and it does not indicates the actual position of the targets. (c-d) Results of the 3D retrieval of the hidden scene on the (x-y) plane and on the (x-z) plane obtained by back-projection imaging algorithm. The black dotted line and the blue dotted line indicate the actual position of the targets and the actual position of the hiding wall respectively [6, 7].

4.4.2 Red Green Blue coloured scenario

The second scenario we investigate is a Red-Green-Blue (RGB) coloured scene with the same experimental setup in Fig. 4.1 by using the PMT single-pixel detector characterized in Section 4.1.2. The target to be retrieved in this case is a rectangular object (Fig. 4.16(b)) where each coloured portion has a rectangular dimension of 20x9 cm^2. In order to recover the coloured scene we use the supercontinuum white-light laser source characterized in Section 4.1.1 with 10 ps pulse duration of 1.5 nJ, 67 MHz repetition rate and 100mW average power in the 450 nm - 700 nm spectral emission range. Since the laser source emits in the all visible spectrum, we select each of the emission wavelength of the RGB colours using band-pass spectral filters centred at 490, 550 and 610 nm (40 nm bandwidth) frequency after the laser source. In this case we run three set of measurements, one per each RGB colour with 20 mW average power each colour by raster scanning. Due to the low power emission, the acquisition time is 10 seconds per mask with an overall acquisition time of 66 minutes.

Figure 4.16 (a) shows the return signal produced by the corresponding colour portion of the hidden target at a time frame of 8 ns. A video of the return signal produced by the hidden object in the red filter measurement is available in the Appendix C.

We then discretize the space in voxels of 1.4x1.4x1 cm^3 for an overall volume of 140x140x100 cm^3 and apply the back-projection algorithm. Figure 4.16 shows the experimental results of the 3D retrieval of the hidden scene after applying the back-projection algorithm. In this case we applying a thresholding of 0.89 over the voxels confidence map. The dotted black line indicates the actual position of the hidden target. As reported in Figs. 4.16 (c-d), the proposed method provides an accurate 3D retrieval in colour of the hidden scene, although with a long acquisition time.

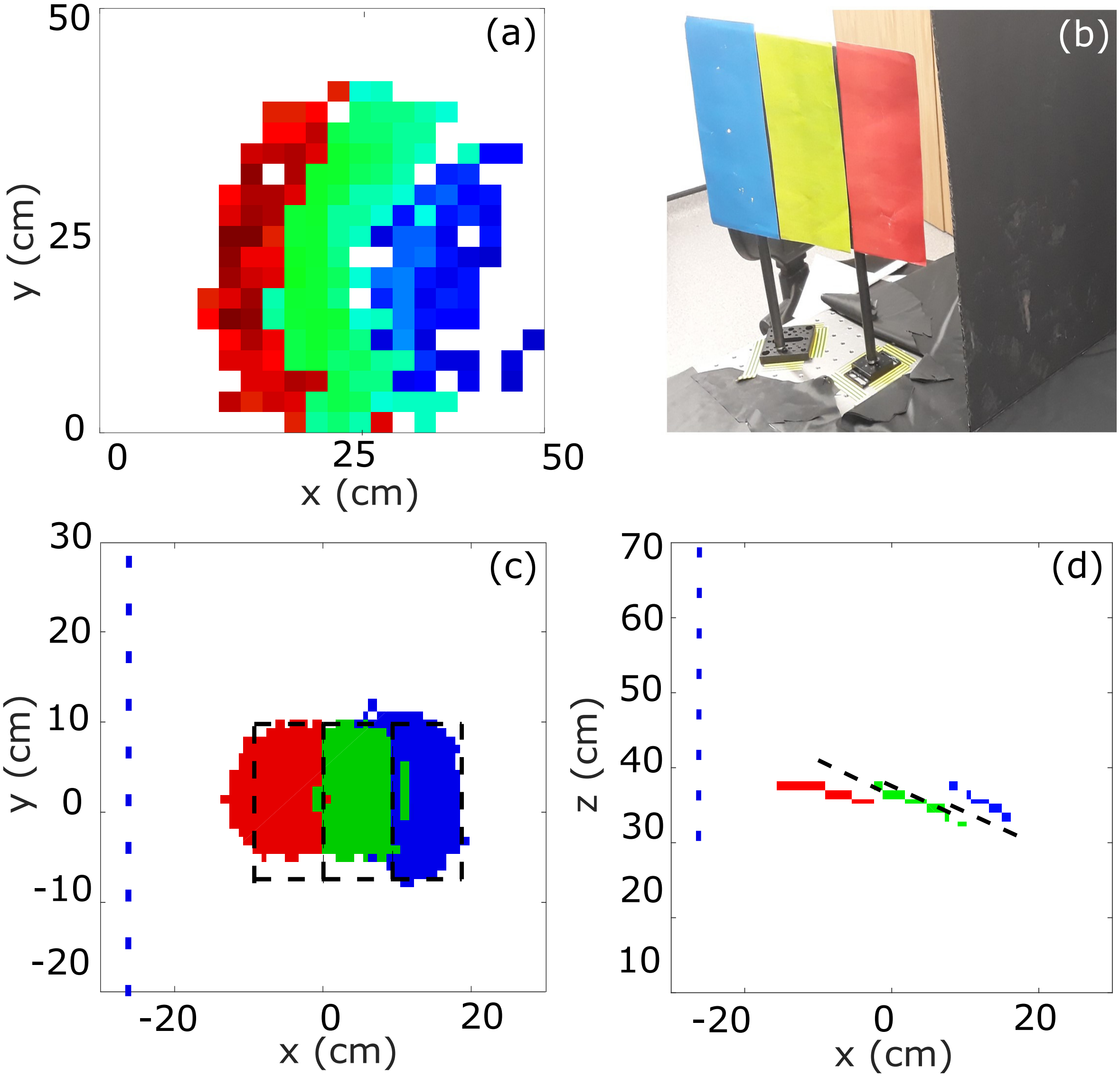

Figure 4.16: **Retrieval of the RGB coloured scenario by using the PMT single-pixel detector.** (a) Return signal scattered back by the hidden targets on the 20x20 pixels FoV at a time frame of 8 ns acquired by raster scanning. The colour indicates the return signal collected by using the corresponding spectral filter. (b) Picture of the hidden RGB coloured target. (c-d) Results of the 3D retrieval of the hidden scene on the (x-y) plane and on the (x-z) plane obtained by back-projection imaging algorithm. The blue colour indicates the recovered target by using the blue spectral filter and so on. The black dotted line and the blue dotted line indicate the actual position of the target and the actual position of the hiding wall respectively [6, 7].

4.4.3 Ultra-fast NLOS single-pixel camera

The previous results demonstrate an accurate 3D retrieval of the hidden scenes, although the acquisition time is not compatible with a non-static hidden scene. In order to reduce the acquisition time to sub-second, we use the high PDE SPAD detector (70% PDE) [2] characterized in Section 4.1.2 and we increase the intensity of the signal collected in each acquisition by applying 20x20 pixels Hadamard patterns where 50% of the pixels are always collecting light from the FoV during each acquisition. In this case the masks are chosen as the first 400 Hadamard patterns. For each Hadamard pattern, one binary mask and its negative are used and combined, leading to a total of 800 patterns. In this case we used the same supercontinuum laser of the previous RGB scenario with a power of 550 mW at 550 nm with the same experimental setup shown in Fig. 4.1. The SPAD sensor used to investigated the scenario has a temporal resolution of 32 ps. Since the SPAD detector has active area of 57x57 μm^2, we use a long working distance microscope objective (Mitutoyo Plan Apo Infinity Corrected Objective, magnification factor 50, numerical aperture 0.55) after the DMD to focus onto the sensor.

In this case, the scenario to investigate is a non-retroreflective rectangular target of 24x10 cm^2 shown in the inset of Fig. 4.17(e). The high sensitivity of the detector and the Hadamard patterns acquisition allows a shorter acquisition time of 1 ms per temporal histogram (i.e. pixel) leading to a total acquisition time of 0.8 seconds.

Figs. 4.17(a-b) show the return signal after weighting each temporal histogram by the corresponding applied Hadamard pattern. Fig. 4.17(a) shows the return signal at a time frame of 5.3 ns. Since the return signal is characterized by a low signal-to-noise ratio due to short acquisition time, we apply a denoising algorithm [8–11] obtaining the filtered return signal shown in Fig. 4.17(b).

We then apply the back-projection algorithm to the filtered return signal dividing the space in 10^6 voxels. Each voxel has a dimension of $1.4x1.4x1cm^3$ for a total voxels space of 140x140x100 cm^3. Figs. 4.17(c-d) show the 3D retrieval on the (x-y) and (x-z) plane after applying a threshold of 0.80 on the reflectivity. To facilitate the visualization the reflectivity is normalized to the absolute maximum. As shown in Figs. 4.17(c-d), the suggested method provides an accurate 3D retrieval even with a sub-second acquisition time, paving the way to real-time 3D NLOS imaging.

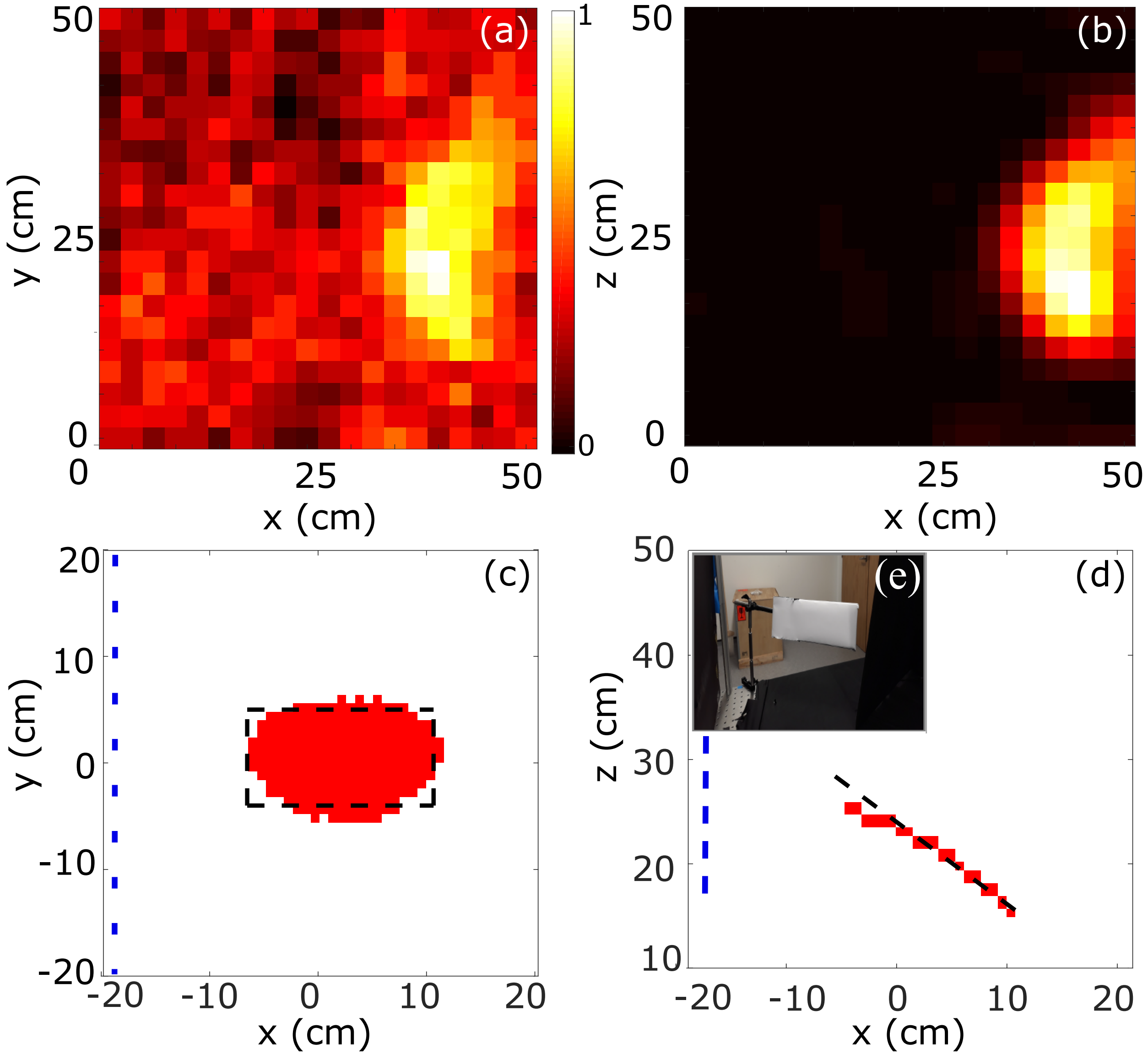

Figure 4.17: **Results of ultra-fast NLOS imaging by using high PDE single-pixel SPAD detector.** (a) Return signal scattered back on the FoV at a time frame of 7.1 ns acquired by applying 20x20 pixels Hadamard patterns. (b) Return signal obtained by applying the denoising algorithm [8–11]. (c-d) Results of the 3D reflectivity of the hidden scene on the (x-y) plane and on the (x-z) plane obtained by the back-projection imaging algorithm on the filtered return signal. The dotted line indicates the actual position of the target [6,7]. (e) Picture of the non-retroreflective white-paper target used for the ultra-fast NLOS imaging. The targets has a rectangular shape of 20x9 cm^2. Using high sensitivity detector (70% quantum PDE) and Hadamard patterns acquisition, the system is able to accurately retrieve the 3D image with a total acquisition time of 0.8 seconds. The black dotted line and the blue dotted line indicate the actual position of the target and the actual position of the hiding wall respectively.

4.4.4 Time-reserval NLOS imaging

We now investigate the 3D retrieval of a hidden scene by considering an experimental scenario where the light propagates in a reverse direction compared to the direction investigated in the previous scenarios. In order to retrieve a hidden scene in a reverse scenario we modify the previous experimental setup by using the DMD in projection.

As reported in Fig. 4.18, a pulsed laser uniformly illuminates the DMD providing a structured illumination onto a 32x32 pixels FoV applying user defined masks. The masks applied in this case are chosen as the first 32x32 Hadamard patterns. The illumination surface has an area of 31x31 cm^2 corresponding to a 8x8 mm^2 pixel area at the scattering surface. For each Hadamard pattern, one binary mask and its negative are used and combined, leading to a total of 2048 patterns. A time-resolving single-pixel detector collects the light passing in a given single-point observation on the scattering surface as opposed to the 20x20 pixels FoV of the previous scenarios. The arrival-time of the return is stored in temporal histograms with 521 time bins of 25 ps duration each. The acquisition time of each temporal histogram is 100 ms leading to a total acquisition time of 3.41 minutes. Collecting the arrival-time of the return at the single-point observation allows to retrieve the 3D information of the hidden scene.

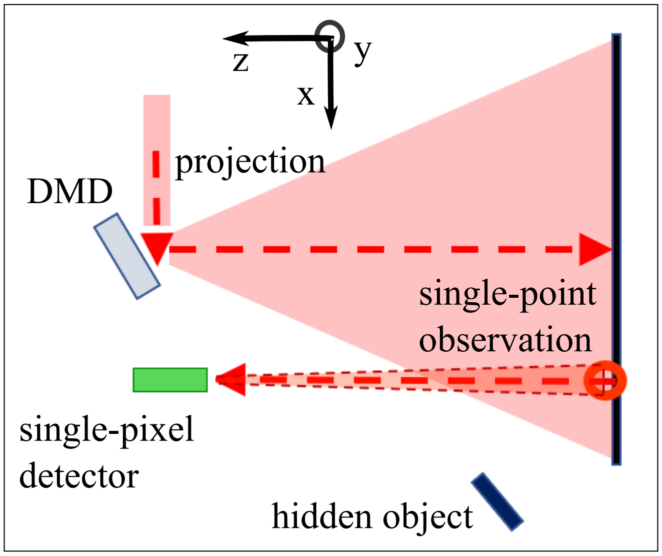

Figure 4.18: **Experimental setup used for the time-reversal NLOS imaging.** As opposed to the previous setup, the DMD provides structured illumination on the scattering surface projecting 32x32 pixels Hadamard patterns. We then collects the temporal back-scattering on a single-point observation area by a time-resolving single-pixel PMT. The signal is stored in temporal histograms and combined with the corresponding pattern to obtain the return signal passing in time across the 32x32 pixels field of view.

Combining each collected temporal histogram with the corresponding Hadamard structured illumination, this configuration can be considered as a "time-reversal" NLOS imaging. Indeed, the reverse scenario can be obtained by backward propagating the light in Figure. 4.1 and swapping the single-pixel detector and the laser. Here, the term "time-reversal" is uniquely intended as the analogy of the two configurations when considering a backward propagation of the light.

Figures 4.19 (a-b) show the analogy between the typical NLOS scenario and its time-reversal considering the Eq. (4.10) which quantifies all the possible locations (x,y,z) of the target patches that contribute to the detected signal. With reference to Fig. 4.19 (a), all the possible location (x,y,z) of the target patches that contribute to the signal collected at a time t_p at a pixel p of the FoV are described by the formula:

$$t_p = \frac{d_{q-Laser} + d_{Obj-q}(x,y,z) + d_{p-Obj}(x,y,z) + d_{sensor-p}}{c} \tag{4.21}$$

Here, $d_{q-Laser}(x,y,z)$ is the distance between the laser spot q on the scattering surface and the laser source. The quantity $d_{Obj-q}(x,y,z)$ is the distance target voxel-laser spot. The quantity $d_{p-Obj}(x,y,z)$ is the distance between the observed pixel and the object and $d_{sensor-p}$ is the distance sensor-observed pixel.

Let's know consider the backward case where the light propagates in a "time reversed scenario". With reference to Fig. 4.19 (b), all the possible locations (x,y,z) of the target patches that contribute to the signal collected at a time t_{sensor} at the single-observation point *Obs* are now described as:

$$t_{sensor} = \frac{d_{p-Laser} + d_{Obj-p}(x,y,z) + d_{Obs-Obj}(x,y,z) + d_{sensor-Obs}}{c} \tag{4.22}$$

Here, $d_{p-Laser}(x,y,z)$ is the distance between the laser spot p on the scattering surface illuminated by the DMD mask and the laser source. The quantity $d_{Obj-p}(x,y,z)$ is the distance between the target's voxel $Obj(x,y,z)$ and the laser spot p. The quantity $d_{Obs-Obj}(x,y,z)$ is the distance between the single-observation point *Obs* on the scattering surface and the target's voxel and $d_{sensor-Obs}$ is the distance sensor-single point observation *Obs*.

Swapping the sensor and the laser position and considering a backward propagation of the light, the time-reversal equation (4.22) is identical to Eq. (4.21) describing a standard NLOS imaging scenario. Since Eq. (4.10) remains unchanged after the backward operation, Eq. (4.10) is time-reversal invariant. This implies that the two scenarios are indistinguishable and the same laws of physics are equally applicable in both cases.

The figure 4.20 shows the portable experimental setup. The laser source used in this scenario is the Toptica FemtoFErb laser emitting 100 fs pulses at central wavelength of 780 nm with a bandwidth

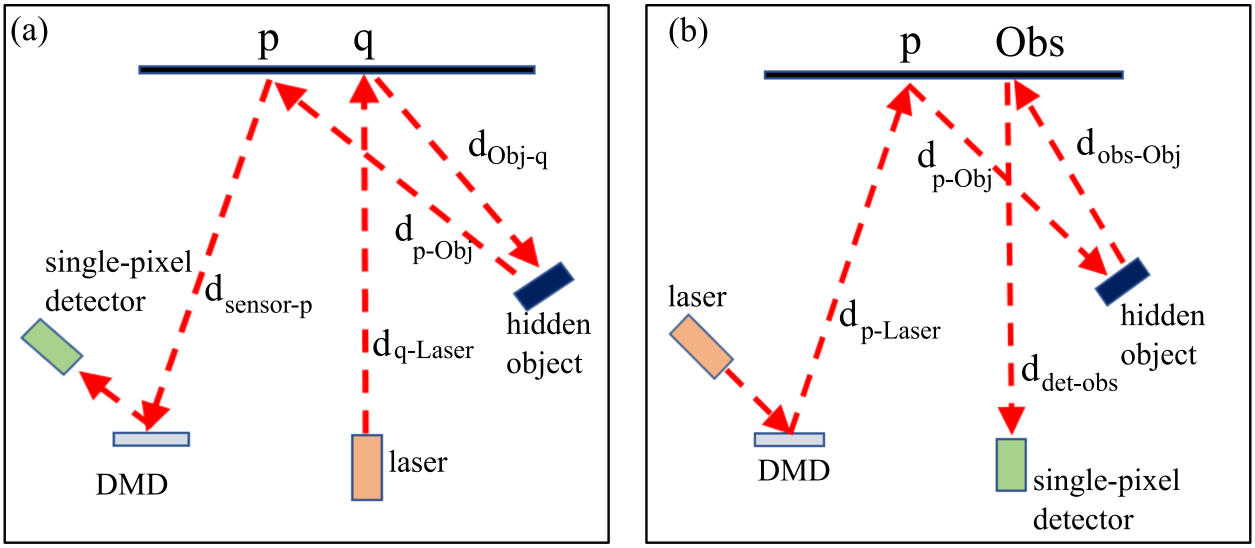

Figure 4.19: **Analogy between the typical single-pixel NLOS imaging scenario and its time-reversal scenario.** (a) The laser sends pulses of light on a spot of the relay wall and the DMD collects the return signal passing towards the 32x32 pixels FoV. The light is then projected on a single-pixel sensor. (b) The DMD projects structured illumination on a 32x32 pixels portion of the relay wall and the single-pixel sensor detects the return signal in a single-point observation of the scattering surface. Swapping the single-pixel sensor and the laser position and using the DMD in pojection, the two scenarios are undistinguishable and they can be described as the same physical laws reversing the flow of the time.

of 10 nm at a 100 MHz repetition rate and 140 mW average power. The single-pixel camera is composed by a DMD (Texas Instruments Discovery 4100 supplied by Vialux, model V-7001) and a single-pixel detector. The time-resolving single-pixel detector is the Picosecond photon detection PPD-900 PMT characterized in Section 4.1.2 with a temporal resolution of 180 ps and a QE of 8%. In this case we used a customized Horiba DeltaHub TCSPC card and the system is mounted in a portable and compact device shown in Fig. 4.20 [12]. The scene to be recovered in this case is a round foam object of 15 cm of diameter (Fig. 4.21(b)).

We then combine each temporal histogram with the corresponding pattern. After obtaining the return signal passing in time across the FoV as a 3D matrix, we apply the back-projection imaging algorithm. In this case the single-observation point has been chosen as the origin $O(x = 0, y = 0, z = 0)$ of the frame of reference. We then divide the space to investigate in 10^6 voxels. Each voxel has a dimension of $1.4x1.4x1cm^3$ for a total voxels space of 140x140x100 cm^3.

Figs. 4.21 (c-d) show the 3D retrieval of the hidden scene on the (x-y) and (x-z) plane. The (x, y, z) axis are oriented as reported in Fig. 4.18. The reflectivity is normalized to the absolute

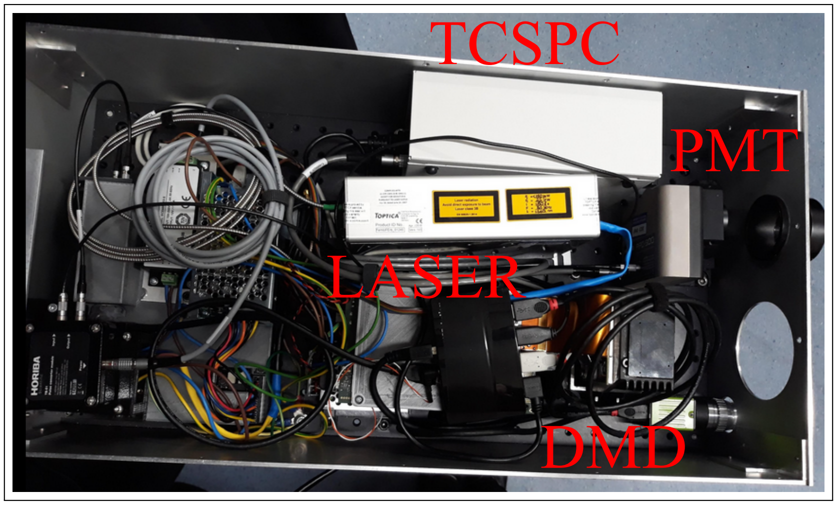

Figure 4.20: **Picture of the portable time-reversal NLOS imaging with single-pixel camera.** A pulsed laser of 100 MHz repetition rate and 780 nm wavelength is sent on a DMD by a fiber (partially visible from the picture). The DMD uniformly illuminates a 31x31 cm^2 portion of the relay wall, providing structured illumination of 32x32 pixels Hadamards patterns by a 35 mm focal length bi-convex lens (not visible in the picture). The distance between the DMD and the relay wall is 162 cm. The TCSPC operating mode, single-pixel PMT detector collects the return signal in a single-observation point of the rely wall. The arrival-time of the photons is then stored in a 521 time bins temporal histogram. The collecting system of the PMT is composed by a camera objective (Nikon, 1.4 aperture) and a focusing lens of 50 mm focal length. In order to collect the return signal in a single-point observation on the scattering surface, a pin hole was placed between the objective and the focusing lens. The distance between the DMD and the detector is 10 cm. The system is mounted inside a box and on a tripod making the system compact and portable. [12]

value to facilitate the visualization and the applied threshold is 0.8. Comparing the actual position of the object (dotted black line) and the retrieval, the proposed method provides a precise 3D reconstruction even in a time-reversal scenario. The results represents an alternative to current 3D NLOS single-pixel imaging systems providing more flexibility and freedom of choice of the optimal setup for the imaging challenge being addressed.

However, the proposed method is affected by some limitations in the spatial resolution of the retrieval. The main limitation is due to the size of the pixels on the observation area. Since the pixels'size on the FoV is 2.6x2.6 cm^2, the temporal resolution is limited to 60 ps measured by considering the 3.6 cm long diagonal of each pixel. This limitation can be overcome decreasing the size of the pixels, at the cost of a greater number of Hadamard patterns and consequently longer acquisition time. A further limitation to the spatial retrieval is the 27 ps temporal resolution of the sensor and the 6 ps temporal width of the laser pulses, inducing respectively a limitation of 0.4 cm and 0.18 cm to the spatial resolution.

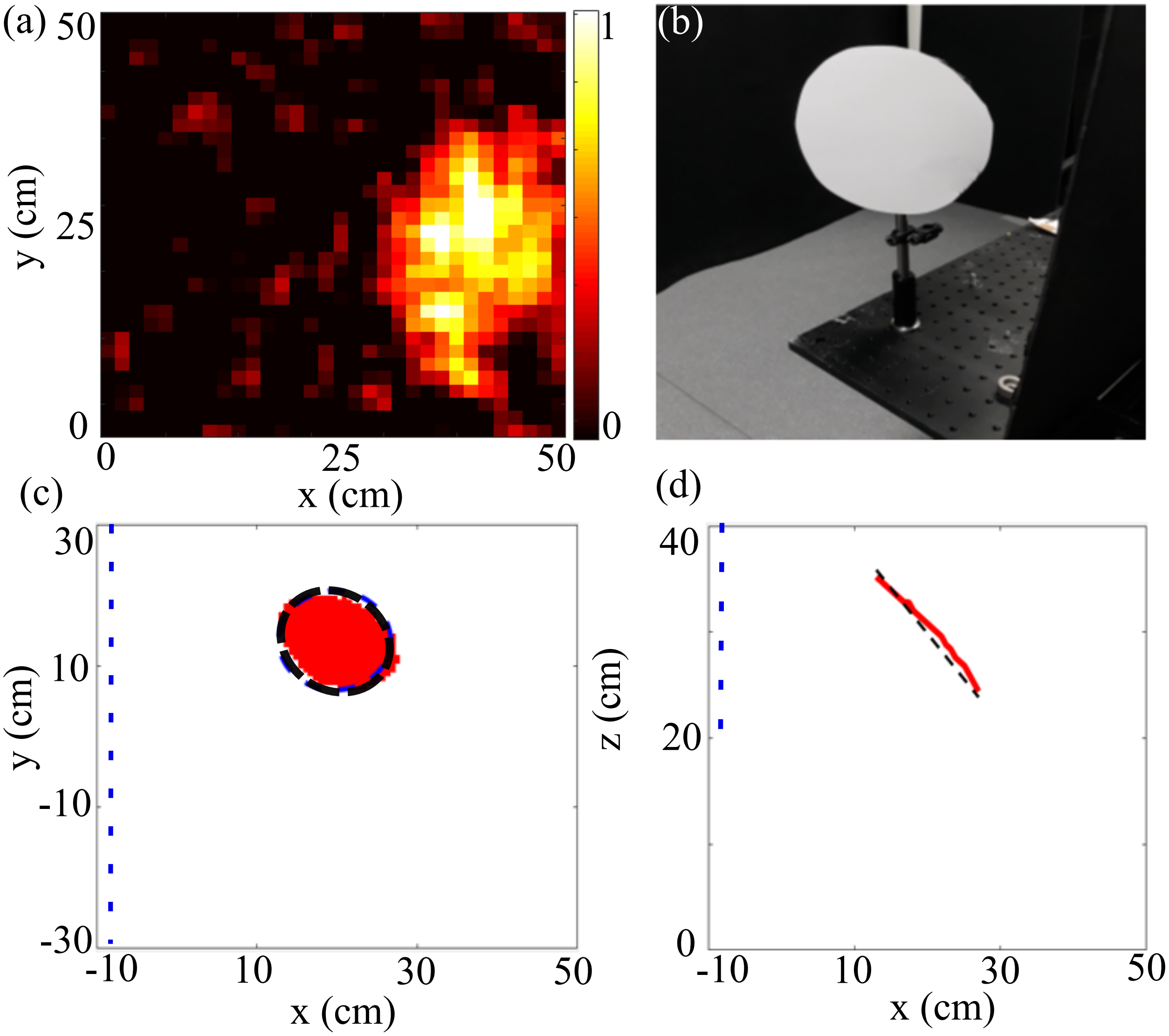

Figure 4.21: Results of the time-reversal NLOS imaging. (a) Return signal scattered back on the 32x32 pixels field of view at a time frame of 8 ns acquired by 32x32 pixels Hadamard structured illumination. The projection surface is a square area of $31x31cm^2$ corresponding to a 8x8 mm^2 pixels area on the relay wall. The total number of applied patterns is 2048 corresponding to the first 32x32 Hadamards patterns and their negative. (b) Picture of the hidden target. The target is a round object of 15 cm of diameter. (c-d) Results of the 3D retrieval of the hidden scene on the (x-y) plane and on the (x-z) plane obtained by the back-projection imaging algorithm. The black dotted line and the blue dotted line indicate the actual position of the target and the actual position of the hiding wall respectively.

