POLITECNICO DI TORINO
Repository ISTITUZIONALE

Analysis of RGB-D camera technologies for supporting different facial usage scenarios

Original
Analysis of RGB-D camera technologies for supporting different facial usage scenarios / Ulrich, L.; Vezzetti, E.; Moos, S.; Marcolin, F.. - In: MULTIMEDIA TOOLS AND APPLICATIONS. - ISSN 1380-7501. - (2020). [10.1007/s11042-020-09479-0]

Availability:
This version is available at: 11583/2842766 since: 2020-08-19T17:47:46Z

Publisher:
Springer

Published
DOI:10.1007/s11042-020-09479-0

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

18 April 2024

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09479-0

Analysis of RGB-D camera technologies for supporting different facial usage scenarios

Luca Ulrich[1] · Enrico Vezzetti[1] · Sandro Moos[1] · Federica Marcolin[1]

Received: 6 August 2019 / Revised: 19 July 2020 / Accepted: 28 July 2020
Published online: 11 August 2020
© The Author(s) 2020

Abstract
Recently a wide variety of applications has been developed integrating 3D functionalities. Advantages given by the possibility of relying on depth information allows the developers to design new algorithms and to improve the existing ones. In particular, for what concerns face morphology, 3D has led to the possibility to obtain face depth maps highly close to reality and consequently an improvement of the starting point for further analysis such as Face Detection, Face Authentication, Face Identification and Face Expression Recognition. The development of the aforementioned applications would have been impossible without the progress of sensor technologies for obtaining 3D information. Several solutions have been adopted over time. In this paper, emphasis is put on passive stereoscopy, structured light, time-of-flight (ToF) and active stereoscopy, namely the most used technologies for the cameras design and fulfilment according to the literature. The aim of this article is to investigate facial applications and to examine 3D camera technologies to suggest some guidelines for addressing the correct choice of a 3D sensor according to the application that has to be developed.

Keywords Sensors · 3D cameras · 3D face analysis · Stereoscopy · ToF · Structured light

1 Introduction

In recent years a considerable number of applications have benefited from the usage of the third dimension [47]; there are several research fields in which 3D is currently successfully used: safety, such as for autonomous driving [22]; orthopedics, for both diagnosis and treatment planning [106]; surgery, as 3D models reconstruction gives the possibility of

✉ Enrico Vezzetti
enrico.vezzetti@polito.it

Luca Ulrich
luca.ulrich@polito.it

[1] DIGEP, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

organizing medical equipment [99], attending the surgeon during the intervention and supporting the post-operative evaluation of the results [70]; 3D printing applications [21], including facial prosthesis [67], dental implants [30] and pelvis prosthesis [13].

The ambition of accelerating the evolution process of cities into interconnected communities brings out other application areas as candidates for heavy 3D usage: land surveying [90], architecture [59], archaeology [42] for research and tourism purposes and also security. Smart cities, urban area equipped with interconnected sensors able to collect data to be used to manage products and services [3], aim to benefit of the spreading of face recognition technology and deep learning techniques to solve problems such as quickly finding missing children and identifying criminals [105] or monitoring public places such as airports [83]. Geometry of the surfaces acquired with sensors capable of capturing depth information can be used for a more accurate face reconstruction [110], to build 3D aging models [82], face manipulation [44] and landmarking [31]. As it will be better explained in the next section, facial applications are shiny examples of this consideration, since the face acquisition can be performed in different conditions depending on various usage scenarios.

3D techniques require a higher computational cost than 2D methods [1], especially if the 3D face model has to be reconstructed from multi-view images [18] or through 3D morphable models obtained from 2D images and 3D scans or even without 3D data [98]. Nonetheless, the robustness given by the opportunity to operate in critical lighting conditions [109], in presence of occlusions [103, 28] and regardless of the orientation of the subject [102] make a 3D approach preferable.

Literature about 3D is varied and fragmented due to lack of a shared methodology for analyzing the field and developing new applications in the face of a growing number of RGB-D cameras on the market. This scientific survey has been conducted to converge on a unique standard and to provide a baseline for the design of the following 3D facial applications in real-time: Face Detection, Face Authentication, Face Identification and Face Expression Recognition.

Total time required by a facial application to be performed is the sum of the acquisition time and the processing time. The first one is the time required to obtain the RGB-D information and depends on 3D cameras; the second one involves the processing of the acquired depth information that is necessary to obtain a result. Since the latter does not depend on 3D cameras, but on other elements of the framework which constitutes an application, for instance the face analysis algorithms, it has not been analyzed in the present work, which focuses on acquisition hardware.

This work aims to be a guide for the right choice of an RGB-D camera depending on the facial application that has to be implemented. The focus is on camera technologies able to provide RGB images and depth maps, namely images on which each pixel has a value representing the distance from the camera; 3D scanners have not been taken into consideration, because they do not work in real-time, since they require a minimum technical time to complete the scan.

The study is structured as follows. Section 2 focuses on facial applications and on 3D sensor technologies; an explanation of the methodologies used for the investigation is provided. Survey results are presented in the third section, while in the final section conclusions have been drawn.

2 Methodological analysis

This survey has been carried out through a two steps analysis. First, a desk research has been performed to qualitatively investigate two aspects of 3D: the available technologies for computing

the depth and the facial applications able to benefit from 3D usage that have been developed up to now. A desk research is a complete review of the literature, including articles and datasheets, indispensable to deeply analyze the functioning and the potentialities of the 3D sensors [100]. Secondly, QFDs (Quality Function Deployments) [66] have been used to quantitatively examine the relationships between two different orthogonal dimensions, namely the qualitative requirements typical of each facial application and the technical specifications of 3D acquisition technologies. Both dimensions have been obtained from the results of the desk research.