4.4.5 Conclusions

The identification of scenes hidden from the direct LOS represents an emerging field of research with application in defence, self-driving vehicles, and surveillance.
In addiction to acoustic systems [102, 103] and speckle correlations [124], current technologies demonstrate the 3D recovery [30] and the real-time tracking [24] of NLOS scenes by exploiting the arrival-time information of the multi-bounce return signal. The 3D information of the hidden scene can then be recovered using NLOS imaging algorithms such as back-projection [107], frequency-wavenumber (or f-k) migration [109] or phasor-field virtual wave methods [125].
Some limitations affect the current NLOS 3D imaging techniques. Indeed the multiple backscattered signal is typically weak and decreases with the inverse of the square distance. Therefore, NLOS methods require high dynamic range, many acquisitions and picosecond time-gated, single photon sensitive detector. Moreover they require complex CI algorithms [30, 112] not compatible with the real-time 3D imaging of non-retroreflective hidden targets. The 3D recovery of NLOS scenes is therefore a current matter of research still impractical for real-time and long range applications.

In this chapter we demonstrated the 3D recovery of hidden scenes with a time resolved single-pixel camera. Combining single-photon sensitive, single-pixel detectors of sub 30 ps temporal resolution with a digital mirror device (DMD) with up to 20 kHz refresh rate, the single-pixel camera approach allows to reduce the acquisition times with good 3D retrieval quality. The proposed method is able to retrieve a 3D image of non retroreflective targets with sub-second acquisition time.
The experimental results demonstrated an accurate 3D information retrieval in colour offering a competitive alternative to conventional cameras. The use of a single-pixel camera provides more flexibility in the choice of the optimal experimental setup without requiring any moving parts such as galvomirrors. The acquisition time can be further reduced fully exploiting the benefits of using a single-pixel camera by applying CI algorithm such as compressive sensing, paving the way to real-time 3D imaging of hidden scenes. Combining next generation of single photon-sensitive detectors, computational algorithm and compressive sensing, the 3D recovery of dynamic NLOS scenes with high spatial resolution could be a reality in the next future.

Chapter 5

Introduction to neural networks

The 3D reconstruction of LOS scenes in real-time represents a crucial task with applications in several real-life scenarios such as autonomous vehicles, robotics and ranging. Current 3D imaging techniques are mainly based on stereovision, holography or arrival-time measurements.

In the next chapters we introduce a data-driven approach for 3D imaging based on arrival-time measurements acquired with a single-pixel, time-resolving detector.

The aim of the neural network (NN) is to find the inverse light transport model transformation F^{-1} that maps the temporal histogram of the return scattered back from the entire scene into the corresponding 3D image.

We now discuss the fundamental principles of NNs and in more details the supervised learning approach adopted to retrieve the 3D image of the scene from the temporal histogram of the return. We infer the inverse transformation F^{-1} training the NN with pairs of temporal histogram and corresponding ground-truth 3D image, as discussed in Chapters 5-7.

5.1 Principles of neural networks

Inspired by human brain neurons, NNs are a series of algorithms designed to extract knowledge and to recognize complex patterns without explicitly providing instructions about the solution of the problem.

Using a data-driven approach, the machine automatically learns its own solution according to the problem to be solved, combining computer science, mathematics and predictive statistics [126].

The most common example of problem solved using NNs is the handwriting recognition, used daily by post-offices to identify delivery addresses. In this case the rules describing the recognition

of handwritten digits are not easy to be expressed by algorithms. At the same time, the recognition of handwritten digits using data-driven approach doesn't require complex NNs or advanced computational resources. Indeed, the handwritten recognition is unconsciously accomplished by humans identifying the rules describing the digits characteristics, as for example a loop at the bottom and a line at the top left for the digit "6". NNs and deep learning approaches solve the handwritten digits recognition using a different strategy. Indeed, NNs automatically infer the rules describing the handwritten digits recognition using a large set of handwritten digits as training examples with a 99.7% of accuracy without human help.
Also known as statistical learning, machine learning and data-driven approaches have been ubiquitous used by many modern websites and providers in everyday life tasks such as spam emails classifier, movie suggestions and online product research. Outside the commercial applications, machine learning has been widely applied in pattern and face recognition in computer vision, security [127], self-driving cars [128], biology [129] and physics [130–133].

Current machine learning techniques require large amount of data in order to provide the correct prediction. Therefore, data-driven approaches have been widely used due to the considerable amount of data that has become available with the increasing use of Internet and social networks. Thanks to recent advances in computing resources such as modern GPUs allowing multiple parallel operations at the same time, machine learning algorithms and data-driven approaches have been also applied to provide costumers suggestions on the online sale. Indeed, current data-driven approaches require the managing and the handling of a continuous steaming of considerable amount of data especially for online recommender systems.
The use of machine learning and NN algorithms offers a series of advantages as opposed to human handcoded rules. The application of data-driven approaches and NN algorithms has played a crucial role in tasks requiring significant amount of time when performed in conventional methods. Moreover the solution inferred using a data-driven approach is not restricted to a specific task or domain, as opposed to the tasks performed by human handcoded rules of "if" and "else" conditions.
The NNs have to provide a generalized prediction independent from the specific characteristics of the data used in the training. For that reason, NNs usually require a representative statistical samples of data.
In order to test the correctness and the general performance of a NN, the observed dataset is usually split in two distinct parts: the training and the testing data, respectively the 75% and the 25% of the overall available dataset. The training set is used to build the machine learning model, whilst

the testing set is used to evaluate the performance of the model.

Of fundamental importance in NN applications is the structure of the observed data and how the data relate to the problem to be solved. According to the structure of the observed data, the learning process can be separated in two distinct leaning approaches: the supervised and the unsupervised learning. In the supervised approach, the dataset is composed by an input set and an output set. Usually used for handwritten digits recognition, fraudulent activities detection or spam classifier, the supervised learning approach recovers the transformation mapping the input into the output from each provided training example. In the training process, the NN finds the transformation that maps the input into the corresponding output. At the end of the learning process, the NN provides a predicted solution of the output testing data by using only the input testing data. Finally, we evaluate the performance of the NN comparing the predicted solution with the corresponding ground truth i.e. the output set. Once the NN has been trained, the algorithm is able to provide a prediction of the output for a new input never observed by the NN. A typical data structure for a spam classifier is composed by pairs of emails-flags where the flag 1 or 0 classifies the corresponding email as spam or not.

In the unsupervised leaning approach, the dataset is composed by only the input data and no output data are used in the training. The unsupervised learning approach is usually used in cluster identification problems. Principal news providers such as Google employ the unsupervised learning in order to separate the news in groups according to the topic or to separate costumers in groups with similar preferences.

5.2 Architecture of neural networks

Artificial neural networks (ANN) and deep learning represent two of the best approaches currently used for problems complex to be solved with conventional methods without requiring human intervention. The computer automatically figures out the best solution according to the diversified dataset provided during the learning process.

The typical scenario is a set of m-dimensional inputs $\{x\}$ mapped into a set of n-dimensional outputs $\{y\}= f(\{x\})$ by an unknown transformation f. The purpose of the NN is to find the unknown transformation f usually represented by a function or a matrix. The solution f found using NNs is encoded in the *learning parameters* weights w and bias b characteristic of the problem. Once the system has been trained, the algorithm is able to provide its own solution of the problem using single or multi-layer of learning neurons. As in the human brain, the neurons represent the unit structure of the ANN and they are interconnected by the learning parameters. We now describe the different types of neurons and learning architecture commonly used in NNs.

5.2.1 Perceptrons

Inspired by human brain, the most fundamental structure of the NNs is the perceptron, an artificial neuron developed by Franck Rosenblatt in the 1950s [134]. As shown in Fig. 5.1, a perceptron takes a series of binary inputs x_1, x_2, x_3 and computes a single binary output y by a set of weights w_j. The weights w_j are usually real numbers expressing the importance of each input on the computed output according to the formula:

$$output\ y = \begin{cases} 0 & \text{if } \sum_j w_j x_j \leq b \\ 1 & \text{if } \sum_j w_j x_j > b \end{cases} \tag{5.1}$$

where the quantity b is a given parameter of the neuron identifying the threshold condition to activate the perceptron. The previous equation can be easily written as follows:

$$output\ y = \begin{cases} 0 & \text{if } w \cdot x + b \leq 0 \\ 1 & \text{if } w \cdot x + b > 0 \end{cases} \tag{5.2}$$

The output of the perceptron is therefore 0 (inactive state) or 1 (active state) according to whether the weighted sum $\sum_j w_j x_j$ is less than or greater than a given threshold b. In general, the output of the neurons is described by a function of z=$w \cdot x - b$ called "activation function". The activation function of a perceptron is then the step function shown in Fig. 5.2.

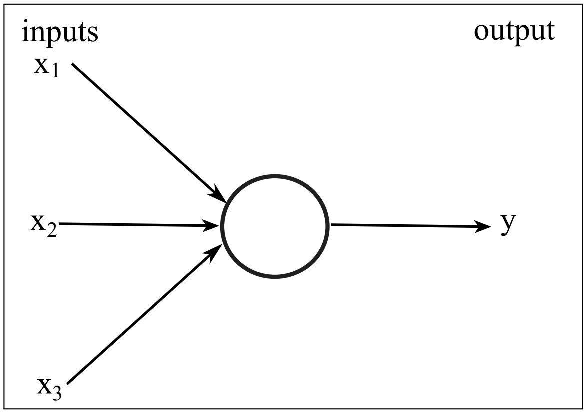

Figure 5.1: **Graphical representation of a perceptron.** As the fundamental structure of ANNs, the perceptron makes decision weighting a series of binary inputs x_1,x_2 and x_3 and providing a binary outcome y at the output.

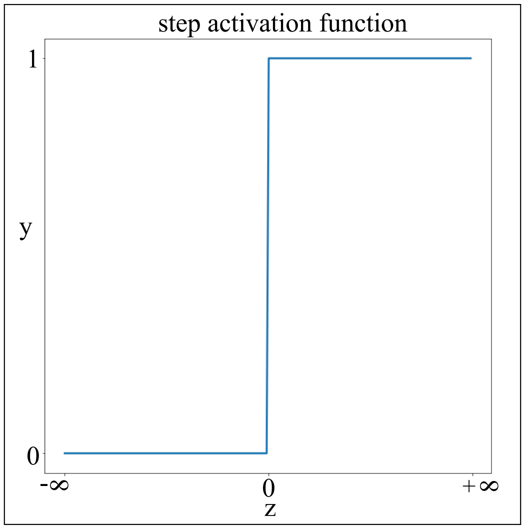

Figure 5.2: **Step activation function for z=wx-b.** The activation function of a perceptron of inputs *x* and weights *w* is the step function.

5.2.2 Sigmoid and hyperbolic tangent activation functions

Perceptrons are highly sensitive to any changes applied to the weights and to the threshold of the neuron. A small change in the parameters of any perceptron of a NN may cause the flipping of an output, drastically changing the final outcome of the NN. The step function can be therefore modified considering a smoother activation function such as the sigmoid function.
Fig. 5.3 shows the sigmoid function described as:

$$\begin{cases} \sigma(z) = \frac{1}{1+e^{-z}} \\ z = wx - b \end{cases} \tag{5.3}$$

Using the sigmoid function as the activation function of a neuron, we then obtain a more complex type of neuron called the sigmoid neuron. This type of neuron takes in input values between 0 and 1 and is less sensitive to the changes of the neurons parameters.
Considering a sigmoid function as the activation function of a neuron, the sigmoid neuron approximates the behaviour of the perceptron. For large negative values of z, the quantity e^{-z} is infinite and the neuron output is zero. For large positive values of z, the quantity e^{-z} is zero and the neuron output is 1, as described in the perceptron model.
A sigmoid neuron and a perceptron differ in the accepted range of values of the output y. A perceptron accepts only binary values, whereas a sigmoid neurons accepts any real number between 0 and 1, as used to model the intensity of the pixels in an image recognition NN.

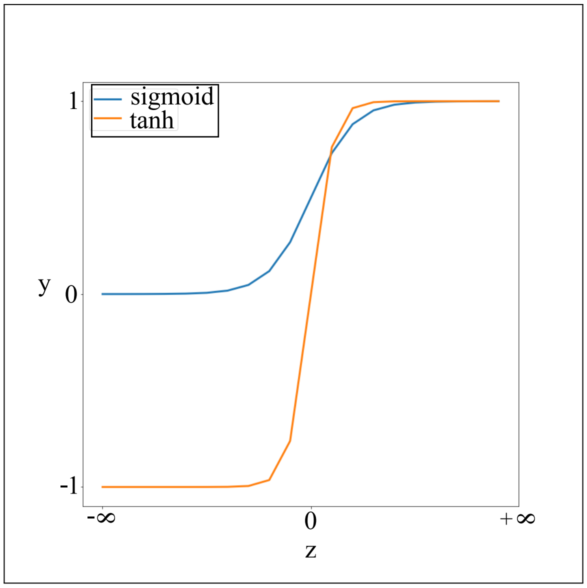

Figure 5.3: **Sigmoid and hyperbolic tangent activation function for z=wx-b.** (blue) Using a sigmoid as the activation function of the perceptrons, we can obtain a sigmoid neuron, less sensitive to changes of the neutron parameters than a perceptron. Small changes in the neuron parameters *w* and *b* induce a small change in the neuron output. (Orange) Hyperbolic tangent activation function defined from -1 for low z values, to +1 for high z values.

Since nonlinear activation functions allow the NN to learn much more complicated functions mapping the input $\{x\}$ into the output $\{y\}$, a wide range of nonlinear activation functions can be used to define the neurons characteristics.

One of the most common activation function is the hyperbolic tangent defined as

$$\begin{cases} tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \\ z = wx - b \end{cases} \tag{5.4}$$

As shown in the orange graph in Fig. 5.3, the hyperbolic tangent is similar to the sigmoid but it goes from -1 for low z values, to +1 for high z values.

The complete list of the possible neurons activation functions used in ANNs can be found in the Keras documentation website [135].

The perceptron and other types of neurons represent a device that makes decisions weighting up a series of observed inputs and computing the value of a given activation function. Therefore, we can simulate a wide range of different decision-making models varying either the activation function or the values of the *learning coefficients b* and *w*.

5.2.3 Multi-layer perceptrons

In order to create more complex decision-making models and simulate problems applicable to real-life scenarios, more elaborated structures can be created consecutively adding multiple perceptrons.

Figure 5.4 shows an example of NN made by multiple layers of perceptrons. In this case each column of perceptrons, or layer, computes the output taking in input the outputs of the previous layer. Therefore each layer has its own characteristics parameters *w* and *b* for every unit of the layer. With reference to Fig. 5.4, each neuron of the first layer is then characterized by four input weights plus one bias parameters, for a total of five parameters per neuron and therefore fifteen learning coefficients for the first layer.

The NN models discussed so far represent examples of feed-forward NNs where the information is moving forward the network. The input of a layer is always taken from the output of a previous layer. In recurrent NNs instead, the input of a layer depends on its own output using feed-back loops. Here, the neurons get activated for a limited amount of time, stimulating other neurons that get activated later in time and creating a cascade of firing neurons.

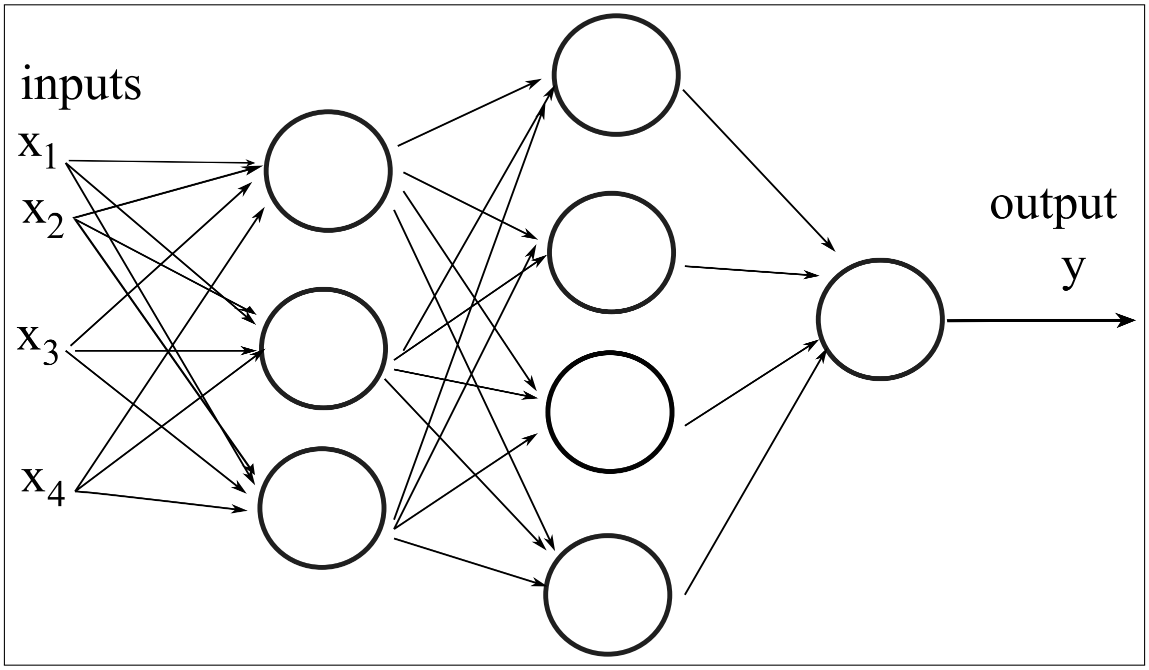

Figure 5.4: **Graphical representation of a multi-layers perceptron.** More developed NN model describing plausible real-life scenarios can be simulated building more complex decision-making models of many consecutive layers. Each layer of perceptrons takes in input the outputs of the previous layer and so on, building a characteristic set learning parameters for each layer.

According to how the neurons of the layer are connected to the neuron in the next layer, different type of layers can be defined. In a dense layer, every input neuron is connected to every output by linear combination of multiple real-value weights and a non-linear activation function. The NN used to recover the 3D image of the scene from the return temporal histogram is composed by a series of dense layers.

One of the most relevant benefits in using NNs is the computational universality of the perceptrons as they can be used to compute any logical function [136]. As an example of the computational universality of the multi-layer perceptrons, we report the implementation of the NAND gate using perceptrons with a chosen set of learning coefficients. The logical operation of a NAND gate for two binary inputs x_1 and x_2 is reported in Tab. 5.1. Figure 5.5(a) shows the equivalent of the NAND gate using a perceptron where the number inside the circle of the perceptron represents the bias $b = 3$ and the numbers above each arrow represents the corresponding weights -2,-2.
Using the given set of learning coefficients, the perceptron computes a NAND gate. The inputs 00, 01 and 10 have 1 as output, whereas the inputs 11 has 0 as output. Since any logical operations can be computed using a set of multiple NAND gates, we can use networks of perceptrons to compute any logical function.
As an example of the computational universality of the perceptrons, we report the bitwise sum of two binary inputs $x_1 \oplus x_2$ using a set of multi-layer perceptron with weights -2,-2 and bias $b = 3$. The logical operation of a bitwise gate for two binary inputs x_1 and x_2 is reported in Tab. 5.2. With reference to the notation introduced for the NAND gate, Fig. 5.5(b) shows the equivalent of the bitwise gate using a set of NAND gate multi-layer perceptrons. In this case two outputs are required in order to compute the bitwise sum. One output computes the sum of the two inputs, whereas the second output computes the carry bit set to 1 when both the inputs x_1 and x_2 are 1. Using the given set of learning coefficients, the perceptron computes a bitwise sum gate. The inputs 00, 01 and 10 produce a carry bit of 0 and a sum bit of 0,1,1 respectively. The input 11 produces a carry bit of 1 and a sum bit of 0 as output.

Since the perceptrons can simulate any computational function, NNs can be as powerful as any other computing device. However, the most revolutionary benefit of using NNs is that the system can automatically tune the learning parameters encoding the function f. Usually used for function complex to be found by conventional methods, the system automatically finds the solution of the given problem learning the solution parameters according to the training data.
For the purpose of this thesis, the primary focus here is the supervised learning approach where

the data provided to the NN during the learning process contain either the input $\{x\}$ and the corresponding output $\{y\}$. As happens in the supervised learning approach, the experimental data will be split in two parts. A part of data will be used for the learning process. The remaining part will be used to evaluate how well the NN has learnt by testing the model on data that have not been used in the learning process. Once the system is trained, we will use the NN to predict the most correct solution according to the learning model.

x1	x2	NAND
0	0	1
1	0	1
0	1	1
1	1	0

Table 5.1: Logic function of the NAND gate

x1	x2	sum	carry bit
0	0	0	0
1	0	1	0
0	1	1	0
1	1	0	1

Table 5.2: Logic function of the bitwise sum $x_1 \oplus x_2$

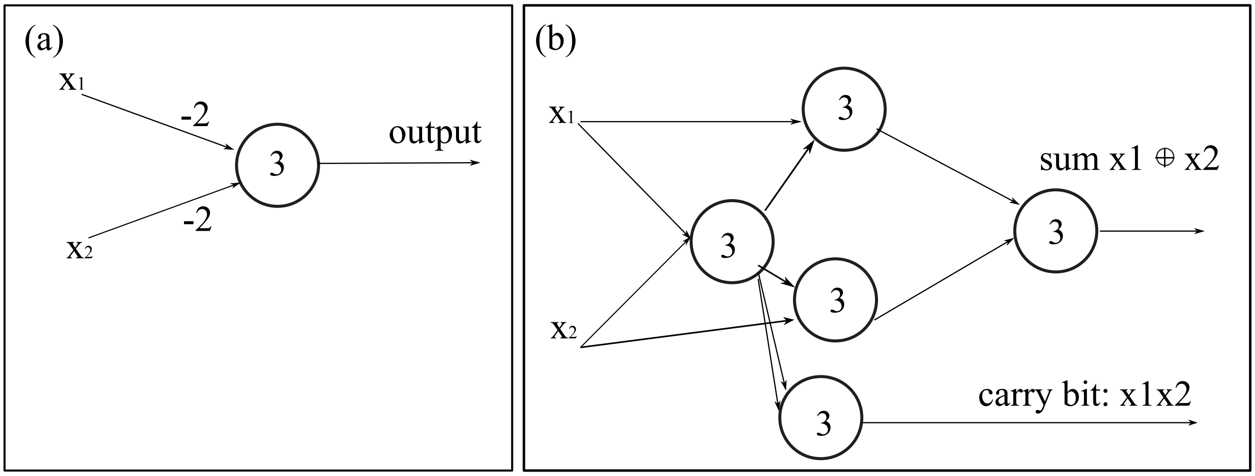

Figure 5.5: **Graphical representation of examples of elementary logic functions using multi-layers perceptrons.** Since the NAND gate is universal for any computational operation, any logical function can be computed using a network of NAND gate perceptrons. (a) NAND logical gate using a single perceptron. With reference to the input-output values shown in Tab. 5.1, the perceptron computes the same operation of a NAND gate using the set of weights -2,-2 and a bias of 3. (b) Bitwise sum $x_1 \oplus x_2$ using multi-layer perceptrons. With reference to the input-output values shown in Tab. 5.2, the first output *sum* $x_1 \oplus x_2$ computes the bitwise sum of the two input x_1 and x_2. The second output computes the carry bit set to 1 when both inputs are 1. The numbers inside the circles represent the bias b, whereas the numbers above the arrows represents the weights w_i of the corresponding perceptron.

5.2.4 Gradient descent

We now describe the learning process of the NNs. The purpose of the NNs is to find the transformation f mapping the input set $\{x\}$ into the corresponding output set $\{y\}$ encoding the solution by a set of learning parameters $\{w\}$ and $\{b\}$.

In order to quantify the goal of NN, we consider the mean square error between the predicted values $f(x_i)$ and the corresponding ground truth $y(x_i)$ of the training examples defined as the following *cost function*:

$$J(\mathbf{w},\mathbf{b}) = \frac{1}{2m}\sum_{i=1}^{i=m} \|y(x_i) - f(x_i)\|^2 \qquad (5.5)$$

Here, x_i and $y(x_i)$ are respectively the i-th input and the corresponding ground truth of the training dataset and i goes from 1 to the total number m of the training data. The function $f = \sigma(\mathbf{w}x + \mathbf{b})$ is the transformation to be found by the NN and defined by the learning parameters $\mathbf{w}$ and $\mathbf{b}$ denoting the collection of all the weights and the biases of the NN. The previous formula quantifies how well the predicted solution fits the observed data comparing the output $f(x_i)$ predicted by the model with the corresponding ground truth $y(x_i)$.

In mathematical terms, the goal of the NN is to find the combination of weights $\mathbf{w}$ and biases $\mathbf{b}$ in order to minimize the cost function $J(\mathbf{w},\mathbf{b})$. During the training process we then apply minimization algorithms in order to find the weights and the biases so that $J(\mathbf{w},\mathbf{b}) \approx 0$.

The most common example used for minimization problem using gradient descent is the convex function $J(w_1,w_2) = w_1^2 + w_2^2$ where w_1 and w_2 are the two independent variables. Figure 5.6 shows its graphical representation. In this case we consider a two variables simplified version of a typical NN cost function where w_1 and w_2 represent the learning parameters of the learning architecture.

In order to find the values $\tilde{w_1}$ and $\tilde{w_2}$ that minimize the cost function $J(w_1,w_2)$, we apply gradient descent as the algorithm simulating the motion of a ball rolling down the slope of a valley from a randomly chosen starting point. The red curves in the plot represent the contour lines along which the cost function $J(w_1,w_2)$ has a constant value for different combination of w_1 and w_2.

As opposed to the mean square error, other functions such as the cross-entropy can be used to model the cost function to be minimized. The cross-entropy is defined as

$$J = -\frac{1}{m}\sum_x \big(ylog(a(x)) + (1-y)log(1-a(x))\big) \qquad (5.6)$$

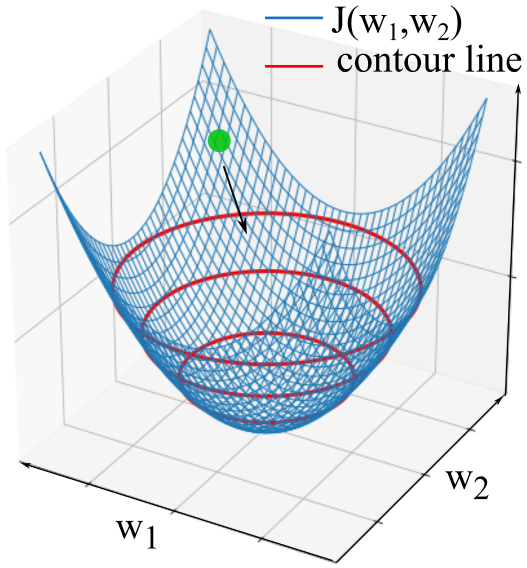

Figure 5.6: **Convex cost function of w_1 and w_2.** The gradient descent can be considered as an algorithm simulating the motion of a ball rolling down the slope of a valley. The red circles represent the contour lines along which the cost function has a constant value of 0,0.2, 0.4 and 0.6 for different combinations of the independent variables w_1 and w_2. Using minimizing algorithms we then compute the values $\tilde{w_1}$ and $\tilde{w_2}$ that minimize the cost function $J(w_1, w_2)$.

where

$$\begin{cases} a = \sigma(z) \\ z = wx - b \end{cases} \tag{5.7}$$

Since the contribution to the cost function is low if the actual output of the neuron is close to the desired output for all training inputs, the cross entropy represents a good cost function, especially in the slow learning case.

The gradient descent algorithm is then used to minimize the cost function. The gradient descent algorithm iteratively computes the derivatives of the cost function respect to the learning parameters. It then updates the values of the learning parameters at each steps according to the steepest direction of the slope of the cost function, or in this case the valley. Formally, when we move the ball by an amount Δw_1 and Δw_2 respectively in the w_1 and w_2 direction, the cost function changes by the amount ΔJ

$$\Delta J \approx \nabla J \cdot \Delta w \tag{5.8}$$

where

$$\nabla J = \left(\frac{\partial J}{\partial w_1}, \frac{\partial J}{\partial w_2}\right)^T \tag{5.9}$$

is the gradient vector of the cost function $J(w_1,w_2)$ respect to the independent variables w_1 and w_2. T is the transpose operation and the quantity

$$\Delta w = \Big(\Delta w_1, \Delta w_2\Big)^T \tag{5.10}$$

is the ball displacement along w_1 and w_2. Here, Eq. (5.8) relates the changes in $w = (w_1,w_2)$ to the changes in the cost function $J(w_1,w_2)$.

In order to guarantee the decreasing of the cost function at each iterative steps, we choose Δw as follows:

$$\Delta w = -\eta \nabla J \tag{5.11}$$

where the learning rate η of the algorithm is a small positive parameter and quantifies how fast the cost function decreases. From the previous equation, we obtain the following equation

$$\Delta J \approx -\eta \nabla J \cdot \nabla J = -\eta (\|\nabla J\|)^2 \leq 0 \tag{5.12}$$

that guarantees the decreasing of the cost function at each iterative steps. Using

$$w \to w' = w - \eta \nabla J \tag{5.13}$$

as the updating rule for the learning parameters at each iteration steps, we therefore obtain the minimum of the cost function.

In a more general scenario, J is a function of many m variables $w_1, w_2, .., w_m$ where

$$\Delta w = \big(\Delta w_1, \Delta w_2, ..., \Delta w_m\big)^T \tag{5.14}$$

and

$$\nabla J = \big(\frac{\partial J}{\partial w_1}, \frac{\partial J}{\partial w_2}, .., \frac{\partial J}{\partial w_m}\big)^T \tag{5.15}$$

Expressing the updating rules at each iteration steps by the original learning parameters **w** and **b**, we obtain the following equations expressing the gradient descent algorithm:

$$w_k \to w_k' = w_k - \eta \frac{\partial J}{\partial w_k} \qquad b_l \to b_l' = b_l - \eta \frac{\partial J}{\partial b_l} \tag{5.16}$$

Updating the learning parameters at each iteration step according to the previous equations, we obtain the best transformation f that maps the set of inputs x into the set of outputs y.

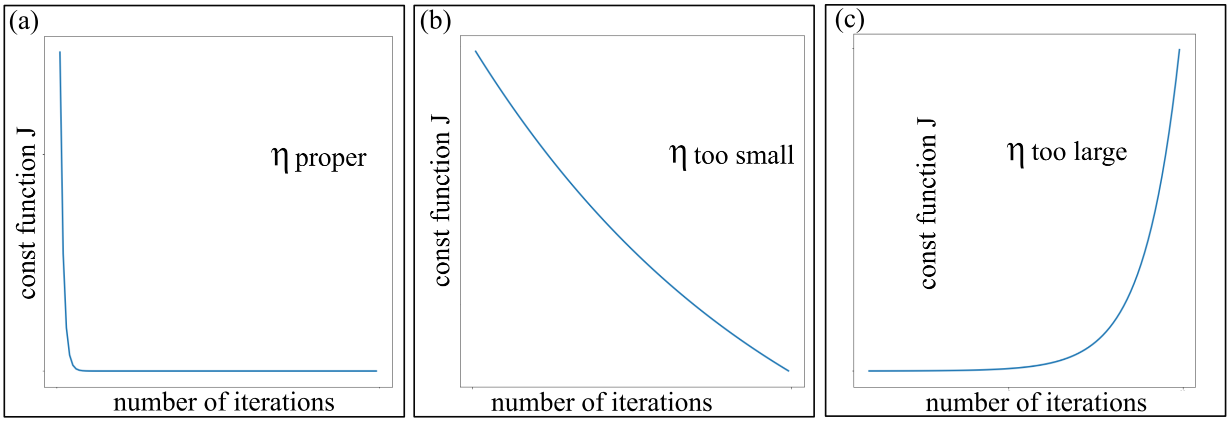

Figure 5.7: **Trend of the cost function *J* as a function of the number of iterations for different values of the learning parameter η.** The learning rate has to be properly chosen in order to minimize the cost function in reasonable time according to the purpose of the neural network. (a) Using a proper learning rate η, the cost function increases at each iteration step and it reaches the minimum value in a reasonable amount of time. (b) Using a too small learning rate η the learning occurs slowly and the cost function takes considerable amount of time and number of iterations to converge to the minimum value. (c) If we use a too large learning rate, the gradient descent may skip the local minimum and the cost function may even increase with the number of iterations.