2.1 Desk research on facial applications

The opportunity of understanding and extracting information from human face has interested many researchers in past decades, giving birth to a new discipline called "Face Perception" [107]. Human brain has the ability of figuring out characteristics such as identity, age, sex and mood [15, 65, 29], a skill that infants already possess from birth [69], and develops during growth [40].

Since the recent spread of Computer Vision outcomes have highlighted that the utilization of technologies able to emulate human behavior is desirable, the idea of automatizing the face perception process has come up. Nevertheless, human brain functioning is highly complex and nowadays the possibility of reconstructing a model able to replay its behavior is remote. That explains why, in literature, all the applications related to the automatic recognition of specific features on human face are studied individually.

In this paper, the main facial applications have been considered: Face Detection, Face Recognition, with the two declinations in Face Authentication and Face Identification, and Face Expression Recognition (Fig. 1).

2.1.1 Face detection

Face Detection [68] aims to detect a face shape inside an image or inside a frame in the case of a video stream. It is often used to crop the image for further processing, typically another facial application, so that algorithms can focus on the region of interest; nonetheless, Face Detection could be used stand-alone in applications such as counting number of people in a room [111], the automatic selection of a region of interest containing a face to insert a tag on it (like on Facebook), to avoid gatherings or otherwise monitor the crowd [61].

There exists various 2D techniques [46, 108], that aim to achieve a good trade-off in terms of accuracy and speed. Some common operations are to localize and discard the background for improving computational speed focusing on the area of the image that carries the relevant information, normalizing the image with rotation and scaling operations not to reject false negatives and finally extracting the facial features necessary for Face Detection [86].

3D techniques benefit of the intrinsic advantages in using the depth information, such as lighting, pose and occlusion independence, to perform the detection through the analysis of surface curvature [27] or other geometrical features [63]. This is the family of methods considered in this paper for the sake of robustness, an essential characteristic for real-time video data streams, thus sensors should provide high quality depth maps in a wide range of functioning.

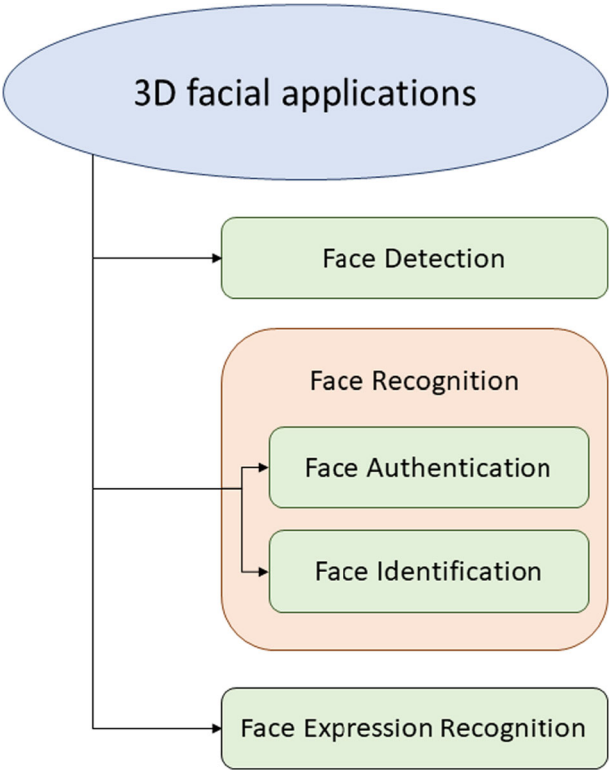

Fig. 1 3D facial applications considered in the present work: Face Detection [86, 27, 46, 63], Face Authentication [41], Face Identification [79, 17, 49] and Face Expression Recognition [36, 37]

2.1.2 Face authentication

A taxonomy clarification is mandatory to deepen the discussion about Face Recognition applications. Face recognition aims to recognize a face detected into an image or into a frame, comparing it with another face or with a set of faces contained into a database.

Face Authentication belongs to biometric systems, that are solutions implemented to control the access to a private area using specific features of individuals [7] and it is not uncommon that information obtained with different biometric systems are often fused together to further improve robustness [85].

Fingerprints [57] and iris scans [43] are two of the most famous biometric systems for recognizing a person, but Face Authentication is becoming a more and more common solution in the case of identity certification for personal devices, especially for laptops and smartphones [41], and to fulfil payments. The high degree of security requested to protect a personal device implicates the need of a great deal of skill in the recognizing process and consequently 3D cameras must provide the best images possible in terms of quality, so that the facial authentication algorithm can minimize false positives and false negatives having as much features as possible retrievable from depth maps provided as input data.

In last years, the spread of personal mobile devices equipped with RGB-D cameras has been the cause of increased usage of face to perform user authentication, to such an extent that a new taxonomy has been forged, the selfie biometrics [81].

2.1.3 Face identification

In this article, Face Identification refers to that variety of applications performing Face Recognition [88] without authentication purposes described in the previous section. Some examples can be found in the fields of security, for criminals identification [79], marketing, to target specific customers or at least some of their features such as age and gender [17], and healthcare, for a health monitoring through a comparison between the current status of a patient and an image of the same patient in good health [49]. Some of the applications benefits of the technology development in terms of portability to recognize criminals [38], patients [32] or other individuals [80].

In Face Identification applications images or frames must be accurate enough to compare different facial features, the result must be provided in a reasonable time, the frame rate should be sufficiently high to detect all the people in the camera field of view (FOV), especially those ones in motion, and an adequately wide operating range to accomplish the target if working on a video data flow.

Face recognition algorithms working with 2D data must be carefully used stand-alone due to their vulnerabilities to spoofing attacks. Indeed, some other methods as liveness detection must be added to obtain a reliable face recognition technique. Furthermore, the technological improvement has made 3D data usage promising since depth map details are more and more refined and robust to spoofing attacks [2].

2.1.4 Face expression recognition

Face Expression Recognition [4] aims to identify the face within a frame and understand humans emotion by observing different parts of the face and analyzing the Action Units proposed by Paul Ekman in his works [36, 37].

The need of such an application is due to the spread of the concept of human-computer interaction in a variety of fields [9]: marketing [91], smart TV [62], videogames [64], psychiatry [16], evaluation of users' engagement [73, 72]. One of the most important fields of application is robotics, since the capability to automatically understand human's mood [11] significantly improve human operator safety during the interaction.

Face Expression Recognition is a critical task since some expressions are ambiguous and difficult to be recognized even by a human observer. Geometrical analysis is the basis of this application, so input images quality should be detailed enough to identify them and to perform further analysis. Recent researches show how landmarks and facial units can be the starting point to detect facial emotions [51] as well as geometrical descriptors can be used as input information to feed a CNN [74].

2.2 Desk research on RGB-D camera technologies

The interest in the applications mentioned above has received a further impulse since the advent of low-cost 3D sensors, i.e. devices able to detect the third dimension. The Microsoft Kinect release on the market in 2010 is one of the milestones related to the diffusion of these devices. This sensor has been designed and developed for the specific purpose of recognizing human body actions to perform an original type of human-machine interaction aimed at controlling characters, vehicles, or whatever object movements inside a videogame.

Several types of 3D sensors have been released on the market during last years and technology is the most suitable characteristic for grouping up sensors according to the similarity of their main parameters (Fig. 2).

All the 3D sensors mentioned above are also known as RGB-D cameras, because they provide two types of data: RGB and D (depth). RGB refers to the color model thanks to which every color can be displayed using three primary color red, green and blue; in other words, it identifies the color images. Depth information is retrievable through depth maps, images on which each pixel has a value representing the distance from the camera. This type of data is an advance compared to 2D data in terms of reliability and suitable for real-time applications. Indeed, it is possible to analyze the depth map without building a mesh; every 3D object is identified with x, y coordinates and the depth value instead of set of vertices, edges, and faces. The result is a more responsive acquisition system at the cost of accuracy. The present work focuses on 3D sensors because it is necessary to understand which technology can preserve high quality depth data working in real-time. This is due to the focus on technologies and data which will be largely adopted in the near future, when the accuracy of the third dimension will be exploited for several purposes and analyzing data real-time will be core for most of the acquisition systems [24].

Some of the applications mentioned above can have a considerable computational cost; nonetheless, 3D cameras and the devices that potentially can integrate them must be able to acquire information in real-time but the processing can be performed by systems located remotely. This solution can be planned at designed time before implementing a facial application, allowing not to be constrained by device capabilities in terms of processing, although they still must guarantee to maintain the 3D camera frame rate and to be connected with the remote system.

The way each technology provides the depth map is described in the following paragraphs.

2.2.1 Passive stereoscopy

Passive stereo requires the presence of at least two cameras for acquiring different images of the same object or environment from different points of view [93, 20, 35, 71, 84, 34].

To understand the distance of each point detected by this type of camera, the triangulation (or computational stereopsis) process must be performed, solving the so-called correspondence

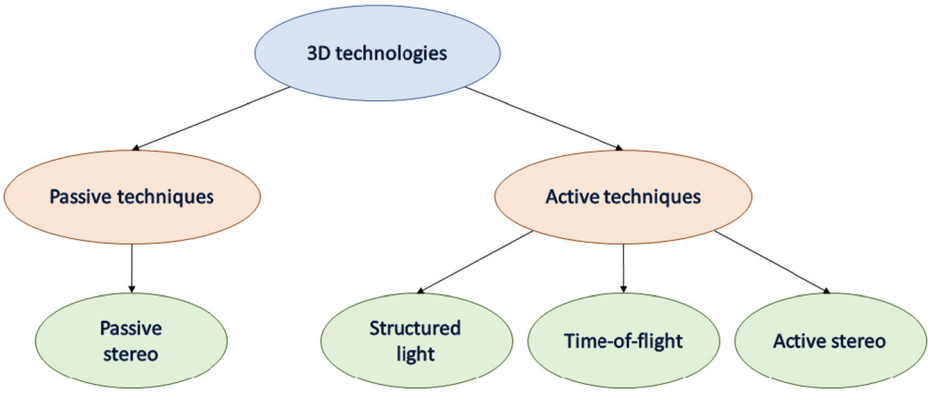

Fig. 2 3D Technologies [14]

problem. Given the camera parameters calibration, the conjugate points, i.e. the two pixels representing the same point on the scene that are positioned on the two different acquired frames, must be found.

The main drawback of stereo cameras is the need of a scene lacking occlusions, therefore the shape of the object can be detected from both the cameras, and this is not trivial, since the object geometry can be complex enough that some parts are visible from a camera and hidden to the other one, such as alae, namely the two points that lie on the right and on the left of the nose and are commonly considered the landmarks for computing nose width [101]. In addition, the scene must not be featureless since the correspondence problem can be solved only if the same features can be found by both the cameras.

Price of these cameras can vary from 150 $ to 700 $ depending both on the features and the release on the market time.

2.2.2 Structured light

Structured light depth cameras have been studied to overcome the issue of reliability of correspondences [54, 56, 6, 8, 39, 75, 77, 95]. If there are two or more cameras filming an object, however close they may be, they will frame different parts of the object and not all the points of the object will be visible from all the cameras. Furthermore, if cameras are too close to each other, disparity will not be large enough to make the triangulation process possible.

The technology consists in projecting a pattern on the object using a transmitter and, successively, evaluate the deformation of the pattern on the object detected by a receiver. This solution allows to put transmitter and receiver close each other, since the distance is computed without the need of the disparity and consequently the occlusions issue is minimized.

The projected pattern can assume different configurations to perform the correspondences estimation according to design concepts. Adopted strategies are wavelength multiplexing, range multiplexing, temporal multiplexing, and spatial multiplexing [87].

This type of camera can be considered quite cheap compared to the other technologies: price is usually not higher than 200 $ with a few exceptions.