In order to retrieve an accurate solution of the problem to be solved, the learning process has to be iterated multiple times. The number of *epochs* of the model defines the number of times that the learning algorithm updates the parameters applying gradient descent to the entire training dataset. The learning rate η has to be properly chosen in order to minimize the cost function J in a reasonable amount of time according to the complexity of the training process.

Figure 5.7 shows the cost function as a function of the epochs for different learning rates. An appropriate learning rate η is such that the cost function is quickly moving forward its minimum at each iterations step (Fig. 5.7(a)). Using a too small learning rate, the cost function slowly converges to its minimum at each iterations step and the learning process requires considerable amount of time and number of epochs (Fig. 5.7(b)). On the contrary, the gradient descent may skip the local minimum and the cost function may even increase using a too large learning rate η (Fig. 5.7(c)). Additionally, the learning rate can be gradually tuned during the learning process in order to improve the convergence of the cost function. We can indeed use a large η when the cost function is far from its minimum and a small η when the cost function is close to its minimum, with a consequent improvement in the solution and in the time resources.

However, two of the main disadvantages of using NNs are the risks of overfitting (or high bias) or of underfitting (or high variance). In the underfitting case, the training data are not enough to predict an accurate solution and the system requires more training examples. In the overfitting case, we either reduce the number of epochs of the training process or use a variable learning rate.

In order to evaluate the appropriate number of epochs, the set of available data are usually split in three separate sets of training, validating and testing data accordingly to their role in the learning process. Considering input-output pairs of examples, the training data are generally used in the learning process to update the learning parameters. The validation data are used to evaluate the hyper-parameters of the NN such as the number of epoch of the training process. Finally the testing data never used during either the training and the validation, are used to evaluate the learning model.

We then consider the *Error* quantifying the ability of the model to generalize and produce correct output samples $\{y\}$ for input samples $\{x\}$ never seen during the training as

$$Error(\{x\},\{y\}) = \frac{1}{m}\sum_{i=M}\big(y - f(x)\big)^2 \tag{5.17}$$

where M is the number of the total validation examples, $f(x)$ is the predicted output for the input x and y is the actual output for the same input example.

In order to evaluate the overfitting problem and reduce the number of epochs, we then consider the *Error* on the training examples and on the validation examples as a function of the number of training epochs. In case of overfitting, the *Error* on the training examples is continuously decreasing. On the contrary, the *Error* over the validation examples decreases until a given number of epochs and then it increases, as reported in Fig. 5.8. Since the training error is computed on the data used in the training, the NN indeed continues to learn and to decrease. On the contrary, since the validation error is computed on data not used in the training process, the neural model does not provide a generalized solution for the unseen data and therefore the validating error starts to increase. We then identify the number of training epochs as the number of training iterations just before the validation error starts to increase, maintaining the integrity of the testing examples. Once the validating error reaches the minimum, we can either stop the training process or use a smaller learning rate.

We now discuss the different models of gradient descent that can be used during the learning process according to the scale of the available training examples.

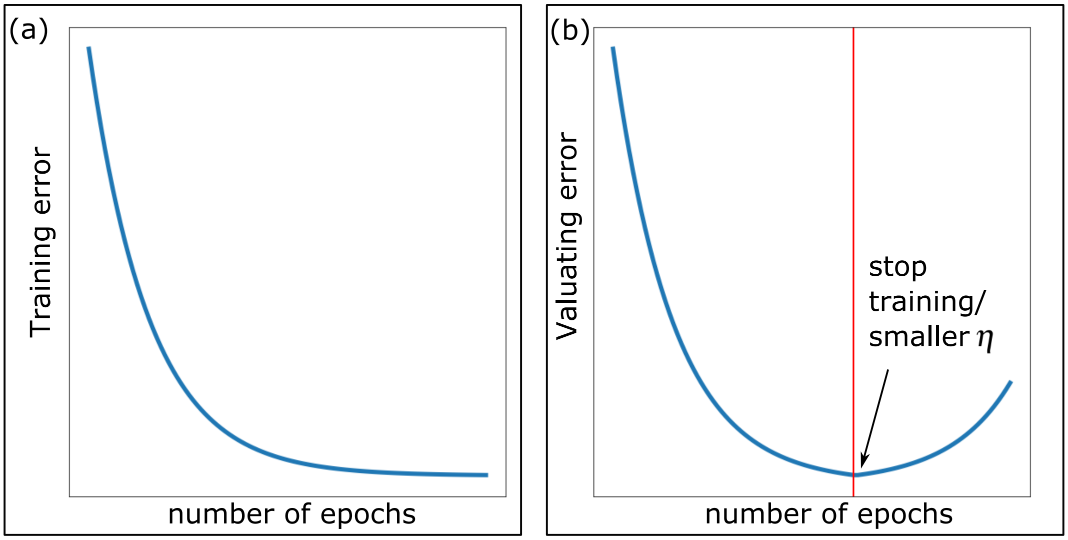

Figure 5.8: **Trend of the training and the validating error as a function of the number of epochs of the training in the overfitting case.** Since the training error is computed summing over the training examples used in the learning process, the training error is continuously decreasing with the number of epochs. On the contrary, the validation error is computed using data that have never been seen during the learning process and it begins to increase after a certain training iteration. We then consider an accurate number of epochs as the number of training iterations just before the validation error starts to increase.

5.2.5 Learning with large datasets

Data-driven approaches such as NN algorithms are usually applied to big dataset used to train the model. Thanks to the increasing development of worldwide web resources available today, modern dataset are characterized by hundreds of millions of training examples, as happens for online recommendation systems of popular websites. Due to the large amount of the available training examples, the application of the gradient descent algorithm on the entire dataset at each iteration is computationally expensive and not efficient in terms of time resources. Moreover, the time required to process large scale data is highly incompatible with the continuous streaming of training data used by online recommendation systems. However, it is possible to select the number of training examples used to updated the learning parameters. According to the number of training examples used to updated the learning parameters, we distinguish three different types of gradient descent: stochastic, mini-batch, or batch gradient descent.

According to the batch gradient descent algorithm, the cost function is described as follows:

$$J(\mathbf{w},\mathbf{b}) = \frac{1}{2m}\sum_{i=1}^{i=m} \|y(x_i) - f(x_i)\|^2$$

Since the cost function is computed summing over the entire training examples, the previous formula can be rewritten as:

$$J(\mathbf{w},\mathbf{b}) = \frac{1}{2m}\sum_{i=1}^{i=m} J_i(\mathbf{w},\mathbf{b}) \tag{5.18}$$

where $J_i(\mathbf{w},\mathbf{b})$ is the contribution of each training examples. Considering the cost function as a sum over the contribution of each training examples, a single iteration of gradient descent described in Eqs. (5.8)-(5.9) is then performed calculating the sum of the m derivative terms over the full set of training data:

$$\nabla J = \frac{1}{2m}\sum_{i=1}^{i=m} \nabla J_i(\mathbf{w},\mathbf{b}) \tag{5.19}$$

Since a single iteration of gradient descent is a sum of m derivatives terms, the learning process with large datasets occurs slowly and a single step of the batch gradient descent may requires considerable amount of time. The computational resources are more prohibitive considering the large number of epochs required to obtain a reliable prediction model in large dataset. However, the learning process can be improved updating the learning parameters at each training examples, as happens in the stochastic gradient descent or in mini-batch gradient descent.

According to the mini-batch gradient descent algorithm, the cost function and the gradient are computed summing over a mini-batch subset of n training examples randomly chosen from the entire dataset. The learning parameters are therefore updated using only the examples of the mini-batch for each gradient descent iteration step. The learning parameters updating rules of Eqs. (5.16) are then expressed as follows:

$$w_k \rightarrow w'_k = w_k - \frac{\eta}{n}\sum_{i=1}^{n}\frac{\partial J_{xi}}{\partial w_k} \qquad b_l \rightarrow b'_l = b_l - \frac{\eta}{n}\sum_{i=1}^{n}\frac{\partial J_{xi}}{\partial b_l} \tag{5.20}$$

Here, the index i goes from 1 to the number of the mini-batch training examples n and the term J_{xi} is the contribution of each mini-batch inputs to the cost function. Since less training examples are used at each iteration of the learning process, mini-batch gradient descent is more efficient than batch gradient descent in terms of computational and time resources. Once the parameters are updated, the NN uses a different subset for the next updating step and so on until the subsets run out. An epoch is then completed iteratively repeating the learning process using all the subset of the dataset.

The mini-batch gradient descent model can be further improved considering mini-batch training examples of n=1 input example, as happens for the stochastic gradient descent. Mainly using for online recommendation systems, in the stochastic gradient descent the cost function and the gradient are indeed computed using a single training example. The learning parameters are therefore updated using a single training example for each gradient descent iteration step. Once the learning parameters are updated, the epoch is then completed iteratively repeating the learning process individually using all the training inputs of the dataset.

5.2.6 Back-propagation algorithm

Introduced in 1986 [137], the back-propagation algorithm computes the expression of the partial derivatives $\partial J/\partial w$, $\partial J/\partial b$ of the cost function J respect to the learning parameters w and b, expressing how quickly the cost function changes with the weights. We now describe the back-propagation algorithm using the following notation.

With reference to the four layers NN of Fig. 5.9, the leftmost layer of the NN is called *input layer* and it contains the input neurons. The two middle layers are called *hidden layers* and the rightmost layer containing the *output neurons* is called *output layer*.

We denote the term w^l_{jk} of the NN as the weight connecting the j-th neuron of the $l-1$-th layer to the k-th neuron of the l-th layer. According to this notation, the weight w^3_{34} connects the third neuron of the second layer to the forth neuron of the third layer.

We use a similar notation for the biases parameters b^l_j identifying the bias of the j-th neuron in the l-th layer.

We then define the quantity a^l_j as the output of the layer l with input x as follows:

$$a^l_j = \sigma\big(\sum_k w^l_{jk} a^{l-1}_k + b^l_j\big) \tag{5.21}$$

where the sum is defined over the k neurons of the (l-1)-th layer. The previous equation expresses the relation between the output a^l_j of the j-th neuron of the l-th layer and the k-th output a^{l-1}_k of the previous layer. The previous equation can then be expressed as the activation function of the weighted input z^l_j as follows:

$$a^l_j = \sigma(z^l_j) \tag{5.22}$$

where the weighted input

$$z^l_j = \big(\sum_k w^l_{jk} a^{l-1}_k + b^l_j\big) \tag{5.23}$$

is indeed the input of the activation function a^l_j for the learning parameters w and b. Considering Eqs. (5.21)-(5.23), we rewrite the activation array a^l and the weighted input array z^l for the l-th layer in the following matrix form

$$a^l = \sigma(w^l a^{l-1} + b^l) \qquad z^l = w^l a^{l-1} + b^l \tag{5.24}$$

where w^l is the weight matrix of the l-th layer whose element in the j-th row and k-th column is the weight w^l_{jk} and b^l is the vector containing all the biases of the l-th layer.

Using the notation introduced so far, we discuss the back-propagation algorithm used to apply gradient descent and update the learning parameters at each iteration.

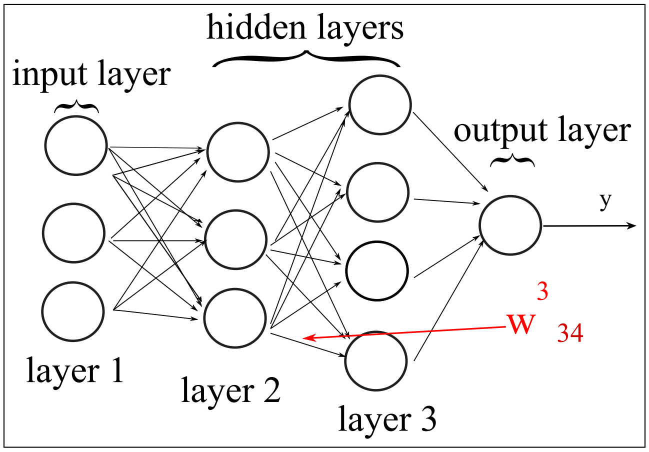

Figure 5.9: **Graphical representation of a network of multi-layer perceptron.** The first column containing the input neurons is called *input layer*. The two middle layers are called *hidden layers* and the rightmost layer containing the *output neurons* is called *output layer*. The notation w^l_{jk} denotes the weight connecting the k-th neuron in the $l-1$ layer with the j-th neuron in the layer l. In this case the weight w^3_{34} connects the third neuron of the second layer with the forth neuron of the third layer.

The purpose of the back-propagation algorithm is to calculate the derivatives $\partial J/\partial w$, $\partial J/\partial b$ of the cost function J respect to the learning parameters w and b. We then introduce the error δ^l_j in the j-th neuron state in the l-th layer as

$$\delta^l_j = \frac{\partial J}{\partial z^l_j} \tag{5.25}$$

where J is the quadratic cost function described in Eq. (5.5)

$$J = \frac{1}{2m}\sum_x \left(\|y(x) - a^L(x)\|^2\right)$$

where x and y are respectively the ensemble of the training inputs and the corresponding ground truth. Considering δ^l as the vector of the errors associated with the l-th layer, the back-propagation algorithm provides a method to compute the error δ^l for each layer and relate the error to the derivatives $\partial J/\partial w$ and $\partial J/\partial b$.

Considering a NN of L layers, the error δ^L over the last layer is:

$$\delta_j^L = \frac{\partial J}{\partial a_j^L}\sigma'(z_j^L) \tag{5.26}$$

Considering $\nabla_a J$ as the gradient vector whose component are the partial derivatives $\partial J/\partial a_J^l$, the previous formula can be rewritten as

$$\delta^L = \nabla_a J \odot \sigma'(z^L) \tag{5.27}$$

where the operator $\odot$ represents the elementwise product between two vectors. Considering the quadratic cost function of Eq. (5.5), we then obtain

$$\delta^L = (a^L - y) \odot \sigma'(z^L) \tag{5.28}$$

Considering Eq. (5.25), the error δ^l on the neurons of the l-th layer is then related to the error δ^{l+1} of the neurons in the (l+1)-th layer:

$$\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l) \tag{5.29}$$

where $(w^{l+1})^T$ is the transpose of the matrix w^{l+1}.

The previous formula expresses how the error on a given layer propagates backwards through the network to the error of the previous layer. Applying the formula backwards through the layers of the network, it is possible to compute the error on all the layers of the network from the error δ^L on the last layer combining Eq. (5.26) and Eq. (5.29).

Once we compute the errors δ^l for all the layers of the NN, the partial derivatives of the cost function J respect to the learning parameters are:

$$\frac{\partial J}{\partial b_j^l} = \delta_j^l \qquad\qquad \frac{\partial J}{\partial w_{jk}^l} = a_k^{l-1}\delta_j^l \tag{5.30}$$

The second equation of (5.30), can be rewritten as

$$\frac{\partial J}{\partial w} = a_{in} \times \delta_{out} \tag{5.31}$$

where the quantity a_{in} is the activation of the neuron input to the weight w and δ_{out} is the error on the neuron output from the weight w.
Considering the sigmoid function of Eq. (5.3) as the activation function σ, the term $\sigma'(z^l)$ in the Eqs. (5.27)-(5.29) is close to zero for very low or very high values of z^l, as happens for the error δ^l and the updating terms $\partial J/\partial w$, $\partial J/\partial b$. In this case, the neuron has saturated or stopped learning and any weight in input to a saturated neuron is learning slowly. Take into account Eq. (5.31), a similar consideration can be obtained for low activation a_{in} of the input neurons.
Since the learning process is slow when the input neuron has a low activation a_{in} or the output neuron has saturated, we can then design the activation functions of the NN accordingly to the desired learning properties.

The training process of NNs can then be summarized in the following steps:

1. **Weights and biases:** initialize the weights w and the biases b to random numbers according to a given distribution.

2. **Activation functions:** set the activation functions for each layer of the NN

3. **Feedforward:** Apply the feed-forward algorithm to find the weighted inputs $z^l = w^l a^{l-1} + b^l$ and the activations $a^l = \sigma(z^l)$ for each layer $l = 2, 3, ...L$

4. **Output of the error:** compute the error vector $\delta^L = \nabla_a J \odot \sigma'(z^L)$ on the last layer

5. **Back-propagation error:** Compute the error $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$ on each $l = L-1, L-2, .., 2$ layer

6. **Gradient of the cost function:** compute the gradient of the cost function $\frac{\partial J}{\partial b^l_j} = \delta^l_j$ and $\frac{\partial J}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j$ in order to update the learning parameters.

7. **Update the learning parameters:** Update the learning parameters according to the updating rules in Eq. (5.20).

8. **Iterations:** Repeat steps 3-7 until the cost function converges to its minimum.

Additionally, the data are usually feature-scaled and randomized in a pre-processing stage. Indeed, the observed data used during the training and the testing are composed by many features identifying a specific characteristic of the data. Since different features can cover a wide range of values at different scale, the observed data are usually normalized in pre-processing in order to be comparable and obtain an accurate prediction. Moreover the features-scaling reduces the amount of time required by the NN to minimize the cost function during the training process.

In order to prevent the over-fitting and obtain a general solution of the problem to be solved, the observed dataset is generally split in three subsets of training, validation and testing data of respectively the 60%, 20% and 20% of the overall available data. As discussed before, the training data are indeed used to update the learning parameter in the training process, the testing data are used to test the correctness of the solution provided by the NN and the validation data are used to evaluate the hyper-parameters such as the number of epochs, the learning rate and the best network architecture.

We now describe how NNs and data-driven approaches have been used in science and in particular in physics, drastically changing the interpretation of the observed data and the scientific research approach.

5.3 Data-driven approach in physics

NNs and data-driven approaches are a current methodology of investigation in most research fields with applications in traffic monitoring, speech and face recognition [138–140]. Deep learning approaches can indeed be applied in computational biology, biomedicine and genomics to identify hidden structures in molecules and genes [141–144].

The application of convolutional and recurrent NNs has been widely demonstrated in 3D imaging microscopy and high-content screening technologies to improve the spatial resolution over a large field of view and depth [145–147]. Moreover the unsupervised learning approach can be used to identify clusters and extract information about the galaxies morphology analysing the data of the emission spectrum [148, 149].

Due to the increasing progress in big data analytics and computing hardware, NNs and data-driven approaches can be efficiently applied for solving complex problems in many fields of physics such as Bose-Einstein condensates [132], many bodies problems [130] and quantum state tomography [131].

Particle Physics benefits from the application of NNs to isolate the signal of interest from a wide range of signals generated by many different background particles. As happens for high-energy particles collisions, the classification of specific particles is particularly difficult due to the low signal to background ratio. ANN approaches and deep learning techniques are often applied to better discriminate the signal from background classes by more complex functions [150]. Additionally, machine learning and deep learning techniques have proven useful application in new rare particles decay [151–153], jet classification [154], fast simulation by generative adversarial networks [155] and tracking reconstruction algorithms [156]. Thanks to the considerable amount of data analysed by data-driven approaches, deep learning techniques have also been applied in gravitational waves studies for real-time detection [157], noise transient classification [158] and gravitational wave signal denoising [159]. For the purpose of this thesis, the primary focus here is the application of the NN approach in Optics and in particular in CI.

Due to the recent advances in mathematical optimization and computing hardware, the data-driven approach has been widely used in CI imaging applications for solving ill-posed problems or noise affected measurements [160].
Most of the imaging problems solved using an ANN approach usually rely on a target image to be retrieved, an illumination system, a collecting system such as single-photon detectors or conventional intensity recording cameras, and the retrieval algorithm based on a NN model.
Using single images or multiple video frames in inputs, spatio-temporal convolution and deep NNs improve the quality of images limited by the undersampling of the detection systems [161, 162].
As proposed in [163], data-driven approaches can be further used to solve phase retrieval and lensless imaging problem where the relation between the amplitude and the intensity image at the output of the optical medium is highly nonlinear. Additionally, machine learning and data-driven approaches have been demonstrated to be efficient even with extremely low level of photons [164] for ghost imaging of hand-written digits recognition [165].
Data-driven approaches can also be used for the investigation of NLOS scenes. Recent results demonstrated the identification and the 2D tracking of people hidden behind a corner combining a convolutional ANN with the return temporal profile [94]. However, the application of data-driven approaches of hidden scenes are currently impractical due to the finite temporal resolution of the single-photon, time-resolving detectors and to the limited amount of training data. Since ToF techniques require expensive single-photon, time resolving detectors and pulsed light sources, recent results proposed a convolutional NN approach for NLOS 3D localization and objects classification by a conventional 2D camera [166, 167]. In this case [167], an adaptive lighting algorithm identifying the optimal illumination scene patches has been applied in order to maximize the NLOS

return signal. Furthermore, NNs can locate, classify and reconstruct the hidden scene by using conventional cameras [161, 168].
Deep learning approach provides crucial benefits also in imaging through scattering media with applications in non-invasive imaging through human body and autonomous vehicles in foggy environment. The propagation of the light through highly scattering media such as fibres or diffusers is indeed characterized by a strong beam-medium interference effect. As a consequence, the light intensity at the exit plane of the medium is apparently random. Recent results demonstrated the application of a deep learning approach on time resolved measurements in imaging through complex media such as a diffuser [169] and opaque walls [170] [171]. Finally, recent results demonstrated the classification and retrieval of handwritten digits input images propagating through up to a 1 Km fiber from the intensity images of the output speckle patterns [133, 172, 173].

Thanks to the infinite model architectures and to the wide range of activation functions, machine learning algorithms can be used to model infinite systems. Moreover the wide range of available layers makes NNs suitable for simulating highly non linear or ill-posed problem where finding the transformation function mapping the input into the output is computationally demanding.
The problem to be addressed in this thesis is the 3D retrieval of the scene from the return temporal histogram acquired with a single-pixel, time-resolving detector. Due to the lack of information about the scene to be imaged, the removal of any spatial structure imprinted either in detection or in illumination requires solving a very ill-posed inverse problem. Employing a data-driven approach to provide the prior knowledge on the scene, we provide a statistical representation of the possible scenes to be imaged on the basis of which a machine learning algorithm can be trained.

In the next chapter we describe the main current technologies used to retrieve the 3D information of LOS scenes and we discuss their advantages and limitations. We then introduce a new concept of data-driven 3D imaging based on arrival-time measurements with single-pixel, time resolving detector.

Chapter 6

3D imaging via artificial neural networks with a single pixel detector: theory

In this chapter we introduce the concept of Intelligent Lidar (ILidar), an innovative 3D imaging paradigm that allows the 3D recovery of LOS scenes using only the arrival-time measurement of the return photons acquired with a single-pixel, time-resolving detector.
We introduce the main current technologies used to retrieve the 3D information of LOS scenes, discussing their advantages and limitations. We then explain how a supervised-learning data-driven approach can be used to retrieve the 3D image of a scene by arrival-time measurements. The imaging challenges to be addressed by the proposed approach are the compactness, the fastness and the amount of measurements and data required to obtain the 3D image of a scene.

6.1 Introduction to 3D imaging of LOS scenes

The ability to reproduce 3D images of direct scenes has always been a crucial task since 1500's when Leonardo Da Vinci realized the first example of a 3D scene. Indeed recent results [174] demonstrated that the *Gioconda* and one of its copies simultaneously realized, differ slightly in perspective. Since the difference between both paintings simulates the human binocular perspective vision, the two *Mona Lisa* together represent the first stereoscopic image in the world history. In recent years, the constant progresses in computing, sensing and technology [175] allowed the

development of 3D imaging technologies with a variety of applications in medical and biomedical imaging [176–178], microscopy [144], autonomous vehicles, ranging [179–181] and facial recognition [182, 183].
The most familiar example of 3D imaging is the visual perception of the human eyes where the 3D information is inferred combining the two different perspectives simultaneously seen by our eyes set slightly apart in space. The brain can indeed combine the two 2D images seen by each retina from two different lines of sight and extrapolate a 3D image of the scene.

In the last decades, various types of 3D imaging technologies have been developed, addressing different needs and requirements. Most of the promising approaches for the 3D retrieval of direct LOS scenes are usually based on stereo-vision (SV), holography or ToF techniques.
As the eyes of humans and many predators, the stereo-vision technique infers the 3D information of the scene combining the images collected from two slightly different prospective by two sensors spatially separated [184, 185].
In conventional 3D optical holography, the 3D information of the scene is retrieved from the interference pattern produced by a reference beam and a probe beam scattered back from the object [175, 186, 187]. In order to capture the interference pattern on the recording medium, conventional holography usually uses photographic media [188, 189]. Digital holography instead, employs digital cameras such as charge-coupled devices (CCD) or Complementary Metal Oxide Semiconductor (CMOS) sensors [190, 191]. Used in a wide range of applications [105, 192], ToF approaches retrieve the depth information by measuring the arrival-time of the return signal. The ToF information of the return beam can be measured using either a frequency modulated continuous wave (FMCW) light beam or a pulsed light source [193]. In more details, the sensors based on FMCW, infer the depth information from the difference in phase between the emitted and the received signal [75, 194]. The pulsed light ToF approach infers the depth information from the difference between the return arrival-time and a reference signal [69, 195].

ToF approaches retrieve the 3D information by scanning the scene or by structured illumination and/or detection. In the scanning technology a laser spot scans the scene, while a time-resolving single-photon sensitive, single-pixel detector confocally collects the return photons. As happens for LiDAR and Velodyne LiDAR systems [55], the arrival-time and the spatial information provided by the scanning, are then combined to provide the 3D information. Scanning systems require as many consecutive measurements as the transverse spatial resolution of the retrieval. Moreover this techniques requires the use of moving parts such as galvo-mirrors or rotat-

ing systems [196]. Imaging algorithms such as compressive sensing allow to reduce the number of measurements required to obtain an image, without affecting the image reconstruction quality. Recent results demonstrated the recovery of a continuous real-time 3D image with a frame-rate up to 12 Hz [69, 197]. Instead, structured detection technology flash-illuminates the whole scene and acquires the return signal by a spatial resolved (or multi-pixel) detector simultaneously collecting the return signal from different points [196]. Pixelated detection requires as many detectors as the spatial resolution of the image and multiple simultaneous acquisitions.
The current 3D imaging technologies aim to improve the depth resolution of the retrieval and the range distance [198, 199]. Indeed recent results demonstrated an hybrid approach for 3D scenes reconstruction with velocity detection and centimetres depth resolution at 1 Hz of frame rate combining multi-view cameras and scanless Lidar [181]. Combining ghost imaging via sparsity constraint (GISC) technique with time-resolved measurements, this technique can be further improved enabling the 3D recovery of a scene up to 1.0 km range [200].

In order to retrieve the spatial information of the transverse plane, current ToF technologies require the use of spatially resolving (many pixels) sensors or structured illumination, providing prior knowledge about the illuminated transverse plane portion of the scene at each acquisition. Indeed, standard 2D imaging techniques recover 3D images projecting structured patterns onto the scene and detecting only the total intensity of the back-scattered light, for which a single pixel is sufficient, as demonstrated with single-pixel cameras [7, 201]. The transverse information is retrieved summing over all the illumination patterns weighted by the corresponding measured intensity. After obtaining the 2D image reconstruction, each 2D pixel is assigned a the temporal histogram of the return signal collected at the corresponding transverse portion of the FoV. The depth information is then inferred by the ToF information of the temporal histograms, providing a 3D retrieval of the investigated scene.
Exploiting the prior information provided by the structured illumination, the recovery of a 3D scene is then a well-posed problem and widely demonstrated in literature [57, 69]. In contrast, the 3D recovery of a LOS scene using a time-resolving single-pixel detector with no prior knowledge about the scene is a ill-posed problem and the current 3D imaging systems are limited by the many pixels requirement.
The necessity of having a spatially resolving sensor is well understood considering the temporal histogram of the return acquired by single-pixel detector that collects the light from the entire FoV. Indeed, all the possible locations (x,y,z) of the contributions at a given time bin t lay on the surface of the sphere defined by the equation $ct = \sqrt{x^2+y^2+z^2}$ where c in the speed of light. Removing

the prior knowledge about the scene (in this case the coordinates (x,y)), all the spatial points laying of the sphere contribute with equal probability to the same time bin of the histogram and it is not possible to univocally retrieve the 3D image of the scene.

Here, we introduce a new paradigm for 3D imaging of direct scenes using a time-resolving single-pixel detector applying a data-driven approach to recover the spatial information lost using a single-pixel. In more details, we demonstrate a 3D imaging approach called intelligent-Lidar (ILidar), in which a pulsed laser source flood illuminates the whole scene and a time-resolving, single-pixel detector records the temporal histogram of the return from the whole scene. The 3D image of the scene is then recovered from the acquired temporal histogram using a data-driven algorithm based on deep learning approach. Once the NN is trained, the discussed method provides a scanless 3D imaging approach from the return signal of the entire 3D scene.

The proposed approach represents a new concept of 3D imaging obtained from a temporal trace so far unexplored. The proposed method demonstrates the possibility of achieving a scanless and compact single-pixel Lidar system for real-time 3D imaging and pattern recognition from a single temporal trace.

Since the suggested approach needs the acquisition of a single temporal trace, the system requires a single measurement as opposed to many consecutive measurements of any other scanning system. It provides a 3D image of the scene at potentially kHz frame rates limited only by the acquisition time of the temporal histogram. The proposed method can additionally operate in a cross-modality mode, as demonstrated testing the system on a different platform such as a radio-frequency antenna.

In this chapter we describe the current optical 3D imaging technologies for LOS scenes and in more details, the systems based on ToF information. Considering the requirement of pixelated sensors of the current ToF systems, we then introduce the concept of ILidar. We finally describe the NN and the algorithm used to retrieve the 3D image of the investigated scene from the temporal trace of the return collected with the single-pixel sensor.

6.1.1 State of the art

In the last decades various technological principles have been exploited to retrieve the 3D image of direct LOS scenes. The most promising approaches are based on stereo-vision, holography or ToF techniques, providing a better depth resolution when the three approaches are combined [202].