2.2.3 Time-of-flight (ToF)

ToF cameras have been considered only professional-grade until Microsoft released the second version of the Kinect, commonly mentioned as Kinect v2 or Kinect One, since it has been developed for being used with the Microsoft X-Box One console, contrary to the Kinect v1 developed for X-Box 360.

This technology relies on the knowledge of the light speed in the air. Distances can be evaluated projecting an electromagnetic wave on the scene and computing the time in which it has been received from the receiver.

A remarkable advantage of this technology is the opportunity to put transmitter and receiver closer than the transmitter and the receiver needed for structured light depth cameras. Moreover, ToF sensors can reach considerable frame rate, making them suitable for real-time applications [92, 50, 89, 10, 96, 97, 33].

On average, ToF cameras are the most expensive on the market since they were born for industrial applications. Nonetheless prices cover a very wide range: from 80 $ to thousands of dollars.

2.2.4 Active stereoscopy

Active stereo is a vision technique in which stereo and structured light, or laser, are combined to benefit of the advantages of both the technologies [19, 55, 52, 53]. A 3D sensor built according to this technology is equipped with two outdistanced cameras and a projector between them, usually working in IR spectrum. This solution allows to improve accuracy in 3D detection and, above all, permits to extend the operating range [12].

Active stereoscopy cameras are peculiar of Intel which proposes them at a cost between 130 $ and 400 $. Most recent devices cost 150 $ - 200 $.

2.3 Benchmarking

A benchmarking among 3D sensor technologies has been done evaluating the parameters available both in literature and in datasheets. Parameters taken into consideration are:

- Resolution: horizontal and vertical number of pixels
- Frame rate: number of images captured in one second (FPS, Frames Per Second)
- Minimum distance: this parameter establishes the lowest gap for sensor functioning
- Maximum distance: this parameter establishes the greatest gap for sensor functioning
- Range: difference between minimum distance and maximum distance
- Field of view (FOV): this parameter indicates the part of the scene visible through the sensor
- Size: sensors dimensions.

Twenty-six sensors belonging to the four categories explained above have been analyzed to identify strengths and weaknesses of each 3D detection technology (Table 1).

2.3.1 Passive stereoscopy

Six passive stereo sensors have been considered (Table 2). Stereo cameras have quite good ranges of functioning, thanks to good maximum distance values that make most of them suitable for acquisition over 3 m of distance, but a bad minimum distance of functioning. Values regarding minimum distance of functioning reported in this work, directly taken from sensor datasheets too, are often misleading. That value means that it is possible to acquire the depth map, but its quality is very poor, especially in the case of facial application. This is a

Table 1 RGB-D cameras considered in this work

Technology	Cameras
Passive stereoscopy	Stereolabs ZED, Carnegie Robotics MultiSense S7, E-Con Tara – Stereo Vision Camera, Nerian SPI, Roboception rc_visard, DUO 3D Sensor.
Structured light	Intel RealSense F200, Intel RealSense SR300, Microsoft Kinect v1, Asus Xtion PRO LIVE, Ensenso N35–606-16-BL, Orbecc Astra Mini, Photoneo PhoXi 3D Scanner L, Structure Sensor.
Time-of-Flight	Microsoft Kinect v2, IFM O3D303, SICK Visionary-T, Basler ToF camera, PMD CamCube 3.0, MESA SR 4000, MESA SR 4500, Soft Kinect ds325.
Active stereoscopy	Intel RealSense R200, Intel RealSense D415, Intel RealSense D435 (D435i), Intel RealSense Euclid.

technological problem: passive stereoscopy uses disparity between two cameras to retrieve the depth information. If the camera is close to the subject, a lot of points will be present in only one of the images due to occlusions, making them very difficult to merge. Resulting depth images contain too big holes, which make data impossible to use. In particular, a second minimum value is often shown in datasheets and it points out the optimal minimum distance that is usually greater than 50 cm.

On the contrary, resolution is excellent, while frame rate has quite different nominal values (3, 15, 30, 45 FPS).

2.3.2 Structured light

Eight structured light sensors have been analyzed (Table 3). Minimum distance is undoubtedly the strength for this technology, in fact several sensors minimum operating distance is between 20 cm and 40 cm. Frame rate is remarkable too, almost all sensors work at 30 or 60 FPS. Maximum distance and range operating functioning are the weaknesses of this technology, since most of sensors work with an upper limit that is suggested from 1.5 m to 2.5 m. Resolution is remarkable for short range, since only one sensor is 320 × 240 and the others are 640 × 480 or above.

2.3.3 Time-of-flight

Among the 8 ToF sensors analyzed (Table 4), just one of them can be considered suitable for facial applications. Other sensors belonging to this category have a magnificent maximum distance (at least greater than 4 m), a decent frame rate (20–30 FPs), but poor minimum distance (0.5 m) and resolution (640 × 480 is the only remarkable value, all the others are below).

Values in the table are strongly influenced by a single sensor build with the specific purpose of working at close distance. This is the reason why the comments reported above are very important to understand the considerations drawn up in the “Results and discussion” section.

2.3.4 Active stereoscopy

The four active stereoscopy sensors considered (Table 5) are the most recent on the market, launched in 2015 or later.

Table 2 Passive stereoscopy sensors specs

Passive stereoscopy		
Parameter	Best	Worst
Frame rate	60 FPS	3 FPS
Maximum distance	15 m	1 m
Minimum distance	0.23 m	0.5 m
Range	14 m	0.3 m
Resolution	2208 × 1242	640 × 480
Size	57 × 30.5 × 14.7 mm	230x75x84 mm

Table 3 Structured light sensors specs

Structured light		
Parameter	Best	Worst
Frame rate	60 FPS	30 FPS
Maximum distance	5 m	0.49 m
Minimum distance	0.2 m	0.87 m
Range	4.4 m	0.23 m
Resolution	1280 × 1024	320 × 240
Size	80x20x20 mm	279.4 × 71.12 × 66.04 mm

They can be considered the best trade-off between all the parameters, with good minimum distance (around 30 cm except for the worst one), maximum distance (up to 10 m), 30 FPS frame rate and good resolution (two of them reach 1280 × 800).

A special mention is deserved by the best minimum distance found during the desk research (0.11 m), but all the others functioning minimum distance exceed 0.3 m, so structured light sensors must be considered as the state-of-art for minimum distance of functioning yet.

Sensors datasheets report the size including the chassis and the support dimensions. Customer-grade sensors can be integrated in personal devices such as smartphones, tablets and laptops without chassis and support, therefore it is desirable to understand the physical space that each technology requires. Passive stereo and active stereo need a larger space due to the presence of two different cameras for detecting the third dimensions through the disparity, while for what concerns structured light and ToF technologies size can be limited by the possibility of putting transmitter and receiver as close as possible.

A brief recap of main advantages and disadvantages for each technology can be found in Table 6.

2.4 Quality function deployment (QFD)

Once the desk research has been completed, the QFD has been used to integrate two orthogonal dimensions, namely sensors' technical specifications and facial applications requirements. The aim of this stage is to identify their interconnections evaluating how much each technical specification is important in relation to a certain application requirement.

QFD [66] is a method applied to transform qualitative user demands into quantitative parameters and the basic design to implement it is the house of quality. On the vertical axis there are the user desires (What's), on the horizontal axis there are technical requirements

Table 4 ToF sensors specs

ToF		
Parameter	Best	Worst
Frame rate	30 FPS	20 FPS
Maximum distance	60 m	1 m
Minimum distance	0.15 m	0.8 m
Range	59.5 m	0.85 m
Resolution	640 × 480	176 × 144
Size	65x65x68 mm	250x70x45 mm

Table 5 Active stereoscopy sensors specs

Active stereoscopy		
Parameter	Best	Worst
Frame rate	30 FPS	30 FPS
Maximum distance	10 m	2.8 m
Minimum distance	0.11 m	0.4 m
Range	10 m	2.25 m
Resolution	1280 × 800	640 × 480
Size	90x25x25 mm	126x254x345 mm

(How's) that may be useful to satisfy the user desires. A weight between 1 and 5 is given to each user's desire according to the final application that has to be designed. In the other cells of the table a score of 1, 3 or 9 [26] is given according to the contribution that each technical requirement gives to each user desire, namely respectively "weak", "moderate" and "strong". 0 value has been given if there is no relationship. Scores to be attributed to the relationships can vary according to different ways of building a QFD [58]. In this case, 0, 1, 3, and 9 have been considered because they reflect at best the perception that people have with regard to the correlation process and strong correlation is awarded.

Four QFDs have been drawn up, one for each facial application previously explained, and they are structured as follows: qualitative application requirements, namely the main characteristics that an application should have, are listed on the first column and the importance of each qualitative requirement is listed on the second column. On the first row there are the technical specifications (How's), and contrary to the qualitative requirements, that are slightly different between the applications, the technical requirement list is the same for each of the four QFDs.

The considered technical specifications are the depth sensors parameters extracted by the desk research. Specifically, technical requirements are the frame rate, the minimum and the maximum distance to which the sensors work, the range, the FOV, the dimensions and the technology used to build the sensor. A little observation for the resolution is important; if its value is high, this means not only that there are more pixels on the same image, and consequently a higher accuracy, but there is also the possibility of performing a downgrade

Table 6 Advantages and disadvantages of analyzed technologies

Technology	Advantages	Disadvantages
Passive stereoscopy	• Good operative range • Good resolution	• Difficulties in discriminating between minimum distance and optimal minimum distance of functioning • Scene not featureless • Problems with occlusions
Structured light	• Good minimum distance of functioning • Good frame rate	• Low maximum distance of functioning • Low range of functioning
Time-of-Flight	• Good maximum distance of functioning	• Poor minimum distance of functioning
Active stereoscopy	• Good minimum distance of functioning • Good maximum distance of functioning	• Small problems with occlusions

of the resolution to speed up the frame rate for those applications in which real-time is a critical task.

In the final row the relative total score of each technical specification is specified. Relative total score is a percentage of how much a technical requirement is important compared to the others. Its values are computed as follows:

1 For each technical requirement, a total score is computed as a sum of products between the application requirement weights and the corresponding evaluation scores given to the technical requirements.
2 For each total score obtained at point 1 the percentage is computed considering the sum of all the total scores as 100%.

3 Results and discussion

Generical raw data have been translated into values to be put in QFDs (9–3–1-0 score) after a discussion held by a focus group. The focus group has proved to be essential to accurately evaluate technical requirements thanks to the involvement of researchers from several areas and is composed by eleven people, five women and six men: four of them are computer science engineers, and their research field involve computer vision and RGB-D cameras; three are management engineers; two are biomedical engineers, experts in face analysis; one is an electronic engineer; and one is a mathematical engineer, whose competences involve facial feature extraction.

The focus group also assigned weights and scores to each of the requirements as a result of a discussion among all participants, so that everyone has intervened in the debate giving a contribution linked to the specific area of expertise, and the final value has been unanimously assigned.

Results are presented in the following section.

3.1 Face detection

Even if accuracy is something to be taken care of in all contexts, this constraint can be considered not so strict for Face Detection stand-alone applications as other facial applications. Once that the face is detectable, details on facial surface are not required. This does not mean accuracy is not relevant at all: a trade-off between accuracy and resources (computational and storage resources) is always necessary; nonetheless, in Face Detection applications the limit can be set closer to the resources than Face Authentication, Face Identification and Face Expression Recognition applications. Moreover, flexibility should be a strength point for this application, so that it can work in all range, light, pose and occlusions situations (Table 6). Qualitative requirements are:

- Real-time: faces should be detected when an individual enters in the camera field-of-view [5].
- Wide operating range: faces should be detectable both if an individual is getting closer to the camera and moving away.
- Accurate at close distance: faces should be detectable if an individual is close to the camera.