As demonstrated for medical imaging [203], underwater 3D reconstruction [67] and urban traffic [185], the stereo-vision 3D imaging is based on combining two 2D images seen from a different

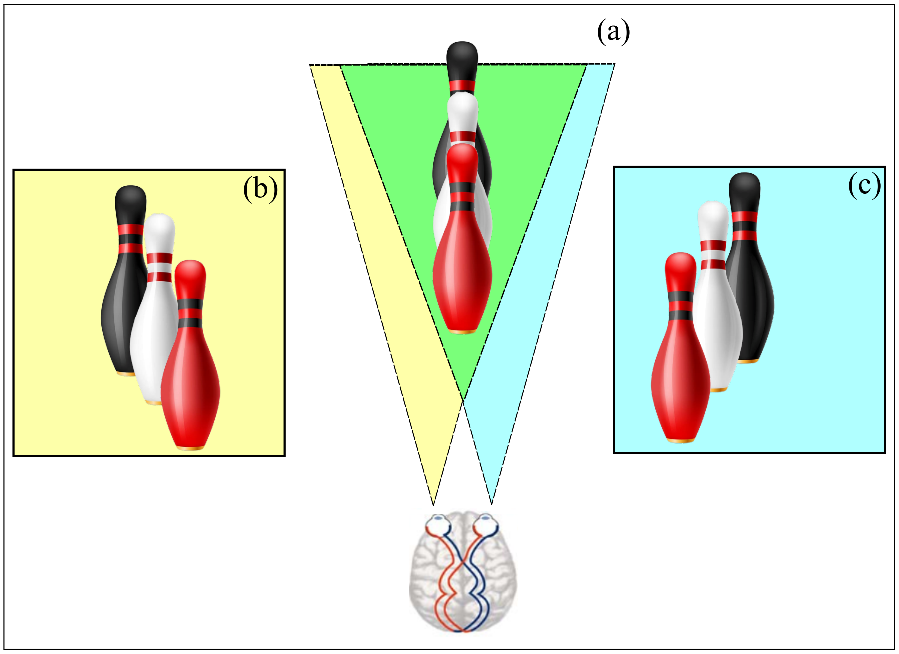

Figure 6.1: **Schematic of stereovision approach for 3D imaging.** a) Since the human eyes are spatially separated, the brain extrapolates the 3D information of the scene combining the 2D images simultaneously seen from two different prospectives by the left and the right eye (b-c) by a triangulation process.

prospective by two sensors spatially separated. In more details, this approach relies on computing the disparity map $D(x_r, x_l)$ given by the difference of the image coordinates between two corresponding pixels x_r and x_l of the left and of the right image [204]. Figure 6.1 shows the holographic approach.

Widely demonstrated in 3D microscopy [205], 3D scenes visualization [175] and retrieval of large scale objects [206], 3D holography is based on recovering the 3D image of the scene by the interference pattern of a reference and a probe beam scattered from the object [186, 187]. In this case the 3D information of the scene is encoded in the phase of the intensity hologram captured using a Charged Coupled Device (CCD) or a Complementary Metal Oxide Semiconductor (CMOS) sensor. Figure 6.2 shows the holographic approach .

Widely investigated in radar imaging systems [207], recent results demonstrated the 3D recovery of LOS and NLOS scenes exploiting the ToF information of the return. Known as LiDAR, this approach combines the temporal information provided by the ToF of the return and the spatial information provided either using a pixelated sensor or scanning the scene. Each pixel of the

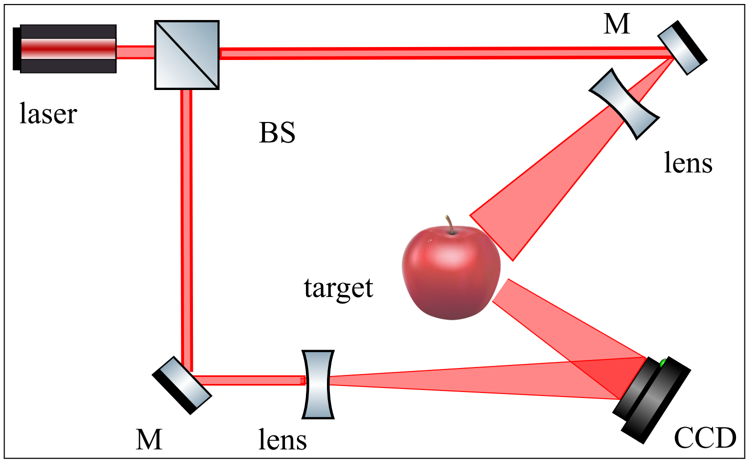

Figure 6.2: **Schematic of holography approach for 3D imaging.** A beam splitter (BS) divides the light beam in two separated beams, known as the object beam and the reference beam. The two beams are then expanded and the object beam illuminates the object to be retrieved. The light scattered by the object travels towards a recording medium where the reference and the object beam are recombined in an interference pattern. In digital holography, the hologram produced by the interference of the two beams is then recorded and stored by a CCD or CMOS sensor rather than a photographic film, as happens for classical holography.

sensor collects the return signal scattered back from a specific (x-y) portion of the transverse plan, encoding the (x-y) information of the scene. At the same time, the depth information is inferred by the time the light takes to go from the laser source to the target and then back to the sensor. A single-pixel is instead used in the scanning approach. Here, a laser spot scans the investigated scene and a single-pixel detector confocally collects the return signal.
The current 3D imaging technologies based on the ToF information rely on either a pulsed or continuous wave (CW) modulated illumination. As reported in Fig. 6.3, the pulsed light approach recovers the ToF information measuring the temporal difference between the arrival-time of the return signal and the arrival-time of a reference. In the CW modulation approach, the ToF information is instead recovered measuring the phase difference between the emitted and the received signal, as discussed in Chapter 3. The current CW modulation ToF cameras technologies available on the market can reach a rate of 120 frames per second with a centimetre resolution and a distance range between 0.5 and 10 metres [208].

Due to the recent advances in single photon detection and TCSPC techniques [2], the pulsed modulation ToF approach is one of the most common method used to retrieve the full 3D image of LOS and NLOS scenes, providing a better depth resolution when combined with the other suggested approaches [202]. This approach can be further improved exploiting the advantage of

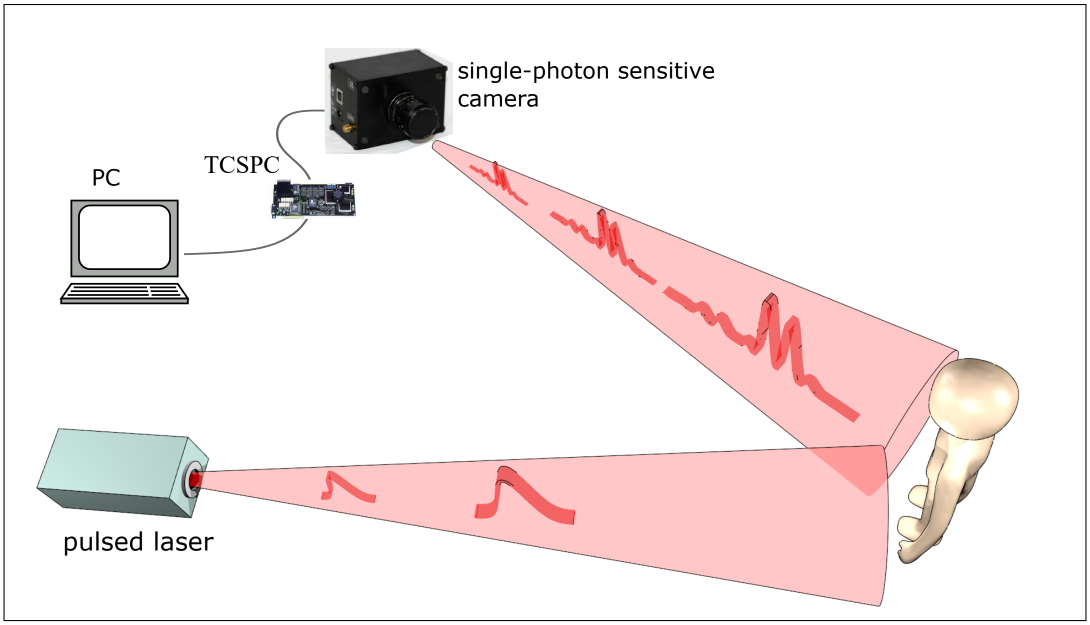

Figure 6.3: **Schematic of ToF approach for optical 3D imaging by pulsed light modulation.** In order to recover the 3D image of a scene, a pulsed light source flood illuminates the scene to be recovered. A single-photon sensitive camera operating in time-correlated single-photon counting (TCSPC) mode collects the return signal scattered back from the objects. The depth of the scene is then inferred by the ToF of the return.

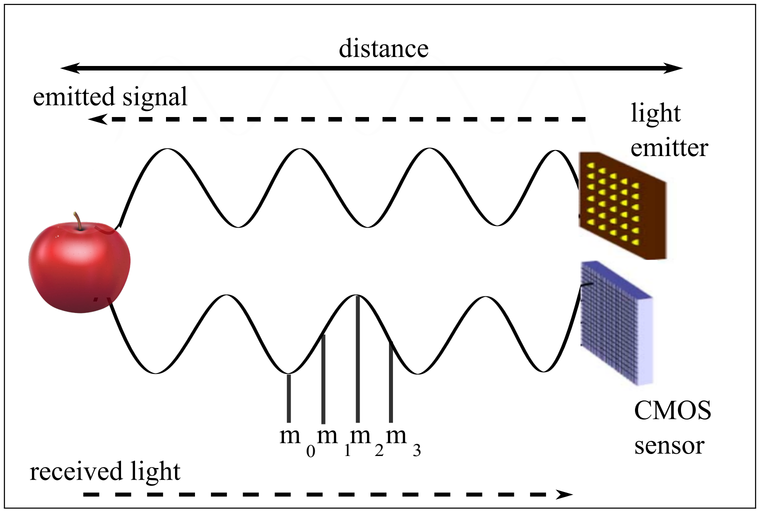

Figure 6.4: **Schematic of ToF approach for optical 3D imaging by CW modulation.** In order to recover the 3D image of a scene, an infra-red sinusoidal modulated beam is flood illuminating the scene to be recovered. A pixelated CMOS sensor collects the return signal. The phase difference is then obtained measuring the amount of light reflected at four points m_0, m_1, m_2, m_3 equally disposed within the modulation period.

using a single-pixel camera [57,69]. The current 3D imaging technologies can be further improved applying deep learning techniques, as demonstrated in 3D microscopy [144], super-resolved fluorescence microscopy [209] and lensless imaging [163].

Despite the recent advances in computational imaging algorithms and single-photon sensitive detectors, the ToF approach requires either many consecutive measurements as it happens for scanning systems, or the use of a pixelated sensor in order to recover the spatial information on the transverse plane, providing prior knowledge about the scene to be recovered. Therefore, the current 3D imaging sensors are usually made by many-pixels or require a scanning imaging system not compatible with compact and portable systems for remote sensing applications. Moreover, the 3D retrieval of a scene requires as many consecutive measurements as the transverse spatial resolution of the image, leading to lower data acquisition and data transfer rate. We describe the ToF detection methods that can be used to retrieve the 3D image of scenes by pulsed illumination, focusing the attention on pixelated detector systems.

6.2 Time-of-flight imaging detection techniques

Current ToF technologies require the use of pixelated detectors to retrieve the spatial information of the transverse plane of the investigated scene. Commonly referred to as cameras or detector array, this technique requires as many single-photon sensitive detectors (or pixels) as the transverse spatial resolution of the image. Indeed, each pixel of the sensor collects the light from a specific portion of the transverse plane of the scene, encoding the (x-y) spatial information of the scene to be imaged, as reported in Fig. 6.5. The depth is then inferred by the ToF information of the return signal collected by each pixel. The 3D image of the scene can then be retrieved combining the 2D information encoded by each pixel and the ToF information.

The spatially resolved single-photon sensor approach can be further improved using a single-pixel camera, representing a cheaper alternative to the spatially resolved single-photon sensor approach. In order to retrieve the 3D image of the investigated scene, a single-pixel camera technology combines a time-resolving, single-pixel detector and a DMD. Applying a series of spatially resolved illumination patterns, the transverse spatial information of the scene is inferred by the DMD, while the corresponding return signal is acquired with a single-photon sensitive, time-resolving detector with no spatial resolution. As discussed in Chapter 2, the transverse spatial information can be retrieved using the DMD either in projection or in detection. Figure 6.6 shows the 3D imaging single-pixel camera approach.

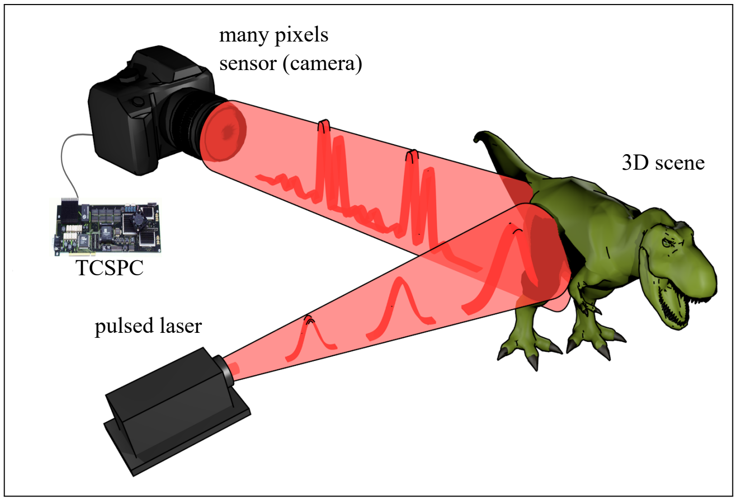

Figure 6.5: **Schematic of pulsed ToF approach for optical 3D imaging by a many pixels sensor.** In order to recover the 3D image of a scene, a pulsed light beam flood illuminates the scene. A single-photon sensitive detector array (cameras) operating in time-correlating single-photon counting mode, collects the return signal. Each pixel of the sensor encodes the transverse spatial information of the scene, meanwhile the depth is inferred by the ToF information of the temporal histogram of the return signal collected by each pixel.

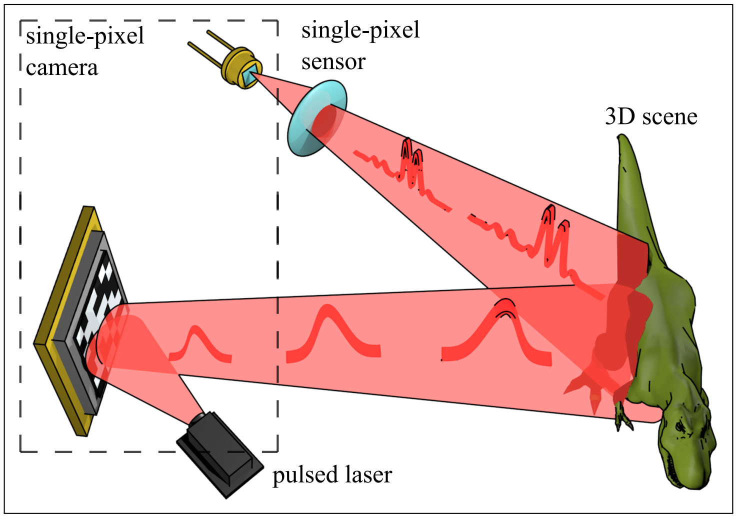

Figure 6.6: **Pulsed ToF approach for optical 3D imaging by a single-pixel camera.** In order to retrieve the 3D image of the investigated scene, the single-pixel camera combines a SLM and a time-resolving single-pixel detector, as opposed to the pixelated sensor approach. Applying structured patterns, the DMD encodes the spatial information of the transverse plane, while the time resolved single-pixel detector collects the back-scattered signal. The 3D scene is then retrieved combining the 2D information provided by the spatially resolved patterns with the corresponding temporal histogram of the return signal.

Since the 3D imaging retrieval of scenes relies on the temporal and on the spatial information, the use of pixelated sensors is a necessary requirement. Indeed the recovery of the 3D image of a scene using a single-pixel, time-resolving detector is a highly unconstrained inverse problem, as discussed in the next section.

6.3 Pixelated sensor requirement

Here, we report the inverse light transport model used to retrieve the 3D image of a LOS scene from the temporal histogram collected with the single-pixel time-resolving detector.

With reference to Fig. 6.7, we consider a typical 3D imaging scenario where a pulsed laser source flash-illuminates the 3D scene and a single-photon sensitive, single-pixel detector collects the return photons from the whole scene in a temporal histogram form. According to the forward imaging algorithm discussed in Chapter 4 for NLOS scenes [105, 107], the target patch i located at coordinates (x_i, y_i, z_i) contributes to return signal collected by the single-point detector at a time t_i. Considering the detector position as the origin $O(x = 0; y = 0; z = 0)$ of our frame of reference, the time t_i is described by the well-known ToF formula

$$t_i = \frac{r_{laser-obj}(x_i, y_i, z_i) + r_{obj-det}(x_i, y_i, z_i, x_\ell, y_\ell, z_\ell)}{c} \tag{6.1}$$

Here, c is the speed of light, $r_{laser-obj} = \sqrt{(x_i - x_l)^2 + (y_i - y_l)^2 + (z_i - z_l)^2}$ is the distance between the laser and the target patch i and $r_{obj-det} = \sqrt{x_i^2 + y_i^2 + z_i^2}$ is the distance between the target patch and the detector.

With reference to the terminology used in Chapter 4.3, the recovery of the ToF information and of the temporal histogram $b(t)$ for a given scene $s(x_i, y_i, z_i)$ is a well-posed problem with a unique temporal solution encoding information about the scene. Indeed, the temporal histogram $b(t)$ of the back-scattering of a scene with spatial distribution $s(x_i, y_i, z_i)$ is described as:

$$b = F(s) \tag{6.2}$$

where F is the light transport function.

We then solve the inverse light transport model in order to retrieve the 3D information of the scene from the temporal histogram of the return photons scattered back by the scene. We neglect the laser position (x_ℓ, y_ℓ, z_ℓ) for simplicity.

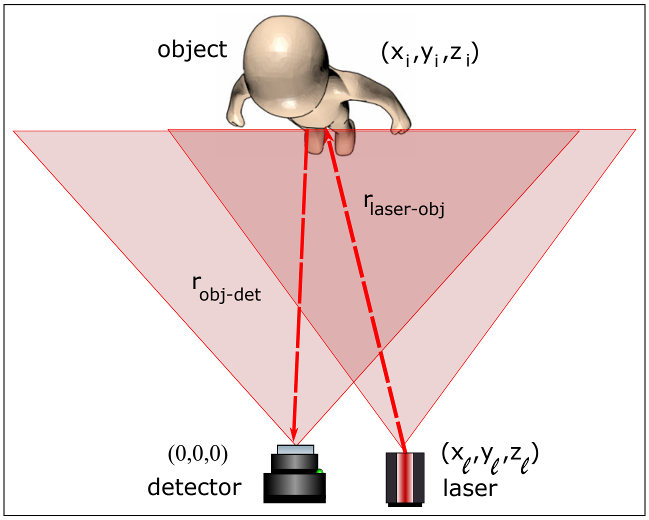

Figure 6.7: **3D imaging by ToF approach with time-resolving sensors.** A pulsed light source flash-illuminates the scene and a time-resolving detector collects the return photons. Applying the standard ToF formula in Eq. (6.1), the target located at coordinates (x_i, y_i, z_i) contributes to the number of photons collected at a unique solution time t_i. On the contrary, the solution of the inverse light transport model with no prior knowledge about the scene is a highly unconstrained and remarkably ill-posed problem.

Since the temporal histogram recorded by the time-resolving detector contains information about the 3D scene, it is possible to infer information about the 3D scene from the temporal histogram of the return photons. Indeed, the 3D image of the investigated scene can be uniquely recovered correlating the ToF information and the (x_i, y_i) spatial information provided either by the scanning or by the pixelated sensor. The depth information is then inferred by the temporal histogram recorded with the time-resolving detector as follows:

$$z_i = \frac{t_i c}{2} \tag{6.3}$$

where t_i is the arrival-time of the contribution to the return signal produced by a target located at coordinates (x_i, y_i, z_i).

However, the 3D recovery of scenes using only the temporal histogram of the return photons collected with a single pixel, time-resolving detector is a highly unconstrained and remarkably ill-posed problem. If no prior knowledge about the transverse spatial information (x_i, y_i) of the scene is available, there are multiple solutions (x_i, y_i, z_i) verifying Eq. (6.1) and the 3D retrieval is not unique. In this case, all the scene patches (x_i, y_i, z_i) located along the surface of the sphere described by the formula

$$t_i = \frac{\sqrt{x_i^2 + y_i^2 + z_i^2}}{c}, \tag{6.4}$$

contribute with equal probability to the same time bin t_i of the recorded histogram. Therefore, it is not possible to univocally retrieve the 3D information of the investigated scene. Considering then the return photons collected at a time t_i by the single-pixel sensor, it is possible to retrieve only the surface of the corresponding sphere and not the unique solution of the actual scene. The solution of the inverse problem is therefore remarkably ill-posed and all the patches laying on the surface of the sphere contribute with equal probability to the return signal collected with the single-pixel detector.

Based on the considerations discussed so far, it is not possible to univocally retrieve the 3D image of a scene using only the temporal information provided by the temporal histogram of the return photons detected with a time-resolving, single-pixel detector. If no prior information about the scene is provided, the lack of constraints on the scene usually provided by a structure illumination, makes the inverse light-transport model highly ill-posed and multiple equally distributed solutions are retrieved. Here, we introduce a new 3D imaging approach for direct LOS scenes based on single-pixel, time-resolving detector providing the necessary scene constraints by a data-driven approach.

6.4 Intelligent Lidar approach

As demonstrated in Section 6.3, the temporal histogram of the return signal acquired with a time-resolving detector, contains information about the 3D scene. However, the temporal histogram information is not sufficient to univocally describe the inverse light transport model. Prior knowledge on the transverse spatial information has to be provided in order to recover the 3D image of the scene by a single-pixel, time-resolving detector.

Here, we introduce a 3D imaging paradigm for LOS scenes by single-pixel, time resolving detector, providing the required prior knowledge about the scene by a data-driven approach. Referred to as ILidar, the suggested approach represents a scanless and compact single-pixel LIDAR system for real-time 3D imaging and pattern recognition using only the temporal histogram of the return from the whole scene. The ILidar 3D imaging modality is depicted in Fig. 6.8. Here (Fig. (a)), a pulsed light source flash-illuminates the scene and a single-pixel, time-resolving detector collects the return from the whole scene in a temporal histogram.

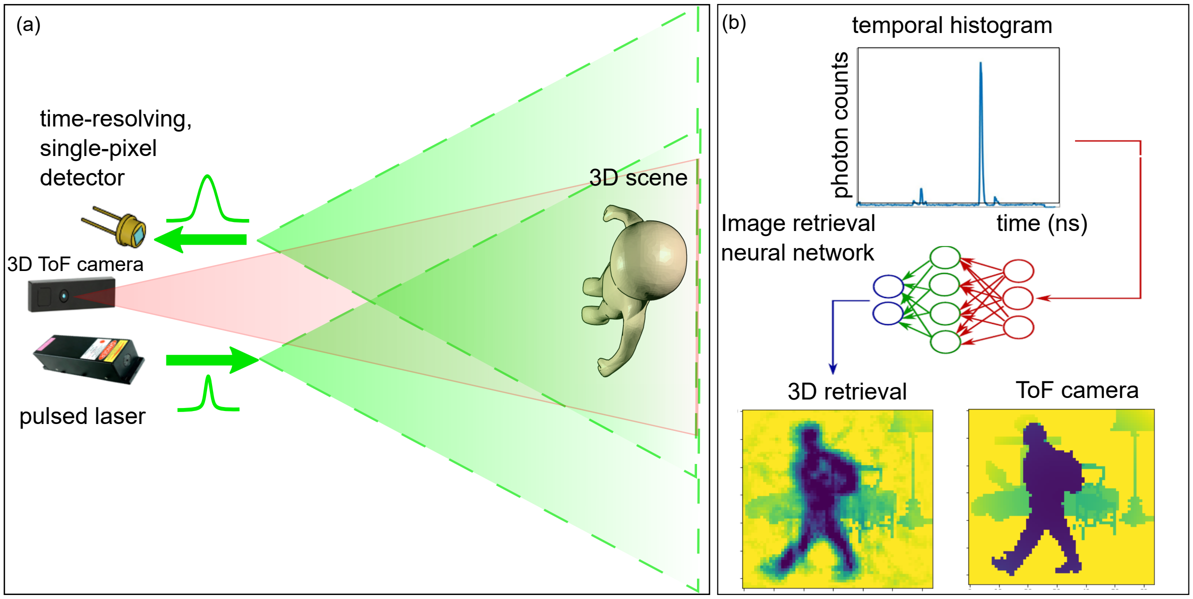

Figure 6.8: **Intelligent Lidar approach: 3D imaging by single-pixel, time resolving sensor.** (a) A pulsed laser source flash-illuminates the scene and a time-resolving, single pixel detector collects the return photons from the whole scene, while a standard 3D ToF camera records intensity-encoded depth 2D images of the scene. (b) Data-driven approach used to recover the 3D image of the investigated scene by arrival-time measurements. The temporal histograms are the input of the NN, while the intensity-encoded depth 2D images are the output of the NN. Training the ANN on pairs of temporal histogram and corresponding 3D image, we find the transformation F^{-1} mapping the temporal histogram into the corresponding 3D image by a supervised artificial neural-network algorithm of fully-connected layers. After the algorithm is trained, it is fed with single-point temporal histograms to retrieve 3D images in a single recording.

In order to retrieve the 3D image of the scene from the corresponding return temporal histogram, we provide a statistical representation of all the possible scenes to be imaged on the basis of which a machine learning algorithm can be trained. The aim of the NN is to find the inverse transformation F^{-1} (see Eq. (6.2)) mapping the temporal histogram b of the return photons into the spatial distribution $s(\vec{r})$ of the targets by a supervised training approach. The image recovery neural-network is depicted in Fig. 6.8 (b). In order to train the NN, we include a standard 3D ToF camera that provides intensity-encoded depth 2D images of the scene. For each acquisition we then acquire the temporal histogram of the return photons provided by the single-pixel, time-resolving detector and the corresponding 3D image of the scene provided by the ToF camera.

Using a supervised-learning approach, the ANN is then fed using temporal histogram-3D image pairs in order to find the mapping function F^{-1}. The temporal histogram acquired by the SPAD is the input layer of the ANN, while the intensity-encoded depth 2D image acquired by the ToF 3D camera is the output layer. Section 6.5 discusses the NN algorithm.

After the algorithm is trained, the proposed approach is able to retrieve the 3D information with just the arrival-time of return and the 3D image recovery algorithm. We then test the discussed approach with series of four different 3D scenes composed by objects freely moving in a room, as described in Section 7.1. However, the proposed approach is affected by some limitations such as the degeneracy problem due to the spatial symmetry of the experimental conditions. Indeed, 3D scenes that are ToF symmetrical produce the same temporal histogram, generating an ambiguity in the 3D image predicted by the NN.

6.4.1 The degeneracy problem

The proposed 3D imaging paradigm is affected by some limitations due to the spatial symmetry of the problem to be addressed. We consider the spatial symmetry of the experimental conditions depicted in Fig. 6.9 (a) where a macroscopic target is freely moving within an uniform background scene. In these experimental conditions, multiple 3D scenes that are ToF symmetrical produce the same temporal histogram, generating an ambiguity in the 3D imaging recovery predicted by the NN.

With reference to Fig. 6.9 and neglecting the laser-targets distance, all the point targets placed on the surface of the sphere $\{x_i, y_i, z_i\}$ verifying the equation

$$t_i = c^{-1}\sqrt{x_i^2 + y_i^2 + z_i^2}, \tag{6.5}$$

contribute to the same temporal return signal acquired by the single-pixel sensor at the time bin t_i. Since no spatial structure is imprinted at any stage, multiple 3D reconstructions are therefore compatible with the same temporal histogram, as reported in Fig. 6.9 (b-e). Here, a target placed in the left side of the FoV and a target placed in the right side of the FoV at the same ToF distance (yellow crosses in Fig. 6.9 (a)), produce exactly the same return temporal histogram. Since the same temporal histogram can be attributed to multiple ToF symmetrical 3D images, the proposed approach is affected by a spatial degeneracy. The spatial degeneracy produces a spatial ambiguities in the corresponding 3D image retrieval. Experimental evidence of the spatial degeneracy of ToF symmetric 3D images is shown in Section 7.5 where two symmetrical 3D images are retrieved from a single-point temporal histogram. The spatial degeneracy induced by the spatial symmetry of the problem is totally removed when information about the spatial structure of the scene is considered in the retrieval process, as it happens for multi-pixel detectors or for single-pixel cameras.

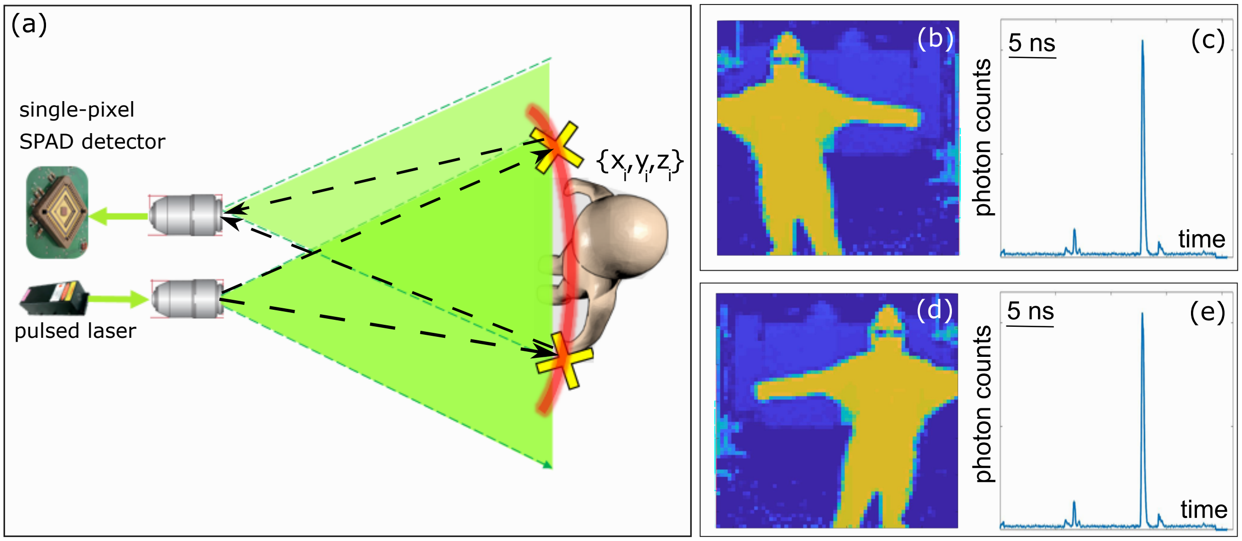

Figure 6.9: **Spatial degeneracy of ToF symmetric 3D scenes.** (a) A pulsed laser source flood illuminates an uniform background scene while a single-pixel, time-resolving detector collects the return scattered back by the entire scene. Due to the spatial symmetry of the problem, all the targets located at the surface of the ellipsoid (or a sphere when neglecting the laser-targets distance) defined by the ToF Eq. (6.5) contribute to the same temporal histogram. Therefore, two isolated targets placed either in the left (b) or in the right (d) side of the FoV have the same return signal (c,e), inducing a spatial ambiguity in retrieval. Here, the red line represents the ellipsoid of all the possible 3D scenes having the same return photons signal. The yellow crosses indicate two ToF symmetric 3D locations of the target.

We now describe the NN algorithm used to retrieve the 3D information of the scene from a single single-point temporal histogram. The aim of the NN is to retrieve the transformation mapping the temporal histogram of the return photons into the corresponding 3D image formerly training the NN with experimental data of 3D image-temporal histogram pairs from the scene.

6.5 3D image retrieval neural network algorithm

In order to infer the inverse light transport function F^{-1}, we use a fully-connected multi-layer NN with feed-forward training approach. Figure 6.10 shows the ANN architecture. The ANN is composed by five sequential layers implemented in TensorFlow [210].

With reference to the terminology introduced in Chapter 5, the mapping transformation F^{-1} can be expressed by the learning parameters **w** and **b** as follows:

$$\mathbf{y} = f(\mathbf{w}\mathbf{x} + \mathbf{b}) \tag{6.6}$$

Here, the input **x** is the temporal histogram, the output **y** is the corresponding 3D image of the scene and f is a generic non-linear function.

In the experimental conditions reported in Chapter 7, the temporal histograms are composed by 1800 time bins, whereas the colour-encoded depth 2D images are 2D matrices of 64×64 = 4096 elements. The 64x64 pixels images are reshaped into a 1D array of 4096 entries in order to train the NN. The input layer is then composed by 1800 neurons corresponding to the dimension of the input vector x, while the output layer has 4096 neurons corresponding to the dimension of the output vector y. The hidden layers are composed by three dense layers of 1024, 512 and 256 neurons for a total of 3553024 trainable parameters, as shown in Fig. 6.10.

The hyperbolic tangent shown in Fig. 5.3 has been used as the activation function of the perceptrons for each layer. The cost function minimized during the learning process is the mean square, defined as

$$J(\mathbf{w},\mathbf{b}) = \frac{1}{m}\sum_{i=1}^{i=m} \|y(x_i) - f(x_i,\mathbf{w},\mathbf{b})\|^2 \tag{6.7}$$

where $f(x_i,\mathbf{w},\mathbf{b})$ is the predicted output array of the i-th input examples x_i and $y(x_i)$ is the corresponding ground truth. Here, the index i goes from 1 to the total number of the training data m.

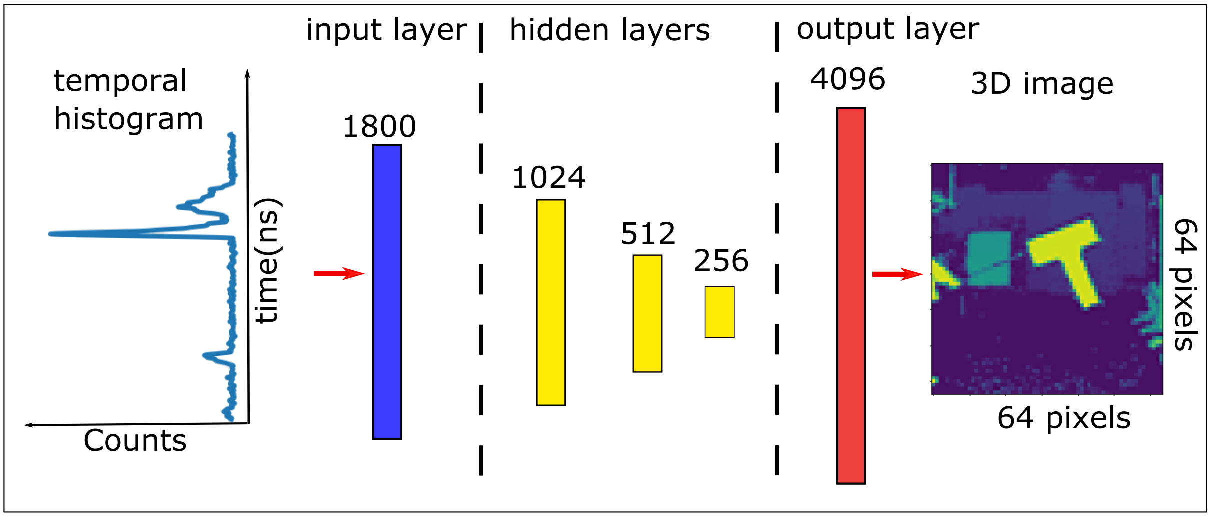

Figure 6.10: **Graphic representation of the ANN.** The ANN is composed by five sequential layers in total. The input layer is made by 1800 neurons corresponding to the experimental temporal histogram dimension. The output layer is composed by 64x46=4096 neurons corresponding to the experimental 3D images dimension. The hidden layers are composed by three dense layers of 1024, 512 and 256 neurons respectively. Using a dense layer, each input neuron is connected to each output neuron by a linear combination of real-values weights **w**, biases **b** and a non-linear activation function f.

During the learning process, the weights are iteratively updated using the ADAM optimization algorithm. Finally, the initial random weights are tensors initialized to zero. Further details about the NN model are provided in Appendix D. Once the system has been trained, we obtain the set of learning parameters **w** and **b** defining the transformation F^{-1} and a 3D retrieval of the investigated scene can be obtained from the return photons temporal histogram.

The concept of Intelligent Lidar represents an innovative 3D imaging paradigm able to retrieve the 3D information of the scene using only the time-arrival information. The suggested approach represents a compact and fast 3D imaging approach leading to a new concept of image made by a temporal profile. Since no scanning system or multiple measurements are required after the training, the complete 3D image can indeed be retrieved by the acquisition of a single temporal histogram on a standard laptop, reducing the memory and the time required for the data storage, handling and transfer with considerable profits for remote sensing applications. Additionally, the proposed 3D imaging paradigm can be extended and applied to completely different platforms, provided that an optical system based training is formerly performed to provide the statistical representation of the possible scenes to be imaged.

Now that this 3D imaging approach has been explained, we test it under experimental conditions. The next chapter (Chapt. 7) will experimentally demonstrate how the proposed paradigm can be used to recover the 3D spatial information from temporal data.

Chapter 7

3D imaging via artificial neural network with a single-pixel detector: experimental results

We now test the ILidar approach in experimental conditions recovering 3D images of various dynamic scenes from the temporal histogram of the return photons by the suggested data-driven approach. The proposed approach represents a new paradigm to recover the 3D image of a scene from arrival-time measurements, paving the way to an innovative, fast and cross-modality 3D imaging technique. In this Chapter we report the experimental setup and the results obtained testing the proposed approach on four different scenarios composed by targets of different shapes and dimensions freely moving within a room.

We test the proposed approach in two distinct cases. In the first case, we test the proposed approach training the NN with 3D image-temporal histogram pairs of the four investigated scenarios separately. In the second case, we train the NN jointly using input-output data of the four investigated scenarios. In both cases, the proposed approach is able to accurately retrieve the 3D image of the investigated scene without requiring any spatial structure either in illumination or in detection.

However, the limited temporal resolution of the experimental detection system affects the performance of the proposed 3D imaging paradigm. Since the proposed approach recovers the 3D image by arrival-time measurements, the algorithm struggles to correctly identify the details of object whose size is comparable with the transverse spatial resolution induced by the temporal resolution of the detection system. This is more evident in a multi-scene training scenario where diversified scene objects are included in the same learning process, as discussed in Section 7.3.2. Additionally, the spatial degeneracy of ToF symmetric 3D scenes affects the retrieval and limits the accuracy and the performance of the suggested method, as discussed in Section 7.5.

Since the 3D image of a scene is recovered by arrival-time measures, the suggested 3D imaging paradigm can be employed in a completely different platform. The proposed approach can indeed recover the 3D information of LOS scenes using a Radio Detection And Ranging (RADAR) transmitter-receiver chip, provided that the system has been previously trained by a 3D imaging optical system.

7.1 Experimental setup

We now experimentally test the suggested 3D imaging paradigm using the experimental setup in Fig. 7.1 (a). Here, a pulsed light source flood-illuminates the whole 2.15x2.15x4 m^3 scene and a single-pixel, time-resolving SPAD detector operating in TCSPC mode collects the return photons from the entire scene. We use a pulsed supercontinuum white-light laser ((SuperK EXTREME/FIANIUM, NKT Photonics) producing $\sim$ 75 ps pulses of 13 nJ with a repetition rate of 19.5 MHz and an average power of 250 mW. Since the laser emits in the all visible spectrum and the sensor has an improved PDE of 70% at 550 nm wavelength , we select the wavelength by a band-pass spectral filter (omitted in the figure) centred at 550 nm with 40 nm bandwidth. The laser source passes through a 10x microscope objective (Olympus plan achromatic, omitted in the figure) with 0.25 numerical aperture, 18 mm focal length and 10.6 mm working distance to flash illuminate the whole scene. The objective has an opening angle of $\approx 30^\circ$ producing a circular illumination pattern of 2.15 m of diameter at a 4 m distance.

Using a 40x microscope objective (Nikon plan fluor, NA=0.75, WD=0.66 mm, omitted in the figure), the return signal is collected with a 50x50μm^2 active area single-pixel, single photon sensitive SPAD detector [2] characterized in the section 3.1.1. The SPAD detector operates in a TCSPC mode and the arrival-time of the return photons is acquired in a temporal histogram form. The return temporal histogram has 1800 time bins of 12.8 ps each with an acquisition time of 250 ms per each histogram. The corresponding 3D images of the scene are acquired with the standard 3D ToF camera (CamBoard pico flexx PDM Technologies) producing 64x64 pixels 2D images of the investigated scene with intensity-encoded depth of centimetres resolution.

We test the suggested approach on four different scenes composed by either people or objects freely moving in a 2.5x3x4 m^3 room. The investigated scenes are a single person, two people, a square and the letter "T" shown in Fig. 7.1 (b). In the (b-c) cases, the targets wear a white vest of non retroreflective material (spunbond meltblown spunbond, SMS), while the "T" and the square (d-e) are covered by white paper. In order to evaluate the role of the background, we test the proposed 3D imaging paradigm on three different backgrounds composed either by a static square object on the left, a square object on the right or by an uniform background .

We collect 10000 temporal histogram-3D image pairs for each scenario for a total of 40000 measurements. We then apply a supervised learning data-driven approach to retrieve the transformation F^{-1} mapping the return temporal histogram into the corresponding 3D image.
We split the ten thousands measurements acquired for each scenario in 9000 and 1000 examples respectively used for the learning and for testing of the NN retrieval algorithm. In the training process, 630 examples are used for validating, corresponding to 0.07% of the training examples.
We reshape the 64x64 pixels images into a 1D array of 64x64=4096 entries to train the NN with 1D array in input and in output. The temporal histograms produced by the single-photons detector are already arranged in a 1D array. The input of the NN is a 1D array of 1800 entries corresponding to the 1800 time bins temporal histograms. The output of the NN is a 1D array of $64 \times 64 = 4096$ entries corresponding to the depth encoded 3D images.

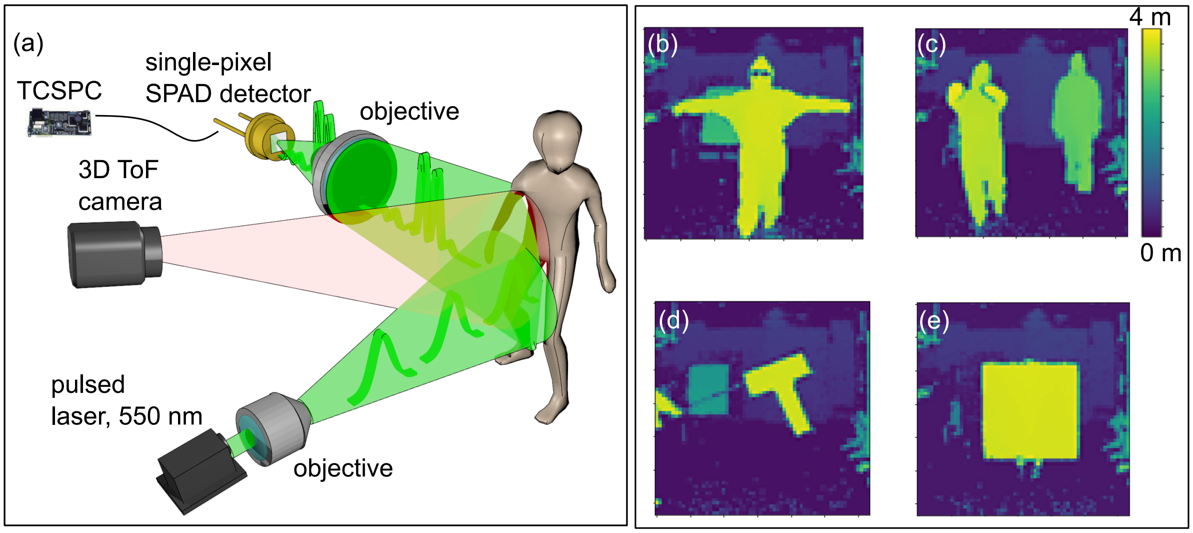

Figure 7.1: **Intelligent Lidar experimental setup.** (a) A pulsed light source of 550 ± 40 nm wavelength and 19.5 MHz repetition rate flash-illuminates the entire scene. A time-resolving, single-pixel SPAD detector collects the return signal, while a 3D ToF camera acquires the 3D image of the scene in colour-encoded depth 2D image. (b) 3D images of the four investigated scenes. The scenes we use to test the proposed approach are composed by a person, two people, a letter "T" and a square object (b-e respectively) freely moving within of 2.5x3x4 m^3 room.

The NN is trained using a mini-batch gradient descent algorithm with a batch size of 64 input-output pairs of training examples performed on a desktop computer equipped with an Intel Core i7 Eight Core Processor i7-7820X at 3:6 GHz and a NVIDIA GeForce RTX 2080 Ti with 11Gb of memory. The examples are usually normalized and randomized before the training process to generalize the predicted solution and compare features covering a wide range of values at different scale. The total number of epochs used to train the NN is 200 and each epoch requires 5 seconds of training for an overall training time of 17 minutes. As discussed in the previous chapter, the number of the epochs has been chosen accordingly to the trend of the Mean Square Error as a function of the number of iterations.

After the NN has been trained using the training data, we obtain the set of learning parameters **w** and **b** defining the transformation F^{-1}. The 3D image of the investigated scene can be retrieved using only the temporal histogram of the return photons and the learning algorithm recovered with the NN. We evaluate the performance of the learning algorithm testing the predicted solution on the testing dataset never seen by the learning algorithm during the training process. Using the proposed 3D imaging paradigm, the 3D retrieval of the investigated scene can be obtained from a single acquisition return photons temporal histogram in 30 μs, leading to a maximum frame-recovery rate of 33 kHz.

7.2 Spatial resolution

Since the proposed approach recovers the 3D image of scenes using only arrival-time measurements, the temporal resolution of the detection system strongly affects the transverse and depth spatial resolution of the 3D retrieval. We therefore consider the temporal resolution of the detection system under the experimental conditions in Fig. 7.1. Figure 7.2 shows the temporal IRF measured collecting the return signal of a round target of 2 cm diameter. The temporal resolution of the detection system is 250 ps, measured as the FWHM of the return signal.

In order to evaluate the transverse spatial resolution of the suggested approach, we now consider the experimental conditions depicted in Fig. 7.3 where a light source flood illuminates the scene to be retrieved. We then calculate the minimum spatial difference that can be resolved by the detector relating the temporal resolution of the detection system with the spatial resolution on the transverse plane. We then consider two isolated points A and B spatially separated by a distance δ on the transverse plane and respectively located at a distance d and D from the sensor. With reference to

Fig. 7.3, the difference Δt in the Tof of the two targets is:

$$\Delta t = \frac{D-d}{c} \tag{7.1}$$

where c is the speed of light. Since the system composed by the two targets and the sensor create a right triangle, the spatial distance δ and the ToF difference ΔT are related as follows:

$$\delta = \sqrt{D^2 - d^2} = \sqrt{(d + c\Delta t)^2 - d^2} = c\Delta t \sqrt{\frac{2d}{c\Delta t} + 1} \tag{7.2}$$

Since the experimental setup has a temporal resolution Δt of 250 ps, the spatial resolution of the system is 78 cm at 4 metre distance on the transverse plane. The depth resolution $\delta_z = c\Delta t$ of 7.5 cm is instead directly obtain by the temporal resolution.

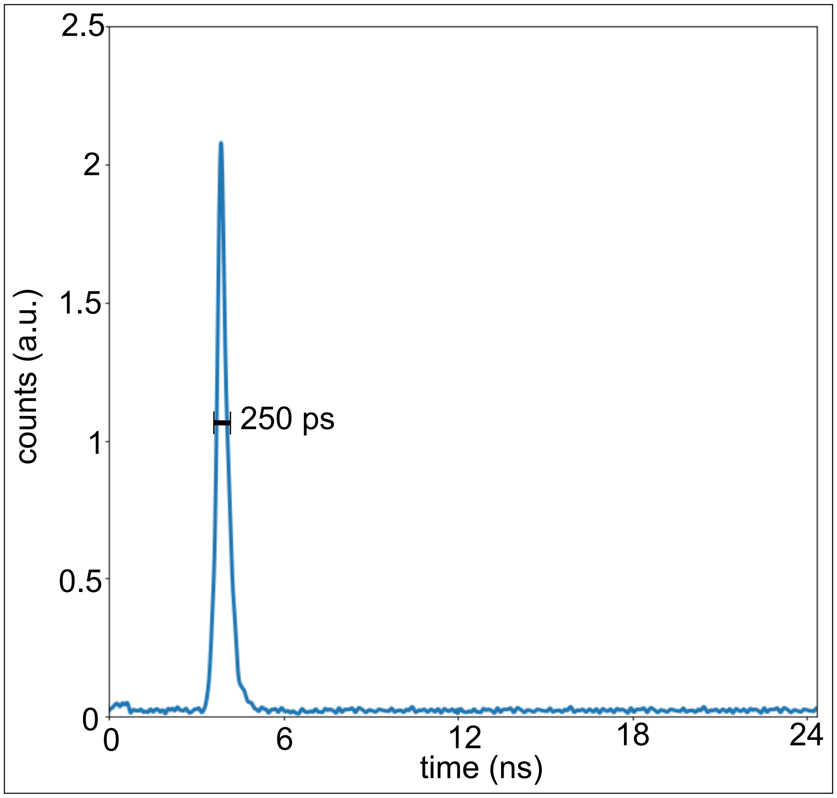

Figure 7.2: **Temporal IRF of the detection system.** The temporal resolution of the experimental setup is 250 ps measured as the full width at half maximum (FWHM) of the temporal return signal scattered back by a centimetre dimension round target.

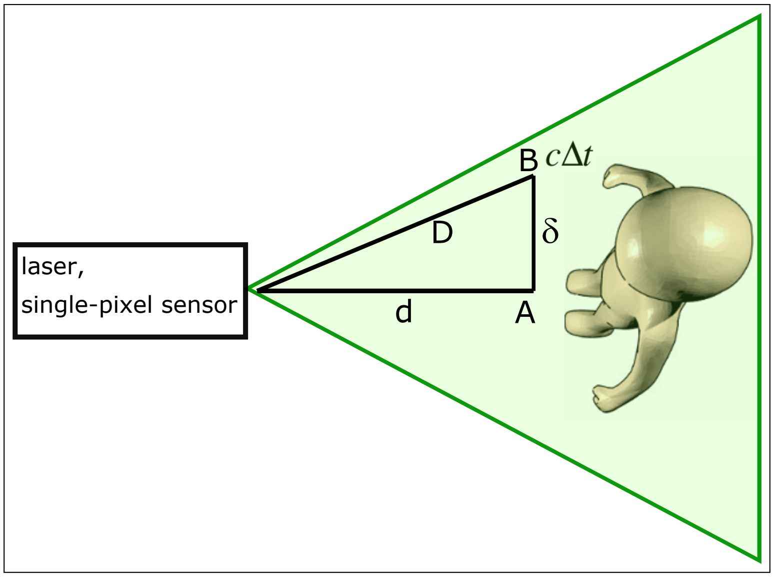

Figure 7.3: **Geometry of the spatial resolution of the single-pixel 3D imaging system.** In order to evaluate the spatial resolution of the retrieval, we consider the temporal resolution of the detection system in Fig. 7.1. Considering the geometry of two targets A and B spatially separated by a distance δ along the transverse dimension, we obtain a depth resolution of 7.5 cm and a transverse resolution of 78 cm at 4 metre distance.

7.3 Experimental results

We now report the experimental results for the four different dynamic scenarios composed by a person, two people, a square object and a letter "T" freely moving in a room with static objects in the background. We test the proposed 3D imaging paradigm on three different backgrounds composed by either a static square object on the left, a square object on the right or by an uniform background.

We then acquire ten thousand pairs of the temporal histogram of the return photons and the corresponding 3D image for each investigated scenario for a total of four thousand 3D image-temporal histogram examples. We then train the NN using the temporal histogram and the corresponding 3D images data in order to recover the transformation mapping the return photons into the corresponding 3D scene. Once the network has been trained, the ground truth 3D images of the scene are used for comparison to evaluate the retrieval obtained using only the temporal histogram.

We test the proposed approach in two distinct cases. In the first case we treat each investigated scenario separately and we train and test the NN using data specific of the scene. Referred to as "specific-scene data", the NN training data and the NN testing data belong to the same specific scene. In the one person specific-scene data, we train the NN using only the one person data and we test the recovery algorithm only on the one person data and so on for each scenario. Treating each scene separately, we test the 3D recovery on scene-specific data for each of the investigated scenario, as discussed in Section 7.3.1.

In the second case the data of the four investigated scenes are not treated as separate. Referred to as "multi-scene data", the data of the four scenes are joint and used in the same training process. In the multi-scene data, the NN retrieval algorithm is tested on data of the four scenarios, generalizing the 3D retrieval to more realistic scenes containing objects of different shapes and sizes.

As demonstrated in the experimental results in Sections 7.3.1-7.3.2, the scene-specific data approach is less complex and recovers more details than the multi-scene counterpart. At the same time, the scene-specific data training can recover only objects of the same type of those used in the training dataset and the retrieval is limited to the specific scene used in the training. On the contrary, the multi-scene data is more general and complex than the scene-specific data counterpart at the cost of a 3D image reconstruction poorer in details.

7.3.1 Training with scene-specific data

We now test the 3D imaging paradigm treating each of the four investigated scenes separately. Therefore, we train and test the learning algorithm using scene-specific data for each scenario. Figures 7.4-7.7 show the experimental results for the four separate scenarios respectively composed by a person (Fig. 7.4), two people (Fig. 7.5), a square object of 1.2x1.2 m^2 (Fig. 7.6) and a 40x40 cm^2 letter "T" target (Fig. 7.7). Each row ((a)-(d)) represents an input-output example of the same investigated scene.

The first column represents the temporal histogram of the return photons acquired with the single-pixel, time-resolving SPAD sensor. The second column represents the 3D image predicted by the NN from the temporal histogram of the return photons. Finally, the third column is the ground truth 3D image of the corresponding scene used only to evaluate the predicted 3D reconstruction. The 3D retrieval and the corresponding 3D image are colour-encoded depth images.

As shown in the temporal histogram in the first column of Figs. 7.4 and 7.7, two peaks appear in the arrival-time measurements, referred to as peak 1 and peak 2 in the figures. The first peak is due to the target freely moving within the scene, in this case the person and the letter "T" shape object. The second peak is the return signal produced by the static square object in the background, as shown in the ground truth image. Thus, the suggested approach provides an accurate 3D retrieval for dynamic and static scenarios, as demonstrated by the 3D recovery of the person freely moving (peak 1) and of the square object in the background (peak 2). Full movies with the targets freely moving within the investigated scene are shown in the links in the Appendix E.

As demonstrated comparing the second and the third column of the figures, the proposed approach provides a precise 3D image of the investigated scenes for dynamic and static scenarios using only the return arrival-time measurements recorded with a single-pixel time-resolving detector. The results highlight the accurate 3D retrieval of the investigated scene in the depth and in the transverse dimension.

Additionally, the suggested paradigm can be applied on completely different platforms provided that a prior training is performed using optical systems such as a conventional ToF 3D camera.

Since the temporal resolution of the experimental setup is 250 ps (Fig. 7.2), the system has a spatial resolution of 7.5 cm and 78 cm along the transverse and depth direction at 4 m depth distance. The limited temporal resolution of the experimental conditions leads to the loss of the details in the 3D retrieval. The 3D retrieval algorithm indeed do not recover well-defined targets and people shapes. The specific-scene data training provides 3D images of a limited statistics of scenes. Indeed, it can only recover objects of the same type of those used in the training dataset and the retrieval is limited to the specific scene used in the training. In order to evaluate the ability of the model to generalize and predict correct 3D image for diversified scenarios, we now join the datasets of the four investigated scenarios and we train the NN on multi-scene data. We then test the multi-scene 3D retrieval algorithm on each scene.

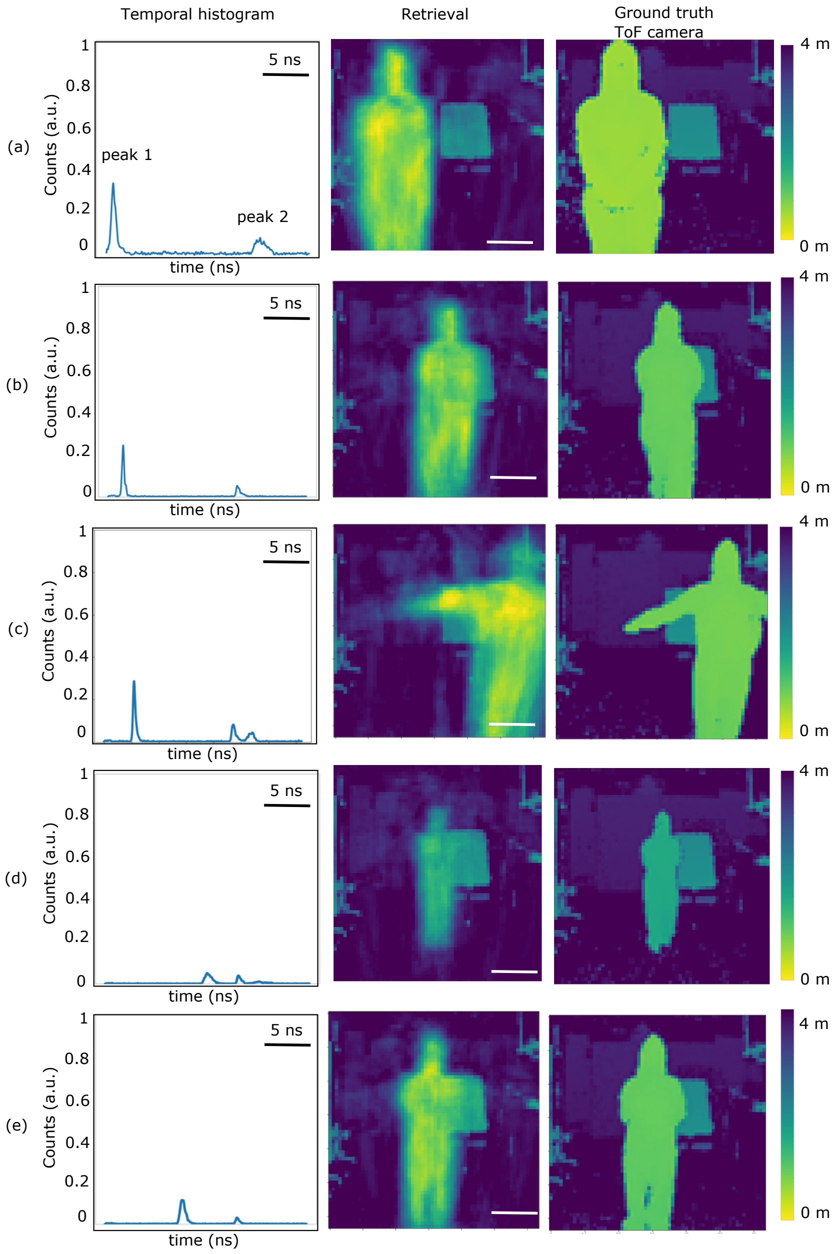

Figure 7.4: **3D image of one person scene specific data with single-pixel, time-resolving detector.** Each row ((a)-(e)) is an input-predicted output example of the one person scene, training the NN only on one person scene data. The first column is the return of the entire scene and acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the corresponding arrival-time measurement in the first column. The third column is the ground truth of the scene acquired by a standard ToF camera to evaluate the accuracy of the proposed approach. The depth dimension covers a 4 metres range and is encoded in the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metre at 4 metres depth distance.

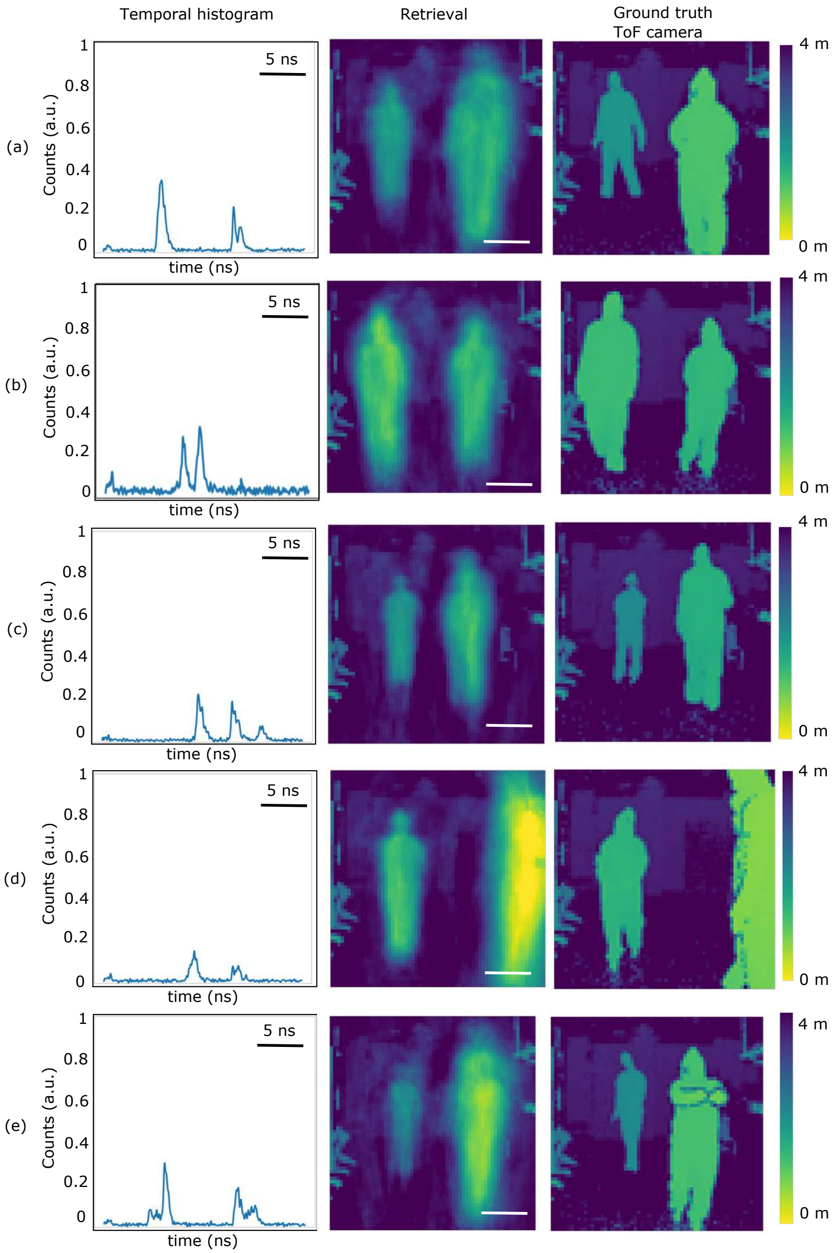

Figure 7.5: **3D image of two people scene specific data with single-pixel, time-resolving detector.** Each row ((a)-(e)) is an input-predicted output example of the two people scene, training the NN only on two people scene data. The first column is the return signal acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the corresponding arrival-time measurement in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera to evaluate the accuracy of the proposed approach. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metre at 4 metres depth distance.