- Accurate at far distance: faces should be detectable if an individual is far from the camera.
- Able to discriminate faces among other elements in the environment: the core of the application, if a face is present in the scene, then it should be detected.
- Integrable into a smartphone: sensors should allow to be put into a smartphone, a tablet, or a laptop to perform Face Detection.
- Portable: this requirement suggests having a sensor small enough to be easily carried by the user.
- Small output data: the detected face should be reported without spending too much resources in terms of memory, for reasons of storage and computational speed. Nonetheless, to preserve a level of accuracy that allows to detect faces is mandatory.
- Robust to light: faces should be detected whatever light conditions are (i.e. in the dark, in a sunny day…).
- Head pose invariant: faces should be detected whatever the individual relative orientation with respect to the camera is.
- Robust to occlusions: faces should be detected in presence of occlusions (i.e. glasses, scarfs…).

Sensors parameters relative importance is shown in Fig. 3. Radar shows that the resolution is the most important parameter, followed by the maximum distance of functioning, since Face Detection applications must detect subjects that do not necessarily position themselves in front of the camera.

3.2 Face authentication

The minimum error rate in Face Authentication is required. User is aware of the sensitivity of this application so that real-time is not strictly required, but speed should be high enough to compete with other type of authentication (for instance, the insertion of a PIN code); nevertheless, speed must not sacrifice accuracy in any way, since for Face Authentication this is the main requirement on which to focus on. (Table 7). Qualitative requirements are:

- Fast enough to unblock a device: this application does not require real-time, unblocking speed should not be annoying for the user.
- Accurate at close distance: face should be recognized from a distance as close as a smartphone, a tablet or a laptop typical user is.
- Able to detect facial features: facial landmark for face analysis must be detected.
- Integrable into a smartphone: sensors should allow to be put into a smartphone, a tablet, or a laptop to perform Face Authentication.
- Robust to light: faces should be recognized whatever light conditions are (i.e. in the dark, in a sunny day…).
- Robust to occlusions: faces should be detected in presence of small occlusions (i.e. glasses).

Sensors parameters relative importance is shown in Fig. 4. Radar shows that resolution and minimum distance of functioning are the most important technical requirements to satisfy, coherently with the most-common usage scenarios: a user that must unlock his personal device. Subsequently, frame rate and dimensions can be considered influential, since a user

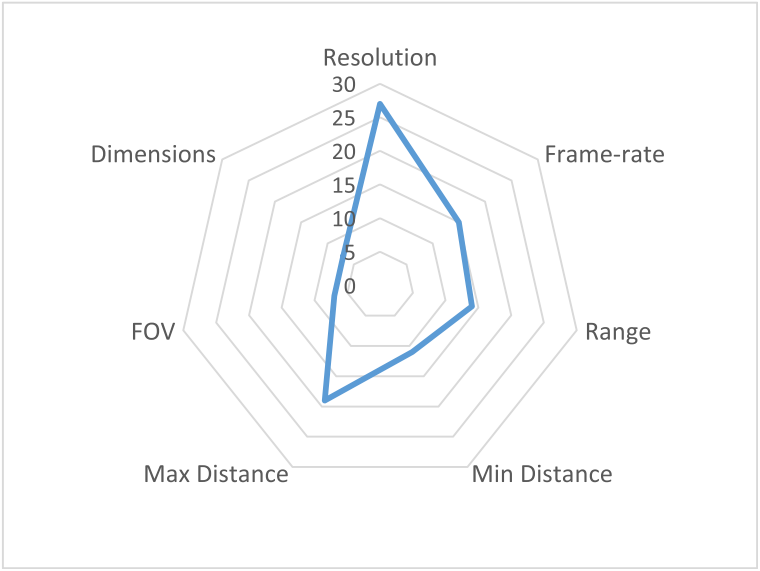

Fig. 3 Specs relative importance for Face Detection

must not wait too much time to be authenticated, otherwise another authentication method would be preferable, and the system should have the possibility of being integrated in personal devices such as smartphones, tablets and laptops.

3.3 Face identification

This application requires to council the accuracy for face analysis and the robustness to work in different range, light, pose and occlusions situation. Close distance is not considered so

Table 7 Face Detection

Application requirements	Application requirements weights	Technical requirements						
		Resolution	FPS	Range	Min distance	Max distance	FOV	Dimensions
Real-time	5	9	9	0	0	0	3	0
Wide operating range	5	0	0	9	1	9	0	0
Accurate at close distance	3	9	3	1	9	0	0	0
Accurate at far distance	3	9	1	3	0	9	0	0
Able to discriminate faces among other elements in the environment	5	9	0	3	3	9	1	0
Integrable into a smartphone	4	1	0	3	3	3	3	9
Portable	3	0	0	0	0	0	0	9
Small output data	5	9	9	0	0	0	3	0
Robust to light	4	3	1	3	3	3	3	0
Head pose invariant	5	3	1	3	3	3	0	0
Robust to occlusions	5	3	3	1	1	1	0	0
Relative total score		27%	15%	14%	11%	19%	7%	7%

relevant since Face Identification is different from Face Authentication as it has been previously explained (Table 8).

Qualitative requirements about Face Identification are:

- • Real-time: a subject should be identified before he leaves the field-of-view of the camera [23].
- Wide operating range: faces should be identified both if an individual is getting closer to the camera and moving away.
- Accurate at close distance: faces should be identified if an individual is close to the camera.
- Accurate at far distance: faces should be identified if an individual is far from the camera
- Able to detect facial features: facial landmark for face analysis must be identified.
- Integrable into a smartphone: sensors should allow to be put into a smartphone, a tablet or a laptop to perform Face Identification.
- Portable: this requirement suggests having a sensor small enough to be easily carried by the user.
- Robust to light: faces should be identified whatever light conditions are (i.e. in the dark, in a sunny day…).
- Head pose invariant: faces should be identified whatever the individual relative orientation with respect to the camera is.
- Robust to occlusions: faces should be identified in presence of occlusions (i.e. glasses, scarfs…).
- Robust to different face expressions: faces should be identified whatever the individual mood is.