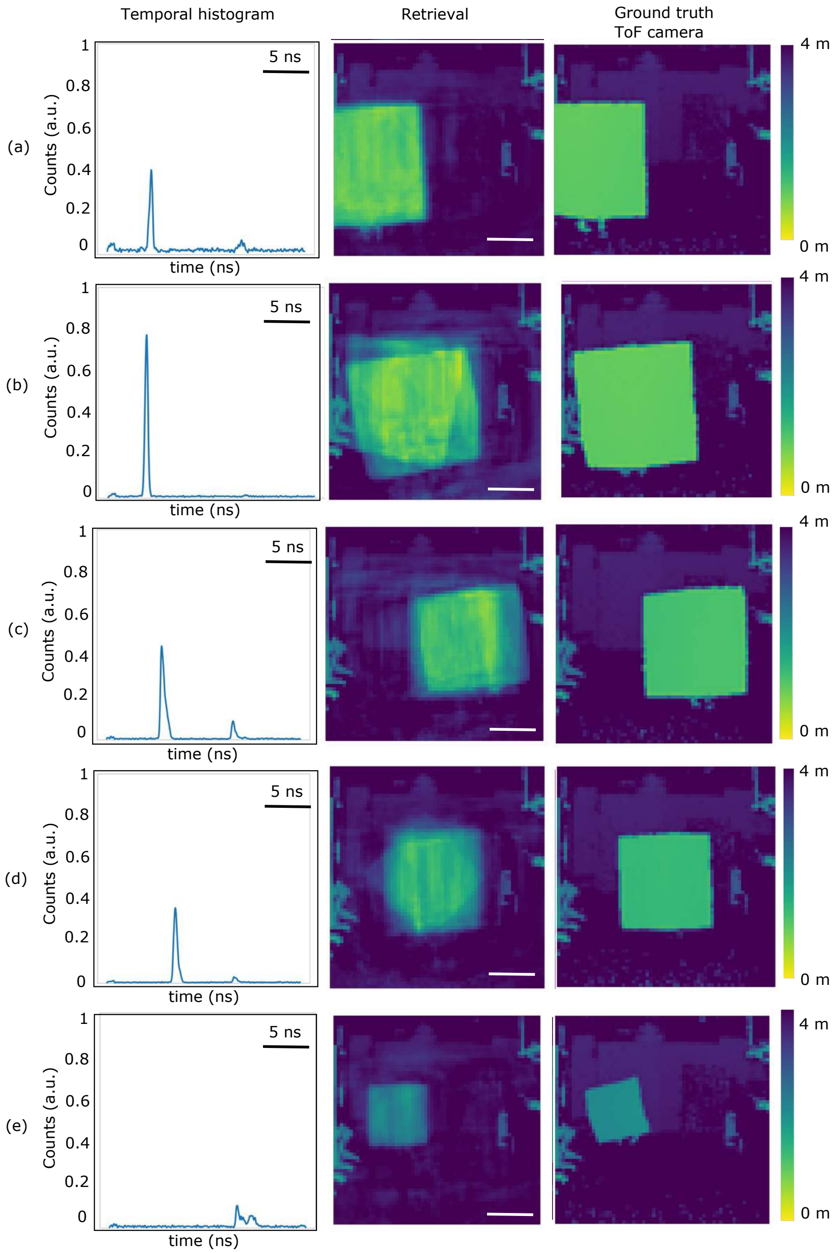

Figure 7.6: **3D image of square object scene specific data with single-pixel, time-resolving detector.** Each row ((a)-(e)) is an input-predicted output example of the 1.2x1.2 m^2 square object scene, training the NN only on square object scene data. The first column is the return acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the corresponding arrival-time measurement in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera to evaluate the accuracy of the proposed approach. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metre at 4 metres depth distance.

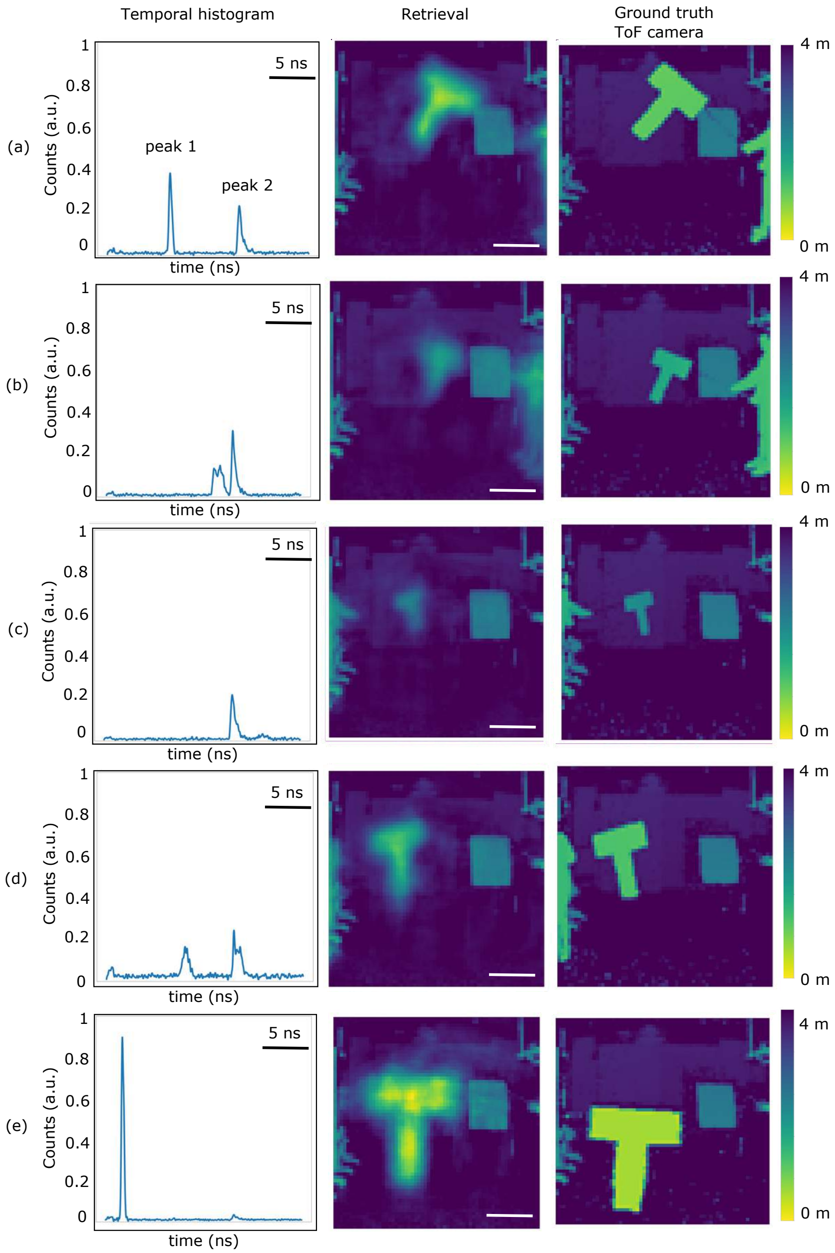

Figure 7.7: **3D image of letter "T" object scene specific data with single-pixel, time-resolving detector.** Each row ((a)-(e)) is an input-predicted output example of the 40x40 cm^2 letter "T" scene, training the NN only on letter "T" scene data. The first column is the return acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the corresponding arrival-time measurement in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera to evaluate the accuracy of the proposed approach. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metre at 4 metres depth distance.

7.3.2 Training with multi-scene data

In order to test the performance of the proposed 3D imaging approach on a more general and realistic scenario, we now test the 3D imaging paradigm joining the data of the four investigated scenes and training the NN using data of multiple scenes. We then test the 3D retrieval algorithm on multi-scene data.

As in the previous case, the scene to be recovered is composed by static and dynamic targets freely moving within a 2.5x3x4 m^3 room. Training the NN on dataset including multi-scene, the suggested approach provides the 3D recovery of more realistic and generalized scenes at the cost of a retrieval algorithm more complex than the scene-specific data counterpart.

We train the NN on multi-scene data using the same learning model of the specific-scene counterpart (Fig. 6.10). The training dataset is composed by 9500 temporal histogram- 3D image pairs of each scene for a total of 38000 training examples. Here, 2.660 examples corresponding to the 0.07 % of the training examples are used for validating. The remaining 2000 examples (500 for each scene) are used for testing the NN retrieval algorithm. As in the scene-specific counterpart, each input and output example is composed by a 1D vector of 1800 and 64×64 = 4096 entries respectively. The total number of epochs used to train the NN is 400 and each epoch requires 28 seconds of training for an overall training time of 3 hours.

Figures 7.8-7.11 show the predicted experimental results obtained training the learning algorithm with multi-scene data and testing the retrieval algorithm respectively on the one person, the two people, the square object and the letter "T" scene. As in the previous results, the first column represents the return arrival-time measurements collected with the single-pixel, time-resolving SPAD detector. The second column is the 3D image of the scene predicted by the ANN and obtained using only the photon arrival-time signal scattered by the entire scene. The third column is the ground truth 3D image of the investigated scene used only for comparison to evaluate the 3D retrieval predicted by the NN. Full movies of the targets freely moving within the scene are shown in Appendix E. As a clarification, the 3D retrieval of the example in the column (d) of Fig. 7.9, predicts only one person since the second person is out of the FoV of the SPAD detector, as demonstrated by the single peak of the corresponding return temporal histogram.

As in the scene-specific data counterpart, the proposed approach correctly recovers the 3D image of the investigated scene using only the arrival-time measurement of the return signal. However, the 3D retrieval is strongly affected by the spatial resolution of the experimental apparatus, determining the minimum resolvable feature and leading to the loss of some details in Figs. 7.8-7.11.

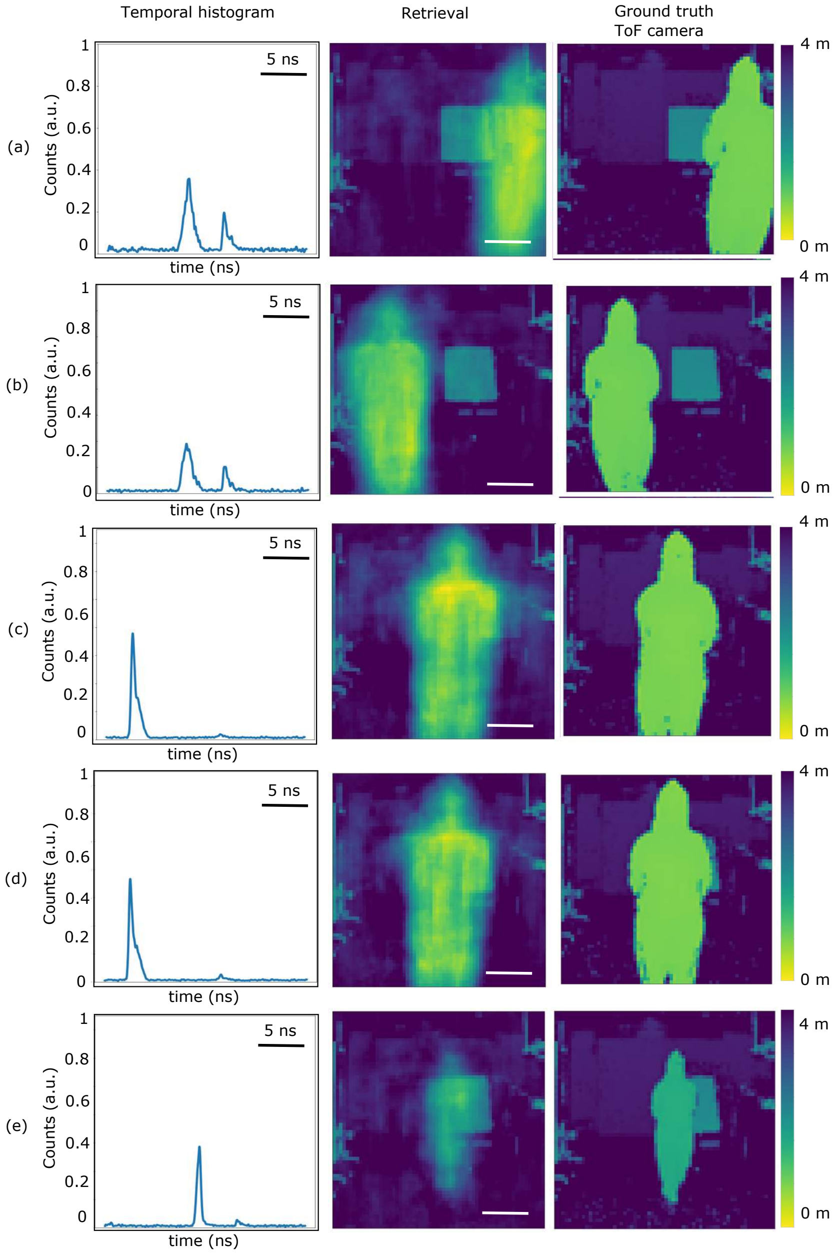

Figure 7.8: **3D image of a person scene from temporal histogram with single-pixel, time-resolving detector by multi-scene training.** Each row ((a)-(e)) is an input-predicted output example of the one person scene, training the NN on joint data of the four scenes. The first column represents the return acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the arrival-time measurements in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metres at 4 metres depth distance.

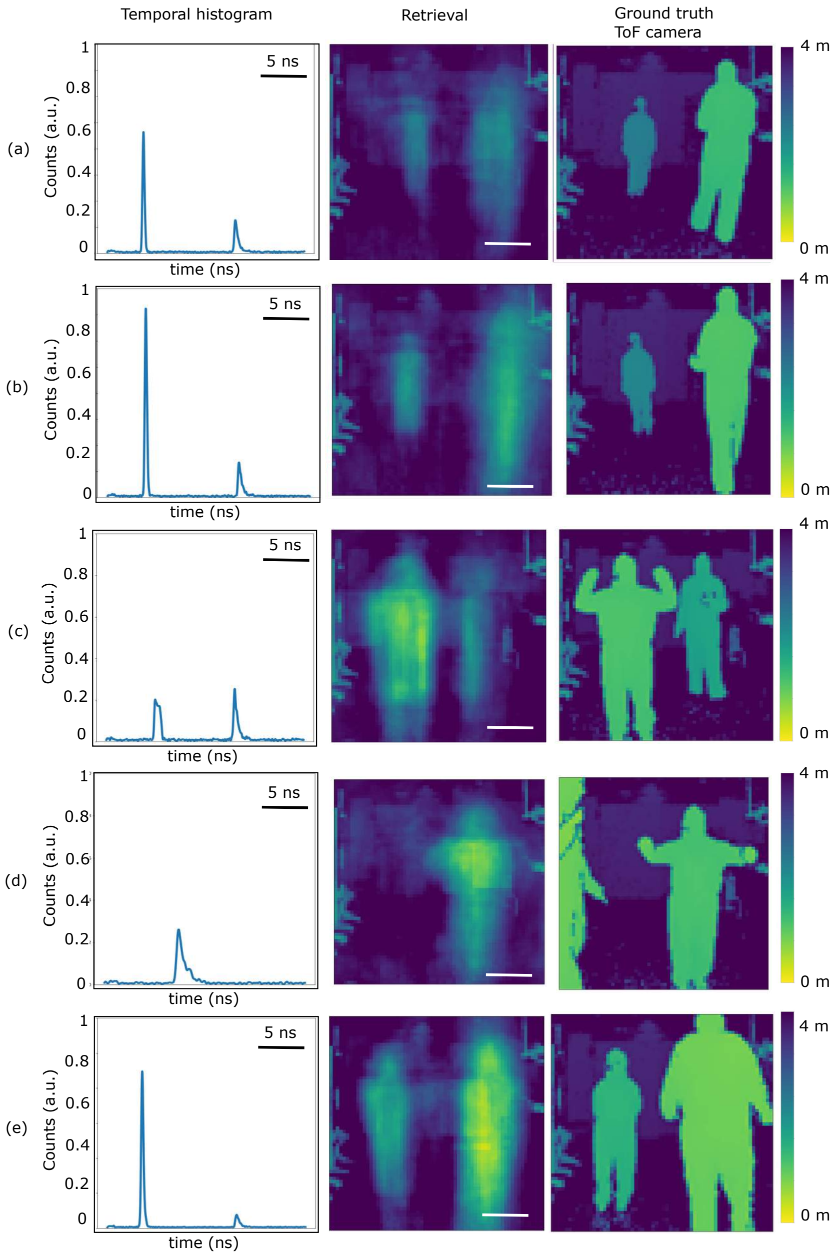

Figure 7.9: **3D image of two people scene from temporal histogram with single-pixel, time-resolving detector by multi-scene training.** Each row ((a)-(e)) is an input-predicted output example of the two people scene, training the NN on joint data of the four scenes. The first column represents the return acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the arrival-time measurements in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metres at 4 metres depth distance.

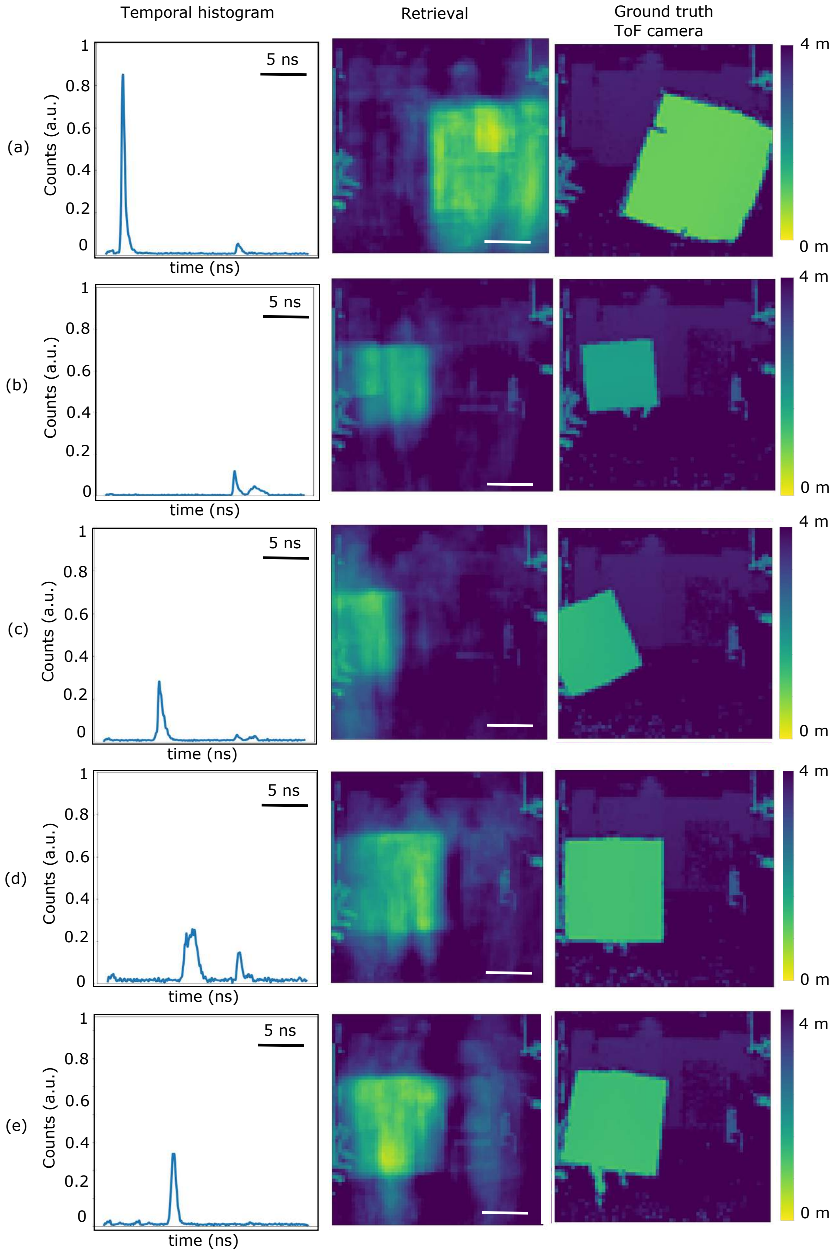

Figure 7.10: **3D image of square object scene from temporal histogram with single-pixel, time-resolving detector by multi-scene training.** Each row ((a)-(e)) is an input-predicted output example of the square object scene, training the NN on joint data of the four scenes. The first column represents the return acquired by the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the arrival-time measurements in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metres at 4 metres depth distance.

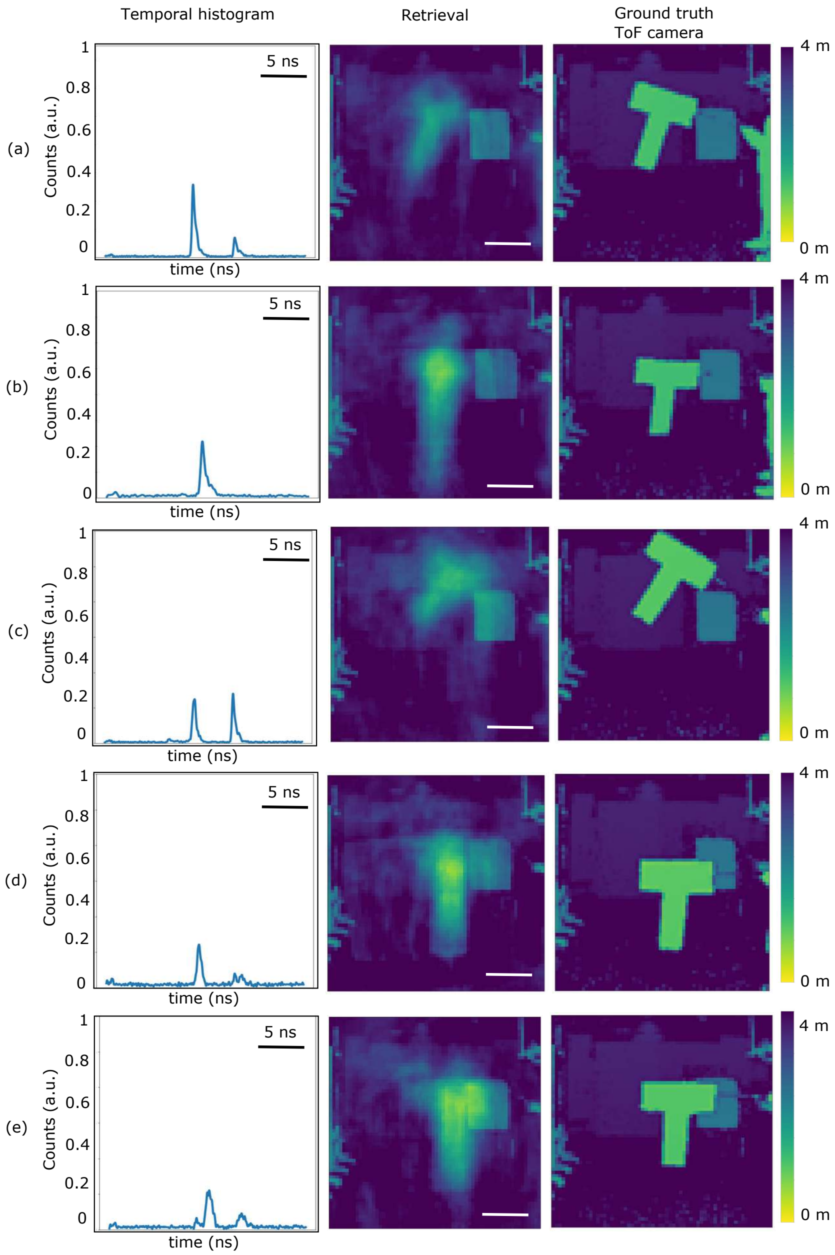

Figure 7.11: **3D image of letter "T" scene from temporal histogram with single-pixel, time-resolving detector by multi-scene training.** Each row ((a)-(e)) is an input-predicted output example of the letter "T" scene, training the NN on joint data of the four scenes. The first column represents the return acquired with the time-resolving, single-pixel detector. The second column is the 3D retrieval predicted by the learning algorithm from the arrival-time measurements in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera. The depth dimension covers a 4 metres range and is encoded by the colour scale of the 2D images. The white scale bar shown in the second column corresponds to a spatial distance of 1 metres at 4 metres depth distance.

The multi-scene training results demonstrate the ability of the suggested 3D imaging approach to generalize and predict correct output samples for multiple scenes. Therefore the suggested 3D imaging approach guarantees the generality of the predicted 3D solution adding more complex and variegate scenes in the training process as in the multi-scene training.

However, the NN algorithm struggles to retrieve the correct shape and is not able to clearly distinguish the shape of different targets due to the limited 78 cm transverse resolution of the experimental conditions and to the generalized training. The limitation induced by the finite spatial resolution is clearly visible in the 3D retrieval of the letter "T" in the rows (d)-(e) of Fig. 7.11. Since the letter "T" has dimension of a 40x40 cm^2, the learning algorithm is not able to clearly distinguish between the person and the object.

Since the 3D image of a scene is recovered by arrival-time measures, the suggested 3D imaging paradigm can be employed in a completely different platform. This leads the way to cross-modality 3D imaging that can be extended to a wider range of compact sensors outside the optical domain such as acoustic or radio waves. We now demonstrate the cross-modality of the approach by a radio-frequency RADAR sensor used as a full 3D imaging device.

7.4 Cross-modality optical 3D imaging with Radar

We now test the system substituting the laser and the single-photon detector with a radio-frequency impulse RADAR transceiver to demonstrate the cross-modality of the proposed 3D imaging approach. Figure 7.12 shows the experimental setup. The cross-modality of the proposed 3D imaging paradigm opens new routes to 3D imaging platforms using a wider range of sensors such as acoustic or radio waves.

In this case the RADAR chip (Novelda XeThruX4) is a single transmitter and receiver channel operating at 7:29 GHz frequency with a bandwidth of 1.4 GHz and pulse duration of 670 ps with a sampling rate of 23x10^9 samples/s. The radar chip emits RADAR pulses towards the investigated scene and collects the return signal in a temporal histogram form, while the conventional ToF 3D camera synchronously acquires the corresponding 3D image. We then train the NN using the same NN model discussed in Section 6.5. The RADAR return signal is the input of the NN, while the 3D is the output of the NN. We acquire 6500 temporal histogram-3D image pairs of which 5580, 420 and 500 examples are respectively used in the training, validating and testing process. Once the ANN has been trained using pairs of the return signal and of the corresponding 3D image, we test the retrieval algorithm using only the temporal return data. The return signal is collected in a temporal histogram of 91 time bins of 293 ps each. As in the optical equivalent, the ToF camera is

acquiring 64x64 pixels images of the scene by a colour-encoded depth map. The data are normalized and randomized to guarantee the generality of the predicted solution and to compare features covering a wide range of values at different scale.
The RADAR transceiver has a frequency bandwidth of 1.4 GHz and a pulse duration of 670 ps measured as the temporal IRF of the RADAR emission. The IRF (Fig. 7.13) has been measured as the FWHM of the temporal return signal of a centimetre-dimension target. According to the formula in Eq. (7.2), the temporal resolution of the chip induces a spatial resolution of 20 cm and 90 cm at 2 metres of distance along the depth and the transverse spatial dimension. Since the system has a lower temporal resolution than the optical equivalent, the quality of the 3D retrieval is poorer than the previous optical counterpart.
Figure 7.14 shows the obtained experimental 3D images predicted from the NN using the return RADAR signal. In this case the scene to be recovered is composed by an isolated person moving back and forward within the investigated scene. The first column represents the temporal trace of the return RADAR signal produced by the person moving in the scene for five investigated examples ((a)-(e)). The second column is the 3D image predicted by the learning algorithm from the temporal trace in the first column. The third column represents the 3D image ground truth to evaluate the performance of the 3D retrieval. Since the experimental setup is affected by the spatial ambiguities produced by the transit-time symmetry of the system, in this case the target is moving only along the z axis dimension of a 3x3x4 m^3 scene. A full video of the 3D retrieval is available in the Appendix E.

Although the quality of the 3D retrieval is lower than the optical equivalent, the suggested approach is still able to recover the 3D location of the target along the depth and the transverse dimensions. Using a data-driven approach, the proposed method represents a compact 3D imaging paradigm from a temporal trace and an optical trained ANN algorithm, transforming an ubiquitous RADAR sensor into a 3D imaging device.
The suggested approach is therefore able to recover the 3D image in cross-modality beyond the optical domain. However, the suggested approach is affected by some limitations such as the spatial degeneracy of the temporal symmetrical 3D scene as discussed in the next section.

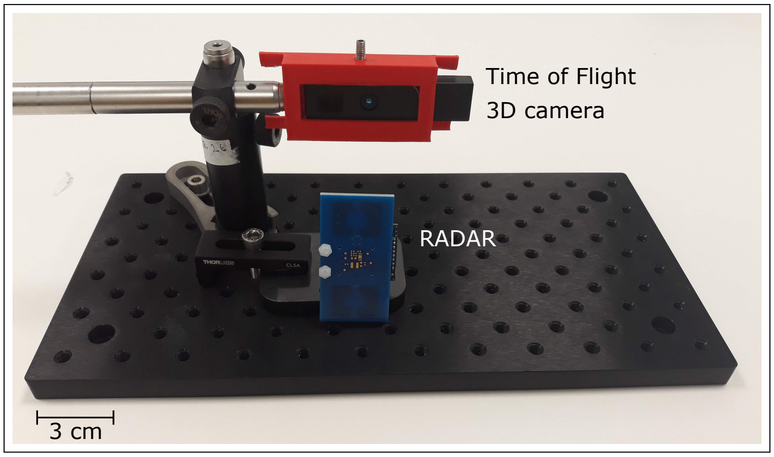

Figure 7.12: **Cross-modality 3D imaging experimental setup by a RADAR chip.** We test the paradigm using a RADAR transmitter-receiver chip and a conventional 3D camera to demonstrate the cross-modality 3D imaging of the proposed method. The RADAR chip emits RADAR pulses towards the scene and collects the return RADAR signal. The 3D camera simultaneously acquires the corresponding 3D image of the scene. Synchronizing the chip and the ToF camera acquisition, we then apply the same learning approach used in the optical equivalent scenario. The entire system is mounted on a 21x9 cm^2 breadboard.

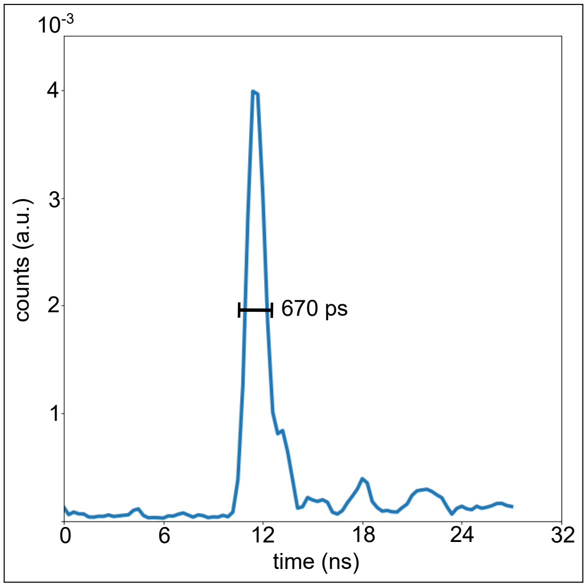

Figure 7.13: **Temporal IRF of the RADAR transceiver.** The temporal resolution of the RADAR transceiver is measured as the FWHM of the temporal signal scattered from a centimetre dimension target.

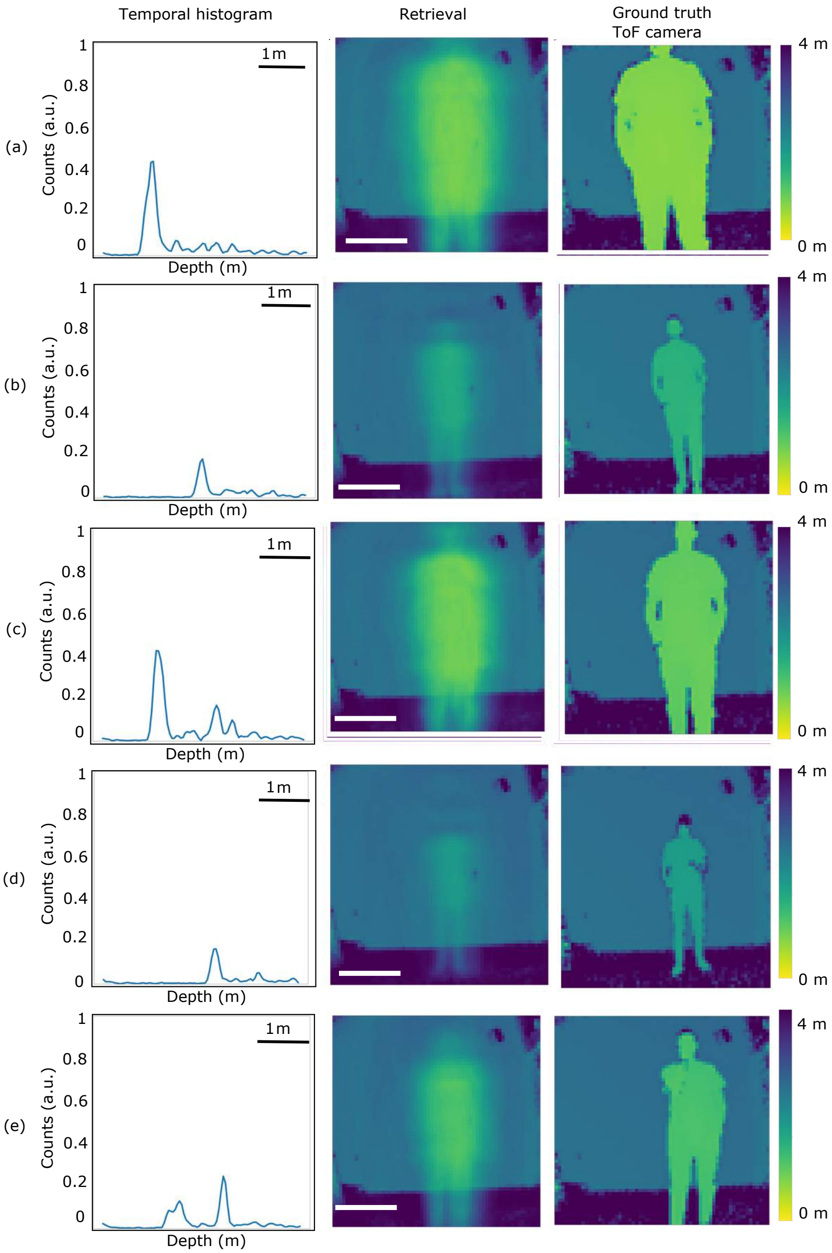

Figure 7.14: **Cross-modality optical 3D imaging with Radar.** Each row ((a)-(e)) represents an input-predicted output example of a person moving back and forward within a 3x3x4 m^3 scene. The first column represents the return RADAR signal from the entire scene and acquired by the RADAR transceiver at 7.29 GHz radiation. The second column is the 3D retrieval predicted by the learning algorithm from the RADAR return in the first column. The third column represents the ground truth of the scene acquired with a standard ToF camera. Due to the wide pulsed duration of the RADAR emitter of 670 ps, the 3D retrieval of the scene has a poorer reconstruction quality than the optical equivalent.

7.5 Limitations of ILidar approach.

The main limitation affecting the proposed 3D imaging paradigm is the spatial degeneracy induced by the symmetry of the problem. We now provide the experimental evidence of the spatial degeneracy of ToF symmetric 3D images theoretically discussed in Section 6.4.1.
We now repeat the experimental measurements discussed in Sections 7.1-7.3 testing the proposed 3D imaging approach on a scene composed by a macroscopic target freely moving within an uniform background scene. Figure 7.15 shows the experimental results obtained using the same procedure and the same NN model of the results discussed in 7.3.1. We acquire 10000 3D image-temporal histogram pairs. The results have been obtained training the NN with 9000 input-output examples and testing the retrieval algorithm on 1000 input examples.

Figure 7.15 shows the effect of the spatial symmetry for five input-output experimental examples (a-e) of an isolated person freely moving within an uniform background scenario. Full movies with the targets freely moving within the investigated scene are shown in the links in the Appendix E. As in the previous cases, the first column represents the temporal histogram of the return photons, whereas the second column is the 3D image predicted by the NN algorithm. Finally, the third column represents the ground truth 3D image to evaluate the 3D recovery in the second column.
In this case, the NN algorithm struggles to identify the correct 3D retrieval since multiple 3D images are both compatible with the same temporal histogram. Considering a target standing either in the right side or in left side of an uniform FoV, the NN algorithm recovers the target in both sides since both configurations are compatible with the same temporal trace.
The spatial degeneracy affects the 3D retrieval in every symmetrical direction of the investigated scenario. The experimental results present only the left-right side ambiguities since the NN has been trained only on vertical objects examples. Due to the spatial symmetry of the experimental geometry, all the isolated targets placed on ToF symmetrical positions respect to the detector location, produce the same return signal. As a consequence, multiple symmetrical 3D images are assigned to the same temporal histogram and symmetrical ambiguities both compatible with the temporal histogram are simultaneously retrieved. However, the background of the investigated 3D scene plays a crucial role in removing the spatial ambiguity produced by the spatial degeneracy of the problem. Introducing a non uniform background in the investigated scene, we can indeed remove the spatial degeneracy and univocally define a temporal histogram-3D image correspondence.

As an example of the key role played by a non uniform background, we now consider the scenario reported in Fig. 7.4 where a person is freely moving within a non uniform background scene. In this case, the presence of a background target placed in locations of the FoV non left-right symmetrical, removes the left-right symmetry and the spatial ambiguities in the 3D retrieval. Considering a background object asymmetrically placed in the background, the person silhouette subtracts signal from the background return peak according to the relative location of the person respect to the background object. As a consequence, if the person is standing in the left side of the FoV as shown in column (a), the return signal collected by the detector contains both the return signals generated by either the person and the background object. On the contrary, if the person is standing in the right side of the FoV as shown in column (b), the person is partially covering the return signal produced by the background target and the temporal histogram contains only partially the return signal of the background object. Since the signal produced by the background object varies according to the location of the person, the left-right spatial ambiguity is removed. Therefore, only the corresponding 3D reconstruction is compatible with the temporal histogram, as reported in the previous experimental results.

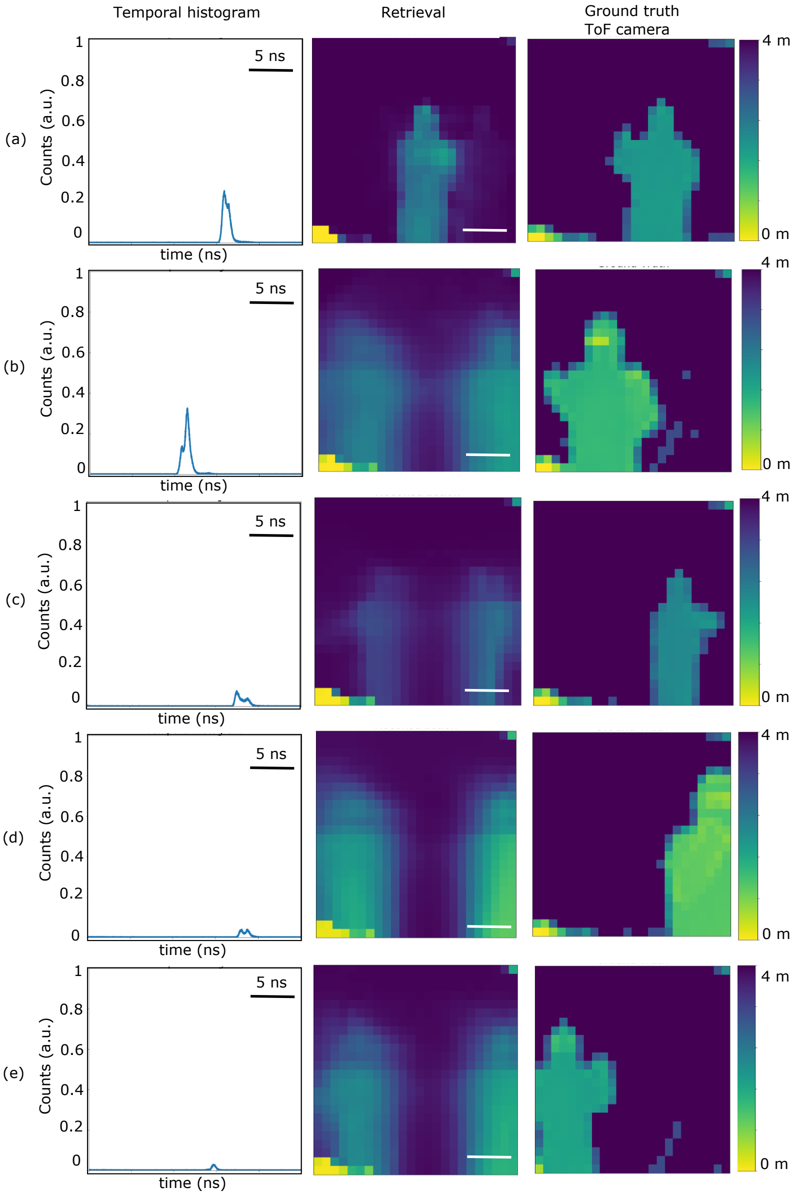

Figure 7.15: **3D image of a person freely moving in an uniform background scene with single-pixel, time-resolving detector.** Each row ((a)-(e)) is an input-predicted output example of the investigated scenario composed by a person freely moving within an uniform background. The first column is the arrival-time of the return. The second column is the 3D retrieval predicted by the learning algorithm from the arrival-time measurement in the first column. The third column is the ground truth of the scene acquired with a standard ToF camera. Due to the spatial symmetry of the problem, all the targets located in ToF symmetrical positions have the same temporal histogram. Considering the spatial symmetry of the experimental conditions, two isolated targets placed either in the left or in the right side of the FoV have the same arrival-time and the NN predicts both the configurations compatible with the temporal histogram.

7.6 Conclusions

Real-time 3D imaging of LOS scenes represents a crucial task with many applications in real-life scenarios such as self-driving cars, remote sensing and machine vision. The most common 3D technologies are based on 3D holography, stereo-vision and ToF 3D imaging. The ToF 3D imaging technology is based on illuminating the investigated scene with CW or pulsed light sources and collecting the return signal. Known as LiDAR, this approach combines the ToF information and the spatial information provided either by scanning the scene, or using a pixelated sensor or structured illumination.
In order to retrieve the full 3D information of the scene by the temporal return signal, current ToF imaging techniques are affected by some limitations such as moving parts, scanning systems or pixelated sensor requirement. It is indeed not possible to retrieve an unique 3D image of the scene exclusively using a single temporal histogram of the return photons since multiple target locations are compatible with the arrival-time measurement. The recovery of the 3D image of a direct scene is therefore a strong ill-posed problem when no spatial information about the scene is provided, as happens in a single-acquisition of arrival-time measurements with a single-pixel sensor.

Here, we demonstrated a new 3D imaging paradigm of LOS scenes by arrival-time measurements acquired by a single-pixel, time-resolving detector. In order to retrieve the 3D image of the investigated scene from the corresponding return temporal histogram, the suggested method employs a data-driven approach to provide the prior knowledge about the scene lost using a single-pixel detector. Employing the suggested approach, we provide a statistical representation of all the possible scenes to be imaged on the basis of which a machine learning algorithm can be trained.
The suggested method represents a 3D imaging paradigm obtained from a single temporal profile, paving the way to new forms of 3D imaging. Since no scanning system or multiple measurements are required, the proposed 3D imaging approach can additionally provide a 3D image of the investigated scene at potentially kHz or even MHz frame rate faster and more compact than any other single-pixel ToF imaging technology. Once the NN is trained, a complete 3D image can be retrieved by the acquisition of a single temporal histogram in 30 μs on a standard laptop, reducing the memory and the time required for the data storage, handling and transfer with considerable profits for remote sensing applications.
As demonstrated training the NN on scene-specific data and multi-scene data, the experimental results demonstrate the versatility and the ability of the proposed 3D imaging approach to generalize and predict correct output samples for variegate multiple scenes. Fully exploiting the benefits of using a data-driven approach, the results can be further improved including more complex and

variegate scenes in the training process. However, the suggested approach is affected by some limitations such as the ambiguities in the 3D retrieval of uniform background scenes, as demonstrated in the experimental results. The suggested approach represents a cross-modality 3D imaging methodology. Once an optical based training has been performed, the suggested method can then be extended to a wide range of compact platforms and devices such as proximity or acoustic sensors, paving the way to a new form of 3D imaging obtained from a single temporal histogram.

Chapter 8

Conclusions and future perspectives

In this thesis we investigated the 3D retrieval of Line-Of-Sight (LOS) and Non-Line-Of-Sight (NLOS) scenes with a single-pixel, single-photon sensitive detector. The use of single-pixel detectors simplifies the hardware complexity of the current Time-Of-Flight (ToF) 3D imaging technologies.

In the first part of this thesis, we investigated the 3D recovery of a LOS scene combining a lock-in amplifier and a single-pixel camera. The single-pixel camera acquires the return light and the lock-in amplifier extracts the amplitude and the phase of the signal. The amplitude of the beam provides the x-y information, whereas the phase encodes the depth information. As reported in the experimental results, the proposed method recovers the retrieval of direct scenes with a depth resolution of 5 mm.

However, some limitations affect the suggested approach. The temporal resources required for high-resolution images are indeed impractical and not compatible with the real-time frame-rates of commercial ToF 3D cameras. Moreover, the combination of a digital-micromirror-device (DMD), a single-pixel detector and a lock-in amplifier results in a bulky device. On the contrary, conventional ToF cameras offer a compact and portable 3D imaging system.

Despite these limitations, the proposed approach offers a tunable depth range. Indeed, the ambiguity-free range distance varies according to the modulation frequency of the reference beam. Moreover, the proposed approach provides the recovery of direct scenes without employing any short pulsed illumination or picosecond temporal resolution sensors.

The second part of the current thesis is dedicated to the 3D imaging of NLOS scenes with a time-resolving, single-pixel camera. As typical NLOS imaging systems, the proposed technique exploits the arrival-time information of the light back-scattered from the hidden targets. The 3D

information is then recovered applying the back-scattered imaging algorithm that models the propagation of the light from the target to the detector. Other computational imaging algorithms for looking around corners employ the frequency-wavenumber (f-k) migrations and the Phasor-Field (PF) approach.
However, the multiply back-scattered signal is typically weak and it decreases with the inverse of the square distance. Therefore, the proposed approach requires high dynamic range and picosecond time-gated, single-photon sensitive detectors. Moreover the 3D image reconstruction requires complex imaging algorithms. The recovery of NLOS scenes is then an interdisciplinary field of research in continuous progress.
The experimental results reported in this thesis demonstrated the 3D recovery of NLOS scenes by single-photon sensitive, high temporal resolution single-pixel camera. The single-pixel camera approach allows more freedom in choosing the optimal single-pixel detector and offers the advantage of imaging with no moving scanning parts.
In this thesis we experimentally investigated the recovery of hidden scenes by testing different scenarios and different single-pixel detectors. We demonstrated the full colour retrieval of a Red-Green-Blue (RGB) coloured targets by using a super-continuum laser to illuminate the scene. The use of a high photon detection efficiency (PDE) SPAD allows the recovery of a NLOS scene with sub-second acquisition times, as demonstrated in the experimental results.

In the last and most interesting part of this thesis, we introduce the Intelligent Lidar (ILidar), a 3D imaging approach by time-resolving, single-pixel sensor. The ILidar paradigm allows the recovery of the 3D information of LOS scenes using only the arrival-time information of the return photons. By combining arrival-time measurements and a neural network (NN) retrieval algorithm, the proposed methodology represents an innovative concept of depth image obtained from a single temporal profile.
The ILidar approach overcomes the limitations of the current ToF imaging systems. In order to obtain the transverse spatial information, standard ToF technologies indeed require either a scanning system with moving parts, or a pixelated sensor. The current 3D imaging technologies are then characterized by many measurements and large amount of data scaling with the transverse spatial resolution of the retrieval.
The suggested ILidar approach produces a 3D image from a single temporal histogram without requiring any scanning system, moving part or pixelated sensor. The retrieval of the scene can be obtained from a single acquisition of the return photons temporal histogram in 30 μs, potentially leading to a maximum frame-recovery rate of 33 kHz.

Since the computational resources are used uniquely in the NN training process, the suggested 3D imaging approach is also efficient in terms of computational resources. After the retrieval algorithm has been trained, the suggested approach recovers the depth image by using just one single temporal trace, requiring less amount of data transferring, storage and handling than scanning or pixelated ToF techniques.
As demonstrated by the experimental results in this thesis, the ILidar approach is suitable for cross-modality 3D imaging. Since the 3D image of a scene is recovered by arrival-time measurements , the suggested 3D imaging paradigm can indeed be employed in a completely different platform provided that the system has been previously trained by a 3D imaging optical system. This leads the way to cross-modality 3D imaging that can be extended to a wider range of compact sensors outside the optical domain such as acoustic or radio waves.

According to the simulation in Chapter 3, the lock-in amplifier approach provides a depth resolution one order of magnitude better than the CW modulated 3D cameras approach. However, the experimental results demonstrated a comparable depth resolution. Future studies will investigate the discrepancies between the theoretical and the experimental depth resolution of the lock-in single-pixel camera.
The application of Computational Imaging (CI) algorithms such as compressive sensing, allows to fully exploit the capabilities of a single-pixel camera. Compressive sensing is a CI algorithm that allows to reconstruct a NxN pixels image using a reduced number of measurements down to a $log_2(N)$. A future hardware implementation of the proposed NLOS 3D imaging approach is the application of compressive sensing techniques to select the number of acquisitions without compromising the quality of the reconstruction. The proposed 3D imaging system combined with compressive sensing techniques, may in future offer a practical solution for the 3D imaging of NLOS dynamic scenes at real-time.
Another 3D imaging algorithm for NLOS is the f-k migration method. This computational imaging method describes the NLOS scene in terms of a wave-equation and analytically solves the NLOS wave equation in order to retrieve the 3D image. Recent experimental results demonstrated that the f-k migration method is easy to implement, fast and efficient in terms of memory, computational resources and resiliency to noise measurements. The f-k migration approach allows high quality 3D recovery of NLOS retroreflective scenes at real-time with a better image quality than NLOS algorithms such as back-propagation or light-cone transform.
In a future computational implementation of the results shown in this thesis, the f-k migration method could be applied to the NLOS measurements discussed in Chapter 4.

In a future implementation, we will evaluate the ILidar performances on a more generalized scenario including more variegate targets of everyday life. The depth range of the 3D ToF camera and the space of the laboratory limit the current 3D imaging retrieval to a 5 metres depth. We will test the ILidar approach on long depth-range scenarios using a longer depth range ToF cameras outside the laboratory.

As discussed in Section 7.2, the temporal resolution of the detector strongly affects the retrieval of details such as arms or legs. A further implementation is to use a single-photon detector with higher temporal resolution.

The ILidar approach can be converted in a eye-safe 3D imaging device by changing the illumination wavelength. In a future implementation, the suggested ILidar approach can be installed in a compact and portable device, as demonstrated by the RADAR cross-modality results. This will allow to test the ILidar approach outside the laboratory in order to evaluate the effects of more variegate scenarios not under the controlled laboratory conditions. The ultimate goal of the ILidar approach is to use the distance sensors already present in the current vehicles to obtain a 3D image of the surrounding area.

Appendix A

Back-projection algorithm for 3D Non-Line-of-Sight Imaging

In this section we provide the pseudo-code used to retrieve the 3D shape of the hidden scenes by back-projection algorithm proposed in [30, 107].

1. Load the dataset:

   - Ensemble of the temporal histograms $s(p_m)$ of the return photons collected at a pixel p_m of the 20x20 pixels field of view.
   - Laser spot chosen a the origin $O(x=0, y=0, z=0)$ of our frame of reference
   - Positions p_m of the 20x20 pixels of the field of view with m=1,..,400.

2. Define the 3D voxels space $V(x,y,z)$ made of 10^6 voxels

3. Initialize the likelihood b of the voxels space to zero:

   - $b_k = 0$ for k=1,.. 10^6 where the index k indicates the k-th voxel v_k

4. Calculate the likelihood b_k of each voxel v_k
   for k=1,...,10^6:

   - Define voxels coordinates $(x,y,z)_k$
   - Calculate the distances $d(v_k, O)$ between the voxel v_k and the laser spot O
   - for each pixel p_m of the field of view with m=1,...,400:

(a) Calculate the distance $d(p_m, v_k)$ between the voxel v_k and the pixel p_m

(b) Calculate the time of arrival $t = (d(p_m, v_k) + d(v_k, O))/c$ of the contribution of the voxel k to the return signal collected at the m-th pixel

(c) Calculate the coefficient β_{km} including the Lambertian reflectance and the distances factor

(d) Calculate the likelihood of the k-th voxel: $b_k = b_k + \beta_{km} s(p_m, O, t)$

5. Apply a Laplacian filter along the z direction of the likelihood map

6. Apply a threshold along the z direction of the likelihood map

Appendix B

Hadamard masks pseudo-code

In this section we provide the pseudo-code used to generate the Hadamard patterns, one binary mask and its negative, to acquire the return signal passing across the filed of view.

1. Define the number of pixels n_p of the Hadamard patterns along the x and y direction

2. Create the n_p^2xn_p^2 pixels Hadamard matrix with entries ±1 and mutually orthogonal rows.

3. for each row i=1,2,..,n_p^2 of the Hadamard matrix

   - Reshape the i-th row in a n_pxn_p matrix to create the i-th binary mask $Mask_i$
   - Create the complementary by multiplying the entries by -1

Appendix C

Non-line-of-sight 3D imaging with a single-pixel camera video.

The videos referred to in this thesis in Chapter 4 have been made available online at the following links:

1. Back-scattered photons produced by an hidden target:

   Back-scattered signal

Appendix D

Artificial neural network model

In this section we provide the pseudo-code used to built the artificial neural network model.

- Load the dataset used in the training and in the validating process of the neural network:

    1. Set of 3D images of the investigated scenes in colour-encoded depth map with dimension p_xxp_y.
    2. Set of temporal histograms of the return photons scattered back by the targets in the scene made by n_{tbins}=1800 and 91 time bins respectively for the optical and radar return signal

- Flatten the colour-encoded depth images (with dimension p_xxp_y) in a one-dimensional array with p_xxp_y entries

- Normalize and randomize the dataset to make the features comparable and to guarantee the generality of the prediction

- Define the architecture of the neural network model as described in section 6.5:

    1. Define input layer made by n_{tbins} input neurons
    2. Define the hidden layers
    3. Define the output layer made by p_xxp_y output neurons

- Train and validate the neural network using temporal histogram-3D image pairs of the investigated scenarios

- Save the neural network model

- Test the neural network model on histogram-3D image pairs of the investigated scenarios not seen during the learning process
- Compare the prediction with the ground truth 3D image

Appendix E

3D imaging via artificial neural network with a single-pixel detector videos

The videos referred to in this thesis have been made available online at the following links:

1. Training with scene-specific data:
   (a) One person scenario.
   (b) Two people scenario.
   (c) Square object scenario.
   (d) Letter T scenario.
   (e) Cross- modality 3D imaging with Radar.
2. Training with multi-scene data:
   (a) Training with joint data.
3. Spatial degeneracy of time-of-flight symmetrical 3D scenes.

Bibliography

[1] Zurich Instruments. Principles of lock-in detection and the state of the art. *CH-8005 Zurich, Switzerland, Accessed*, 2016.

[2] M. Sanzaro, P. Gattari, F. Villa, A. Tosi, G. Croce, and F. Zappa. Single-photon avalanche diodes in a 0.16 μm bcd technology with sharp timing response and red-enhanced sensitivity. *IEEE Journal of Selected Topics in Quantum Electronics*, 24(2):1–9, 2017.

[3] Becker-hickl, hybrid photo detectors. https://www.becker-hickl.com/products/hybrid-photo-detectors. Accessed: 2019-09-19.

[4] Polimi, spadlab. http://www.everyphotoncounts.com/. Accessed: 2019-09-19.

[5] Horiba, picosecond photon detectors. http://www.horiba.com/us/en/scientific/products/fluorescence-spectroscopy/lifetime/tcspc-components/details/ppd-22585/. Accessed: 2019-09-25.

[6] G Musarra, A Lyons, E Conca, Y Altmann, F Villa, F Zappa, MJ Padgett, and D Faccio. Non-line-of-sight 3d imaging with a single-pixel camera. *arXiv preprint arXiv:1903.04812*, 2019.

[7] Gabriella Musarra, Ashley Lyons, Enrico Conca, Federica Villa, Franco Zappa, Yoann Altmann, and Daniele Faccio. 3d rgb non-line-of-sight single-pixel imaging. In *Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP)*, page IM2B.5. Optical Society of America, 2019.

[8] Piergiorgio Caramazza, Kali Wilson, Genevieve Gariepy, Jonathan Leach, Stephen McLaughlin, Daniele Faccio, and Yoann Altmann. Enhancing the recovery of a temporal sequence of images using joint deconvolution. *Scientific reports*, 8(1):5257, 2018.

[9] Mário AT Figueiredo and José M Bioucas-Dias. Restoration of poissonian images using alternating direction optimization. *IEEE transactions on Image Processing*, 19(12):3133–3145, 2010.

[10] Yoann Altmann, Aurora Maccarone, Abderrahim Halimi, Aongus McCarthy, Gerald Buller, and Steve McLaughlin. Efficient range estimation and material quantification from multispectral lidar waveforms. In *2016 Sensor Signal Processing for Defence (SSPD)*, pages 1–5. IEEE, 2016.

[11] Rachael Tobin, Yoann Altmann, Ximing Ren, Aongus McCarthy, Robert A Lamb, Stephen McLaughlin, and Gerald S Buller. Comparative study of sampling strategies for sparse photon multispectral lidar imaging: towards mosaic filter arrays. *Journal of Optics*, 19(9):094006, 2017.

[12] Matthew P Edgar, Steven Johnson, David B Phillips, and Miles J Padgett. Real-time computational photon-counting lidar. *Optical Engineering*, 57(3):031304, 2017.

[13] Joseph N Mait, Ravi Athale, and Joseph van der Gracht. Evolutionary paths in imaging and recent trends. *Optics Express*, 11(18):2093–2101, 2003.

[14] EJ Limkin, Roger Sun, Laurent Dercle, EI Zacharaki, Charlotte Robert, Sylvain Reuzé, Antoine Schernberg, Nikos Paragios, Eric Deutsch, and Charles Ferté. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Annals of Oncology*, 28(6):1191–1206, 2017.

[15] Qianqian Fang. *Computational methods for microwave medical imaging*. PhD thesis, Citeseer, 2004.

[16] Jeffrey H Shapiro. Computational ghost imaging. *Physical Review A*, 78(6):061802, 2008.

[17] Seung-Hyun Hong, Ju-Seog Jang, and Bahram Javidi. Three-dimensional volumetric object reconstruction using computational integral imaging. *Optics Express*, 12(3):483–491, 2004.

[18] M Duguay and J Hansen. Ultrahigh-speed photography of picosecond light pulses. *IEEE Journal of Quantum Electronics*, 7(1):37–39, 1971.

[19] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V Wang. Single-shot compressed ultrafast photography at one hundred billion frames per second. *Nature*, 516(7529):74–77, 2014.

[20] Andreas Velten, Everett Lawson, Andrew Bardagjy, Moungi Bawendi, and Ramesh Raskar. Slow art with a trillion frames per second camera. In *ACM SIGGRAPH 2011 Talks*, pages 1–1. 2011.

[21] G Häusler, JM Herrmann, R Kummer, and MW Lindner. Observation of light propagation in volume scatterers with 10 11-fold slow motion. *Optics Letters*, 21(14):1087–1089, 1996.

[22] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)*, 32(4):1–10, 2013.

[23] Adrian Jarabo, Belen Masia, Julio Marco, and Diego Gutierrez. Recent advances in transient imaging: A computer graphics and vision perspective. *Visual Informatics*, 1(1):65–79, 2017.

[24] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10(1):23, 2016.

[25] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics (ToG)*, 32(4):1–8, 2013.

[26] Isamu Takai, Hiroyuki Matsubara, Mineki Soga, Mitsuhiko Ohta, Masaru Ogawa, and Tatsuya Yamashita. Single-photon avalanche diode with enhanced nir-sensitivity for automotive lidar systems. *Sensors*, 16(4):459, 2016.

[27] Markus Henriksson, Håkan Larsson, Christina Grönwall, and Gustav Tolt. Continuously scanning time-correlated single-photon-counting single-pixel 3-d lidar. *Optical Engineering*, 56(3):031204, 2016.

[28] Ximing Ren, Yoann Altmann, Rachael Tobin, Aongus Mccarthy, Stephen Mclaughlin, and Gerald S Buller. Wavelength-time coding for multispectral 3d imaging using single-photon lidar. *Optics express*, 26(23):30146–30161, 2018.