Sensors parameters relative importance is shown in Fig. 5. Radar shows that the resolution confirms to be the most important technical requirement, indeed, to recognize features is mandatory to apply facial algorithms. All the technical requirements linked to the distance of

Fig. 4 Specs relative importance for Face Authentication

Table 8 Face Authentication

Application requirements	Application requirements weights	Technical requirements						
		Resolution	FPS	Range	Min distance	Max distance	FOV	Dimensions
Fast enough to unblock a device	5	9	9	1	3	3	3	0
Accurate at close distance	5	9	1	0	9	0	0	0
Able to detect facial features	5	9	0	3	9	3	0	0
Integrable into a smartphone	4	1	0	1	9	1	3	9
Robust to light	4	3	1	3	3	3	3	0
Robust to occlusions	3	3	3	1	1	1	0	0
Relative total score		28%	11%	6%	28%	9%	7%	11%

functioning appears right after resolution in the ranking, since the sensor should be able to recognize subjects that could be more or less close to the camera. This result is significantly different from Face Authentication and confirms the choice of splitting Face Recognition applications in Face Authentication and Face Identification.

3.4 Face expression recognition

Qualitative requirements about Face Expression Recognition are very similar to the Face Identification ones since the operating conditions are almost the same (Table 9):

- Real-time: individual expressions should be recognized whenever an event associated to what they are assisting is triggered [76].
- Wide operating range: individuals' expressions should be recognized both if an individual is getting closer to the camera and moving away.

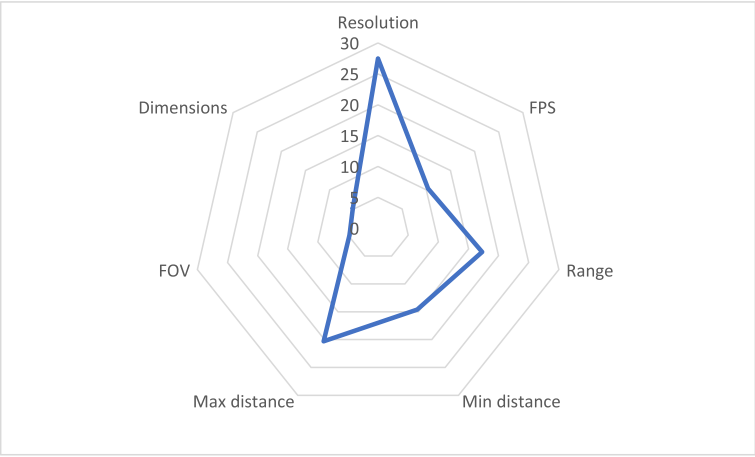

Fig. 5 Specs relative importance for Face Identification

- Accurate at close distance: individuals' expressions should be recognized if an individual is close to the camera.
- Accurate at far distance: individuals' expressions should be recognized if an individual is far from the camera.
- Able to detect facial features: facial landmarks for face analysis must be recognized.
- Integrable into a smartphone: sensors should allow to be put into a smartphone, a tablet or a laptop to perform Face Expression Recognition.
- Portable: this requirement suggests having a sensor small enough to be easily carried by the user.
- Robust to light: individuals' expressions should be recognized whatever light conditions are (i.e. in the dark, in a sunny day…).
- Head pose invariant: individuals' expressions should be recognized whatever the individual relative orientation with respect to the camera is.
- Robust to occlusions: individuals' expressions should be recognized in presence of occlusions (i.e. glasses, scarfs…).

Sensors parameters relative importance is shown in Fig. 6. The radar appears to be very similar to the Face Identification one, but this result should not be surprising. In both cases resolution must be excellent in order to discriminate between different features on resulting images. Data should be retrievable both if the subjects is close or far from the camera, and, regarding the frame rate, data should be available several times per second (this requirement is satisfied by the vast majority of analyzed sensors). Finally, dimensions and field of view are not so much considered, because sensors should be not necessarily portable and can be placed in strategic locations in order to avoid FOV issues.

Table 9 Face Identification

Application requirements	Application requirements weights	Technical requirements						
		Resolution	FPS	Range	Min distance	Max distance	FOV	Dimensions
Real-time	4	9	9	0	0	0	3	0
Wide operating range	5	0	0	9	1	9	0	0
Accurate at close distance	3	9	3	1	9	0	0	0
Accurate at far distance	4	9	1	3	0	9	0	0
Able to detect facial features	5	9	0	9	9	9	1	0
Integrable into a smartphone	3	1	0	3	3	3	3	9
Portable	2	0	0	0	0	0	0	9
Robust to light	5	3	1	3	3	3	3	0
Head pose invariant	5	3	1	3	3	3	0	0
Robust to occlusions	5	3	3	1	1	1	0	0
Robust to different face expressions	5	9	3	0	1	1	0	0
Relative total score		28%	10%	17%	15%	20%	5%	5%

A comparison between facial application specs is reported in Fig. 7. Supplementing the comments already reported, resolution can be universally recognized as the most important parameter, followed by technical requirements linked to the distance of functioning, minimum, maximum and range, depending on the facial application. Frame rate varies from 10% to 15% and this result can be explained as follows: nowadays real-time is a mandatory requirements for facial applications, nonetheless the bottleneck is not the choice of the sensor, but the computationally demanding techniques, thus the focus must be moved on the choice (and the implementation) of the suitable algorithm.

Afterwards technical specifications and facial applications have been analyzed, the most suitable 3D detection technology can now be identified (Table 11).