[29] Susan Chan, Ryan E Warburton, Genevieve Gariepy, Jonathan Leach, and Daniele Faccio. Non-line-of-sight tracking of people at long range. *Optics express*, 25(9):10109–10117, 2017.

[30] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature communications*, 3:745, 2012.

[31] G. Musarra, P. Caramazza, A. Turpin, A. Lyons, C. F Higham, R. Murray-Smith, and D. Faccio. Detection, identification, and tracking of objects hidden from view with neural networks. In *Advanced Photon Counting Techniques XIII*, volume 10978, page 1097803. International Society for Optics and Photonics, 2019.

[32] Alessandro Boccolini, Francesco Tonolini, Jonathan Leach, Robert Henderson, and Daniele Faccio. Imaging inside highly diffusive media with a space and time-resolving single-photon sensor. In *Imaging Systems and Applications*, pages ITu3E–2. Optical Society of America, 2017.

[33] Maria Bondani, Davide Redaelli, Alessandro Spinelli, Alessandra Andreoni, Giuseppe Roberti, Patrizia Riccio, Raffaele Liuzzi, and Ivan Rech. Photon time-of-flight distributions through turbid media directly measured with single-photon avalanche diodes. *JOSA B*, 20(11):2383–2388, 2003.

[34] Reuben S Aspden, Daniel S Tasca, Robert W Boyd, and Miles J Padgett. Epr-based ghost imaging using a single-photon-sensitive camera. *New Journal of Physics*, 15(7):073032, 2013.

[35] Ron Meyers, Keith S Deacon, and Yanhua Shih. Ghost-imaging experiment by measuring reflected photons. *Physical Review A*, 77(4):041801, 2008.

[36] K Goda, KK Tsia, and B Jalali. Serial time-encoded amplified imaging for real-time observation of fast dynamic phenomena. *Nature*, 458(7242):1145–1149, 2009.

[37] Augusto Ronchini Ximenes, Preethi Padmanabhan, Myung-Jae Lee, Yuichiro Yamashita, DN Yaung, and Edoardo Charbon. A 256× 256 45/65nm 3d-stacked spad-based direct tof image sensor for lidar applications with optical polar modulation for up to 18.6 db interference suppression. In *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*, pages 96–98. IEEE, 2018.

[38] Daniele Perenzoni, Leonardo Gasparini, Nicola Massari, and David Stoppa. Depth-range extension with folding technique for spad-based tof lidar systems. In *SENSORS, 2014 IEEE*, pages 622–624. IEEE, 2014.

[39] Lucio Pancheri and David Stoppa. A spad-based pixel linear array for high-speed time-gated fluorescence lifetime imaging. In *2009 Proceedings of ESSCIRC*, pages 428–431. IEEE, 2009.

[40] Nil Franch, Oscar Alonso, Joan Canals, Anna Vilà, and A Dieguez. A low cost fluorescence lifetime measurement system based on spad detectors and fpga processing. *Journal of Instrumentation*, 12(02):C02070, 2017.

[41] Philip A Hiskett, Gabriele Bonfrate, Gerald S Buller, and Paul D Townsend. Eighty kilometre transmission experiment using an ingaas/inp spad-based quantum cryptography receiver operating at 1.55 μm. *Journal of Modern Optics*, 48(13):1957–1966, 2001.

[42] Paul D Townsend. Experimental investigation of the performance limits for first telecommunications-window quantum cryptography systems. *IEEE Photonics Technology Letters*, 10(7):1048–1050, 1998.

[43] E Charbon. Single-photon imaging in complementary metal oxide semiconductor processes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2012):20130100, 2014.

[44] Jeffrey M Roth, TE Murphy, and Chris Xu. Ultrasensitive and high-dynamic-range two-photon absorption in a gaas photomultiplier tube. *Optics letters*, 27(23):2076–2078, 2002.

[45] Becker&hickl. https://www.becker-hickl.com/products/hybrid-photo-detectors/. 2020.

[46] Becker&hickl. https://www.becker-hickl.com/wp-content/uploads/2018/11/db-hpm-06-07-v02.pdf. 2020.

[47] Ravil Agishev, Adolfo Comerón, Jordi Bach, Alejandro Rodriguez, Michael Sicard, Jordi Riu, and Santiago Royo. Lidar with sipm: Some capabilities and limitations in real environment. *Optics & Laser Technology*, 49:86–90, 2013.

[48] E Aprile, P Cushman, K Ni, and P Shagin. Detection of liquid xenon scintillation light with a silicon photomultiplier. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 556(1):215–218, 2006.

[49] M Baszczyk, P Dorosz, W Kucewicz, L Mik, W Reczynski, and M Sapor. Chemiluminescence detection method using sipm with dedicated readout circuit. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3. IEEE, 2017.

[50] Wolfgang Becker. *Advanced time-correlated single photon counting techniques*, volume 81. Springer Science & Business Media, 2005.

[51] Gerald Buller and Andrew Wallace. Ranging and three-dimensional imaging using time-correlated single-photon counting and point-by-point acquisition. *IEEE Journal of selected topics in quantum electronics*, 13(4):1006–1015, 2007.

[52] Wolfgang Becker, Axel Bergmann, and Christoph Biskup. Multispectral fluorescence lifetime imaging by tcspc. *Microscopy research and technique*, 70(5):403–409, 2007.

[53] Wolfgang Becker, Axel Bergmann, Karsten König, and Uday Tirlapur. Picosecond fluorescence lifetime microscopy by tcspc imaging. In *Multiphoton microscopy in the biomedical sciences*, volume 4262, pages 414–419. International Society for Optics and Photonics, 2001.

[54] Becker&hickl. https://www.becker-hickl.com/wp-content/uploads/2019/09/hb-bh-tcspc-1.pdf. 2020.

[55] waymo. waymo, documentation, website=https://waymo.com/lidar/. 2020.

[56] Gregory A Howland, Petros Zerom, Robert W Boyd, and John C Howell. Compressive sensing lidar for 3d imaging. In *CLEO: 2011-Laser Science to Photonic Applications*, pages 1–2. IEEE, 2011.

[57] Matthew P Edgar, Ming-Jie Sun, Graham M Gibson, Gabriel C Spalding, David B Phillips, and Miles J Padgett. Real-time 3d video utilizing a compressed sensing time-of-flight single-pixel camera. In *Optical Trapping and Optical Micromanipulation XIII*, volume 9922, page 99221B. International Society for Optics and Photonics, 2016.

[58] Matthew P Edgar, Graham M Gibson, Richard W Bowman, Baoqing Sun, Neal Radwell, Kevin J Mitchell, Stephen S Welsh, and Miles J Padgett. Simultaneous real-time visible and infrared video with single-pixel detectors. *Scientific reports*, 5:10669, 2015.

[59] Vincent Studer, Jérome Bobin, Makhlad Chahid, Hamed Shams Mousavi, Emmanuel Candes, and Maxime Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences*, 109(26):E1679–E1687, 2012.

[60] Jürgen Hahn, Christian Debes, Michael Leigsnering, and Abdelhak M Zoubir. Compressive sensing and adaptive direct sampling in hyperspectral imaging. *Digital Signal Processing*, 26:113–126, 2014.

[61] Joel Greenberg, Kalyani Krishnamurthy, and David Brady. Compressive single-pixel snapshot x-ray diffraction imaging. *Optics letters*, 39(1):111–114, 2014.

[62] Wai Lam Chan, Kriti Charan, Dharmpal Takhar, Kevin F Kelly, Richard G Baraniuk, and Daniel M Mittleman. A single-pixel terahertz imaging system based on compressed sensing. *Applied Physics Letters*, 93(12):121105, 2008.

[63] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.

[64] Baoqing Sun. Three dimensional computational imaging with single-pixel detectors. *PhD thesis*, pages 43–46, 2015.

[65] Giovanna Sansoni, Marco Trebeschi, and Franco Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009.

[66] Jan W Weingarten, Gabriel Gruener, and Roland Siegwart. A state-of-the-art 3d sensor for robot navigation. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2155–2160. IEEE, 2004.

[67] F Bruno, Gianfranco Bianco, Maurizio Muzzupappa, Sandro Barone, and ARMANDO VIVIANO Razionale. Experimentation of structured light and stereo vision for underwater 3d reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):508–518, 2011.

[68] Rudolf Tanner, Martin Studer, Adriano Zanoli, and Andreas Hartmann. People detection and tracking with tof sensor. In *2008 IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance*, pages 356–361. IEEE, 2008.

[69] Ming-Jie Sun, Matthew P Edgar, Graham M Gibson, Baoqing Sun, Neal Radwell, Robert Lamb, and Miles J Padgett. Single-pixel three-dimensional imaging with time-based depth resolution. *Nature communications*, 7:12010, 2016.

[70] A Prusak, O Melnychuk, H Roth, Ingo Schiller, and Reinhard Koch. Pose estimation and map building with a time-of-flight-camera for robot navigation. *International Journal of Intelligent Systems Technologies and Applications*, 5(3-4):355–364, 2008.

[71] Aamir Saeed Malik. *Depth Map and 3D Imaging Applications: Algorithms and Technologies: Algorithms and Technologies*. IgI global, 2011.

[72] Brent Schwarz. Mapping the world in 3d. *Nature Photonics*, 4(7):429–430, 2010.

[73] samsung. Samsung. documentation, website=https://www.samsung.com/global/galaxy/what-is/tof-camera/. 2020.

[74] Markus-Christian Amann, Thierry M Bosch, Marc Lescure, Risto A Myllylae, and Marc Rioux. Laser ranging: a critical review of unusual techniques for distance measurement. *Optical engineering*, 40, 2001.

[75] Sergi Foix, Guillem Alenya, and Carme Torras. Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.

[76] Thierry Oggier, Bernhard Büttgen, Felix Lustenberger, Guido Becker, Björn Rüegg, and Agathe Hodac. Swissranger sr3000 and first experiences based on miniaturized 3d-tof cameras. *Proc. of the First Range Imaging Research Day at ETH Zurich*, 2005.

[77] Thomas Spirig, Peter Seitz, Oliver Vietze, and Friedrich Heitger. The lock-in ccd-two-dimensional synchronous detection of light. *IEEE Journal of quantum electronics*, 31(9):1705–1708, 1995.

[78] T Oggier, R Kaufmann, M Lehmann, P Metzler, G Lang, M Schweizer, M Richter, B Büttgen, N Blanc, K Griesbach, et al. 3d-imaging in real-time with miniaturized optical range camera. In *Proc. OPTO*, pages 89–94, 2004.

[79] Tobias Möller, Holger Kraft, Jochen Frey, Martin Albrecht, and Robert Lange. Robust 3d measurement with pmd sensors. *Range Imaging Day, Zürich*, 7(8), 2005.

[80] Robert Lange and Peter Seitz. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics*, 37(3):390–397, 2001.

[81] CR Cosens. A balance-detector for alternating-current bridges. *Proceedings of the physical society*, 46(6):818, 1934.

[82] A Gnudi, L Colalongo, and G Baccarani. Integrated lock-in amplifier for sensor applications. In *Proceedings of the 25th European Solid-State Circuits Conference*, pages 58–61. IEEE, 1999.

[83] Leo Gross, Fabian Mohn, Nikolaj Moll, Bruno Schuler, Alejandro Criado, Enrique Guitián, Diego Peña, André Gourdon, and Gerhard Meyer. Bond-order discrimination by atomic force microscopy. *Science*, 337(6100):1326–1329, 2012.

[84] Daniel C Tsui, Horst L Stormer, and Arthur C Gossard. Two-dimensional magnetotransport in the extreme quantum limit. *Physical Review Letters*, 48(22):1559, 1982.

[85] Basler. baslerweb. documentation, website=https://www.baslerweb.com/en/. 2019.

[86] becom group. becom-group, documentation, website=https://www.becom-group.com/en/becom-systems/3d-time-of-flight-cameras/argos3d-cameras/argos3d-p100/. 2020.

[87] Mirko Schmidt, Klaus Zimmermann, and Bernd Jähne. High frame rate for 3d time-of-flight cameras by dynamic sensor calibration. In *2011 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2011.

[88] M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015.

[89] O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics express*, 20(17):19096–19108, 2012.

[90] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking around the corner using transient imaging. In *2009 IEEE 12th International Conference on Computer Vision*, pages 159–166. IEEE, 2009.

[91] M. Gupta, S. K Nayar, M. B Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(5):156, 2015.

[92] J. Klein, C. Peters, J. Martín, M. Laurenzis, and M. B Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific reports*, 6:32491, 2016.

[93] F. Heide, L. Xiao, W. Heidrich, and M. B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014.

[94] P. Caramazza, A. Boccolini, D. Buschek, M. Hullin, C. F Higham, R. Henderson, R. Murray-Smith, and D. Faccio. Neural network identification of people hidden from view with a single-pixel, single-photon detector. *Scientific reports*, 8(1):11945, 2018.

[95] M. O-Toole, D. B Lindell, and G. Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555(7696):338, 2018.

[96] X. Liu, I. Guillén, M. La Manna, S. A.and Le T. H. Nam, J. H.and Reza, D. Gutierrez, A. Jarabo, and A. Velten. Virtual wave optics for non-line-of-sight imaging. *arXiv preprint arXiv:1810.07535*, 2018.

[97] S. A. Reza, M. La Manna, S. Bauer, and A. Velten. Wave-like properties of phasor fields: experimental demonstrations. *arXiv preprint arXiv:1904.01565*, 2019.

[98] S. A. Reza, M. La Manna, and A. Velten. A physical light transport model for non-line-of-sight imaging applications. *arXiv preprint arXiv:1802.01823*, 2018.

[99] S. A. Reza, M. La Manna, and A. Velten. Imaging with phasor fields for non-line-of sight applications. In *Computational Optical Sensing and Imaging*, pages CM2E–7. Optical Society of America, 2018.

[100] F. Adib, Z. Kabelac, D. Katabi, and R. C Miller. 3d tracking via body radio reflections. In *11th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 14)*, pages 317–329, 2014.

[101] C. J Baker and S.O Piper. Continuous wave radar. *Radar, Sonar, Navigation and Avionics*, 3:17–85, 2013.

[102] D. B Lindell, G. Wetzstein, and V. Koltun. Acoustic non-line-of-sight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6780–6789, 2019.

[103] I. Dokmanić, Reza Parhizkar, A. Walther, Y. M Lu, and M. Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191, 2013.

[104] A. O'Donovan, R. Duraiswami, and J. Neumann. Microphone arrays as generalized cameras for integrated audio visual processing. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[105] Brent Schwarz. Lidar: Mapping the world in 3d. *Nature Photonics*.

[106] A. K. Pediredla, M. Buttafava, A. Tosi, O. Cossairt, and A. Veeraraghavan. Reconstructing rooms using photon echoes: A plane based model and reconstruction algorithm for looking around the corner. In *2017 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2017.

[107] F. La Manna, M.and Kine, J. Breitbach, E.and Jackson, T. Sultan, and A. Velten. Error backprojection algorithms for non-line-of-sight imaging. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1615–1626, 2018.

[108] E. Breitbach, F. Kine, M. La Manna, T. Sultan, J. Jackson, and A. Velten. Error backprojection algorithms for non-line-of-sight imaging. 2018.

[109] D. B Lindell, G. Wetzstein, and M. O'Toole. Wave-based non-line-of-sight imaging using fast fk migration. *ACM Transactions on Graphics (TOG)*, 38(4):116, 2019.

[110] R. H Stolt. Migration by fourier transform. *Geophysics*, 43(1):23–48, 1978.

[111] G. F Margrave and M. P Lamoureux. *Numerical methods of exploration seismology: with algorithms in MATLAB®*. Cambridge University Press, 2019.

[112] C. Jin, J. Xie, S. Zhang, and Y. Zhang, Z.and Zhao. Reconstruction of multiple non-line-of-sight objects using back projection based on ellipsoid mode decomposition. *Optics express*, 26(16):20089–20101, 2018.

[113] Q. Chen, S. Kumar Chamoli, P. Yin, and X. Wang, X.and Xu. Imaging of hidden object using passive mode single pixel imaging with compressive sensing. *Laser Physics Letters*, 15(12):126201, 2018.

[114] Emmanuel Candes and Justin Romberg. Sparsity and incoherence in compressive sampling. *Inverse problems*, 23(3):969, 2007.

[115] Justin Romberg. Imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[116] Kathy J Horadam. *Hadamard matrices and their applications*. Princeton university press, 2012.

[117] S. M Seitz, Y. Matsushita, and K. N Kutulakos. A theory of inverse light transport. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1440–1447. IEEE, 2005.

[118] AC Kak. Algorithms for reconstruction with nondiffracting sources. *Principles of computerized tomographic imaging*, pages 49–112, 2001.

[119] C.Y. Tsai, K. N Kutulakos, S. G Narasimhan, and A. C Sankaranarayanan. The geometry of first-returning photons for non-line-of-sight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7216–7224, 2017.

[120] D. Faccio and A. Velten. A trillion frames per second: the techniques and applications of light-in-flight photography. *Reports on Progress in Physics*, 81(10):105901, 2018.

[121] J. T Kajiya. The rendering equation. In *ACM SIGGRAPH computer graphics*, volume 20, pages 143–150. ACM, 1986.

[122] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 117–128. ACM, 2001.

[123] Rodney A Brooks, George H Weiss, and Alan J Talbert. A new approach to interpolation in computed tomography. *Journal of computer assisted tomography*, 2(5):577–585, 1978.

[124] Ori Katz, Pierre Heidmann, Mathias Fink, and Sylvain Gigan. Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations. *Nature photonics*, 8(10):784, 2014.

[125] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 572(7771):620–623, 2019.

[126] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[127] Abdul Adeel Mohammed, Rashid Minhas, QM Jonathan Wu, and Maher A Sid-Ahmed. Human face recognition based on multidimensional pca and extreme learning machine. *Pattern Recognition*, 44(10-11):2588–2597, 2011.

[128] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

[129] Adi L Tarca, Vincent J Carey, Xue-wen Chen, Roberto Romero, and Sorin Drăghici. Machine learning and its applications to biology. *PLoS computational biology*, 3(6):e116, 2007.

[130] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[131] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447, 2018.

[132] Tim Byrnes, Shinsuke Koyama, Kai Yan, and Yoshihisa Yamamoto. Neural networks using two-component bose-einstein condensates. *Scientific reports*, 3:2531, 2013.

[133] Navid Borhani, Eirini Kakkava, Christophe Moser, and Demetri Psaltis. Learning to see through multimode fibers. *Optica*, 5(8):960–966, 2018.

[134] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[135] Keras. Kears documentation, website=https://keras.io/activations/. 2019.

[136] Micheal A. Nielsen. Neural network and deep learning. 2015.

[137] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[138] Zhanyi Wang. The applications of deep learning on traffic identification. *BlackHat USA*, 24, 2015.

[139] Li Deng, Geoffrey Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: An overview. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8599–8603. IEEE, 2013.

[140] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, et al. Deep face recognition. In *bmvc*, volume 1, page 6, 2015.

[141] Maxwell W Libbrecht and William Stafford Noble. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6):321–332, 2015.

[142] Christof Angermueller, Tanel Pärnamaa, Leopold Parts, and Oliver Stegle. Deep learning for computational biology. *Molecular systems biology*, 12(7), 2016.

[143] Polina Mamoshina, Armando Vieira, Evgeny Putin, and Alex Zhavoronkov. Applications of deep learning in biomedicine. *Molecular pharmaceutics*, 13(5):1445–1454, 2016.

[144] Laura Waller and Lei Tian. Computational imaging: Machine learning for 3d microscopy. *Nature*, 523(7561):416, 2015.

[145] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017.

[146] Fuyong Xing, Yuanpu Xie, Hai Su, Fujun Liu, and Lin Yang. Deep learning in microscopy image analysis: A survey. *IEEE transactions on neural networks and learning systems*, 29(10):4550–4568, 2017.

[147] Oren Z Kraus, Jimmy Lei Ba, and Brendan J Frey. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics*, 32(12):i52–i59, 2016.

[148] Andrew Schutter and Lior Shamir. Galaxy morphology, an unsupervised machine learning approach. *Astronomy and Computing*, 12:60–66, 2015.

[149] James Riden. *Unsupervised learning on galaxy spectra*. PhD thesis, Citeseer, 2002.

[150] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for exotic particles in high-energy physics with deep learning. *Nature communications*, 5:4308, 2014.

[151] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep learning and its application to lhc physics. *Annual Review of Nuclear and Particle Science*, 68:161–181, 2018.

[152] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Enhanced higgs boson to τ+ τ- search with deep learning. *Physical review letters*, 114(11):111801, 2015.

[153] Pierre Chiappetta, Pietro Colangelo, P De Felice, Giuseppe Nardulli, and Guido Pasquariello. Higgs search by neural networks at lhc. *Physics Letters B*, 322(3):219–223, 1994.

[154] B Todd Huffman, Thomas Russell, and Jeff Tseng. Tagging *b* quarks without tracks using an artificial neural network algorithm. *arXiv preprint arXiv:1701.06832*, 2017.

[155] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Accelerating science with generative adversarial networks: an application to 3d particle showers in multilayer calorimeters. *Physical review letters*, 120(4):042003, 2018.

[156] M Aaboud, Georges Aad, Brad Abbott, Jalal Abdallah, Ovsat Abdinov, Baptiste Abeloos, Syed Haider Abidi, OS AbouZeid, NL Abraham, Halina Abramowicz, et al. Performance of the atlas track reconstruction algorithms in dense environments in lhc run 2. *The European Physical Journal C*, 77(10):673, 2017.

[157] Daniel George and EA Huerta. Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data. *Physics Letters B*, 778:64–70, 2018.

[158] Massimiliano Razzano and Elena Cuoco. Image-based deep learning for classification of noise transients in gravitational wave detectors. *Classical and Quantum Gravity*, 35(9):095016, 2018.

[159] Hongyu Shen, Daniel George, EA Huerta, and Zhizhen Zhao. Denoising gravitational waves using deep learning with recurrent denoising autoencoders. *arXiv preprint arXiv:1711.09919*, 2017.

[160] George Barbastathis, Aydogan Ozcan, and Guohai Situ. On the use of deep learning for computational imaging. *Optica*, 6(8):921–943, 2019.

[161] Katherine L Bouman, Vickie Ye, Adam B Yedidia, Fredo Durand, Gregory W Wornell, Antonio Torralba, and William T Freeman. Turning corners into cameras: Principles and methods. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2270–2278, 2017.

[162] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4778–4787, 2017.

[163] Ayan Sinha, Justin Lee, Shuai Li, and George Barbastathis. Lensless computational imaging through deep learning. *Optica*, 4(9):1117–1125, 2017.

[164] Alexandre Goy, Kwabena Arthur, Shuai Li, and George Barbastathis. Low photon count phase retrieval using deep learning. *Physical review letters*, 121(24):243902, 2018.

[165] Meng Lyu, Wei Wang, Hao Wang, Haichao Wang, Guowei Li, Ni Chen, and Guohai Situ. Deep-learning-based ghost imaging. *Scientific reports*, 7(1):17865, 2017.

[166] Matthew Tancik, Tristan Swedish, Guy Satat, and Ramesh Raskar. Data-driven non-line-of-sight imaging with a traditional camera. In *Imaging Systems and Applications*, pages IW2B–6. Optical Society of America, 2018.

[167] Sreenithy Chandran and Suren Jayasuriya. Adaptive lighting for data-driven non-line-of-sight 3d localization and object identification. *arXiv preprint arXiv:1905.11595*, 2019.

[168] Matthew Tancik, Guy Satat, and Ramesh Raskar. Flash photography for data-driven hidden scene recovery. *arXiv preprint arXiv:1810.11710*, 2018.

[169] Shuai Li, Mo Deng, Justin Lee, Ayan Sinha, and George Barbastathis. Imaging through glass diffusers using densely connected convolutional networks. *Optica*, 5(7):803–813, 2018.

[170] Meng Lyu, Hao Wang, Guowei Li, and Guohai Situ. Exploit imaging through opaque wall via deep learning. *arXiv preprint arXiv:1708.07881*, 2017.

[171] Guy Satat, Matthew Tancik, Otkrist Gupta, Barmak Heshmat, and Ramesh Raskar. Object classification through scattering media with deep learning on time resolved measurement. *Optics express*, 25(15):17466–17479, 2017.

[172] Piergiorgio Caramazza, Oisín Moran, Roderick Murray-Smith, and Daniele Faccio. Transmission of natural scene images through a multimode fibre. *Nature communications*, 10(1):2029, 2019.

[173] Alex Turpin. Projecting light through complex media with machine learning. In *Computational Optical Sensing and Imaging*, pages CTu3A–5. Optical Society of America, 2019.

[174] Claus-Christian Carbon and Vera M Hesslinger. Da vinci's mona lisa entering the next dimension. *Perception*, 42(8):887–893, 2013.

[175] Bahram Javidi and Fumio Okano. *Three-dimensional television, video, and display technologies*. Springer Science & Business Media, 2002.

[176] Jayaram K Udupa and Gabor T Herman. *3D imaging in medicine*. CRC press, 1999.

[177] Navid Farahani, Alex Braun, Dylan Jutt, Todd Huffman, Nick Reder, Zheng Liu, Yukako Yagi, and Liron Pantanowitz. Three-dimensional imaging and scanning: current and future applications for pathology. *Journal of pathology informatics*, 8, 2017.

[178] Matthew D Hammers, Michael J Taormina, Matthew M Cerda, Leticia A Montoya, Daniel T Seidenkranz, Raghuveer Parthasarathy, and Michael D Pluth. A bright fluorescent probe for h2s enables analyte-responsive, 3d imaging in live zebrafish using light sheet fluorescence microscopy. *Journal of the American Chemical Society*, 137(32):10216–10223, 2015.

[179] Raffaele Schiavullo. Mathematical approach and detecting devices for automotive lidar. 2018.

[180] Stewart Wills. Auto lidar: Optical choices and challenges. 2019.

[181] Neal Radwell, Adam Selyem, Lena Mertens, Matthew P Edgar, and Miles J Padgett. Hybrid 3d ranging and velocity tracking system combining multi-view cameras and simple lidar. *Scientific reports*, 9(1):5241, 2019.

[182] Kartik Venkataraman. Systems and methods for 3d facial modeling, May 30 2019. US Patent App. 15/823,473.

[183] Marcella Peter, Jacey-Lynn Minoi, and Irwandi Hipni Mohamad Hipiny. 3d face recognition using kernel-based pca approach. In *Computational Science and Technology*, pages 77–86. Springer, 2019.

[184] Stephen T Barnard and Martin A Fischler. Computational stereo. Technical report, SRI INTERNATIONAL MENLO PARK CA ARTIFICIAL INTELLIGENCE CENTER, 1982.

[185] Uwe Franke and Armin Joos. Real-time stereo vision for urban traffic scene understanding. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511)*, pages 273–278. IEEE, 2000.

[186] Stephen A Benton, HJ Caulfield, and William T Rhodes. Special problems. In *Handbook of Optical Holography*, pages 349–378. Academic Press, 1979.

[187] Vincent Toal. *Introduction to holography*. CRC press, 2011.

[188] NJ Phillips and D Porter. An advance in the processing of holograms. *Journal of Physics E: Scientific Instruments*, 9(8):631, 1976.

[189] A Graube. Advances in bleaching methods for photographically recorded holograms. *Applied optics*, 13(12):2942–2946, 1974.

[190] Ulf Schnars. Direct phase determination in hologram interferometry with use of digitally recorded holograms. *JOSA A*, 11(7):2011–2015, 1994.

[191] Ulf Schnars, Claas Falldorf, John Watson, and Werner Jüptner. Digital holography. In *Digital Holography and Wavefront Sensing*, pages 39–68. Springer, 2015.

[192] S Burak Gokturk, Hakan Yalcin, and Cyrus Bamji. A time-of-flight depth sensor-system description, issues and solutions. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 35–35. IEEE, 2004.

[193] François Blais. Review of 20 years of range sensor development. In *Videometrics VII*, volume 5013, pages 62–76. International Society for Optics and Photonics, 2003.

[194] Sergi Foix Salmerón, Guillem Alenyà Ribas, and Carme Torras. Exploitation of time-of-flight (tof) cameras. 2010.

[195] Aongus McCarthy, Robert J Collins, Nils J Krichel, Verónica Fernández, Andrew M Wallace, and Gerald S Buller. Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting. *Applied optics*, 48(32):6241–6251, 2009.

[196] Dipl-Ing Bianca Hagebeuker and Product Marketing. A 3d time of flight camera for object detection. *PMD Technologies GmbH, Siegen*, 2007.

[197] Gregory A Howland, P Ben Dixon, and John C Howell. Photon-counting compressive sensing laser radar for 3d imaging. *Applied optics*, 50(31):5917–5920, 2011.

[198] Martin Laurenzis, Frank Christnacher, and David Monnin. Long-range three-dimensional active imaging with superresolution depth mapping. *Optics letters*, 32(21):3146–3148, 2007.

[199] Kenneth David Mankoff and Tess Alethea Russo. The kinect: A low-cost, high-resolution, short-range 3d camera. *Earth Surface Processes and Landforms*, 38(9):926–936, 2013.

[200] Wenlin Gong, Chengqiang Zhao, Hong Yu, Mingliang Chen, Wendong Xu, and Shensheng Han. Three-dimensional ghost imaging lidar via sparsity constraint. *Scientific reports*, 6:26133, 2016.

[201] Baoqing Sun, Matthew P Edgar, Richard Bowman, Liberty E Vittert, Stuart Welsh, A Bowman, and MJ Padgett. 3d computational imaging with single-pixel detectors. *Science*, 340(6134):844–847, 2013.

[202] Sigurjón Árni Guðmundsson, Henrik Aanaes, and Rasmus Larsen. Fusion of stereo vision and time-of-flight imaging for improved 3d estimation. *International Journal on Intelligent Systems Technologies and Applications (IJISTA)*, 5(3/4):425–433, 2008.

[203] Kyoung Won Nam, Jeongyun Park, In Young Kim, and Kwang Gi Kim. Application of stereo-imaging technology to medical field. *Healthcare informatics research*, 18(3):158–163, 2012.

[204] JF Cardenas-Garcia, HG Yao, and S Zheng. 3d reconstruction of objects using stereo imaging. *Optics and Lasers in Engineering*, 22(3):193–213, 1995.

[205] Tong Zhang and Ichirou Yamaguchi. Three-dimensional microscopy with phase-shifting digital holography. *Optics letters*, 23(15):1221–1223, 1998.

[206] Giancarlo Pedrini, Staffan Schedin, and Hans J Tiziani. Lensless digital-holographic interferometry for the measurement of large objects. *Optics communications*, 171(1-3):29–36, 1999.

[207] Pinliang Dong and Qi Chen. *LiDAR remote sensing and applications*. CRC Press, 2017.

[208] M. Schmidt, K. Zimmermann, and B. Jahne. High frame rate for 3d time-of-flight cameras by dynamic sensor calibration. In *2011 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8, April 2011.

[209] Hongda Wang, Yair Rivenson, Yiyin Jin, Zhensong Wei, Ronald Gao, Harun Günaydın, Laurent A Bentolila, Comert Kural, and Aydogan Ozcan. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods*, 16:103–110, 2019.

[210] TensorFlow. Tensorflow. documentation, website=https://www.tensorflow.org/. 2019.