ToF cameras are the best in terms of long range operating functioning [48, 94], but this strength is not feasible for facial applications, and they are weaker than other technologies in terms of resolution, this is the reason why it is the worst choice for the considered facial applications.

Passive stereo technology has resulted to be the most suitable choice for Face Detection applications, due to the trade-off between high resolution and remarkable maximum operating functioning distance [60, 25], followed by the active stereo technology and, in third position, by structured light cameras, because of their too poor maximum operating functioning distance.

Face Detection has been taken into account during all the evaluation process not only as stand-alone application, but also as preliminary step of Face Authentication, Face Identification and Face Expression Recognition.

Scores of these facial applications have been given from a global point of view.

In particular, when the focus group gathered for the evaluation, the main facial application steps were taken into account and this means that they discussed about the face detection step as well as the subsequent steps such as feature extraction or analysis with neural networks.

Face Detection requirements in stand-alone applications are different from Face Detection requirements as preliminary step. The requirements of Face Authentication, Face Identification and Face Expression Recognition definitely consider Face Detection as a part of their

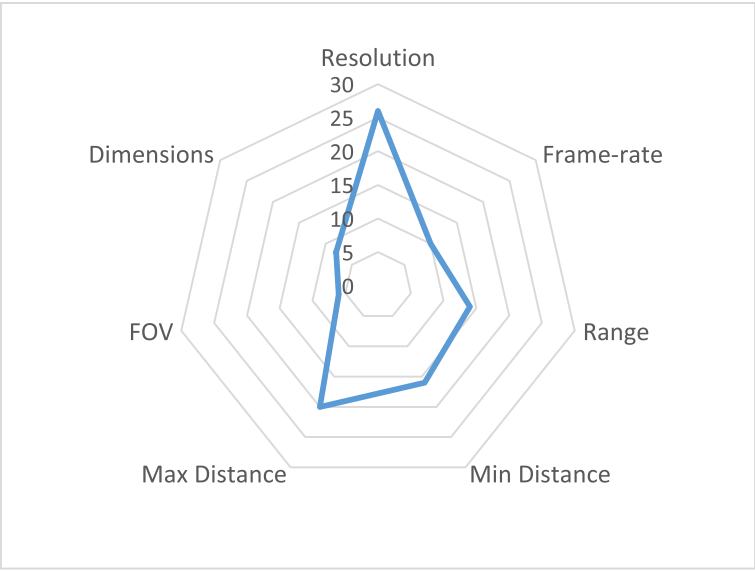

Fig. 6 Specs relative importance for Face Expression Recognition

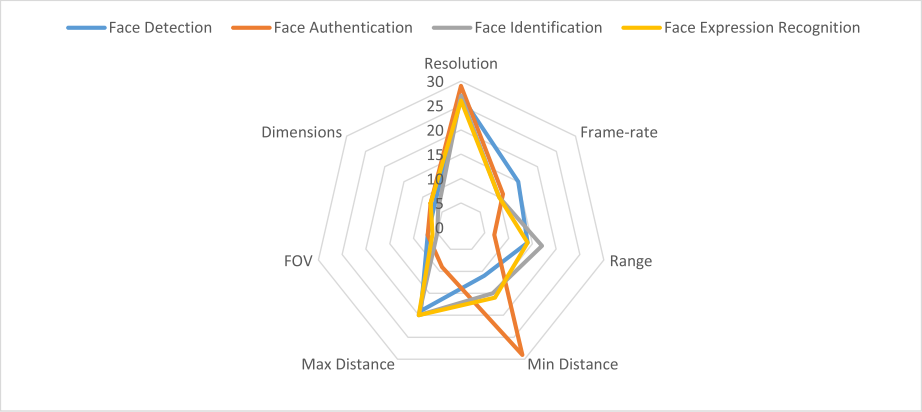

Fig. 7 Specs comparison for the different facial applications

algorithms, but some of the requirements may change based on the application within which they are incorporated.

Going into the detail, if included in a Face Authentication application, Face Detection can accept a higher response time than Face Detection as a stand-alone application such as counting people in a room. Besides, the range need not be wide, because Face Authentication use cases are at short distance.

Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.Considering Face Identification and Face Expression Recognition, the shape of the radars related to these applications are very similar each other and the Face Detection one is not too different. This testify that Face Detection requirements played a role in the evaluation of these facial application requirements. Indeed, they have not been twisted if they are considered as stand-alone or integrated, nonetheless there are some differences. In terms of relative importance, frame rate has a greater value in stand-alone Face Detection applications, while minimum distance acquires importance in Face Identification and Face Expression Recognition.

The situation is inverted in Face Authentication. Since minimum distance is the most important parameter, together with the resolution, the excellent minimum operating functioning distance of structured light technology has resulted to be the best for this application [78, 104]. It is mandatory to observe that an active stereoscopy sensor seems to be the best at close range, but this is false to a broader set of sensors. Since active stereoscopy is the most recent technology, it is wise to bear in mind this result, but the time is not yet ripe to claim that it is the best one for close-range applications and, consequently, Face Authentication.

Face Identification and Face Expression Recognition have resulted to be similar in terms of qualitative requirements, in fact the shapes of their technical specifications relative importance are really close to each other. Active stereoscopy is the most suitable technology for these applications [45], because of the presence of good resolution both at close distance and long distance operating functioning at the same time. Passive stereoscopy is the second-best choice, thanks to its very high resolution and operating functioning at high distance, that is more relevant with respect to close distance. This is the reason why structured light is in third position, in fact the poor maximum operating function distance has been penalizing for this sensor category.

Table 10 Face Expression Recognition

Application requirements	Application requirements weights	Technical requirements						
		Resolution	FPS	Range	Min distance	Max distance	FOV	Dimensions
Real-time	5	9	9	0	0	0	3	0
Wide operating range	4	0	0	9	1	9	0	0
Accurate at close distance	4	9	3	1	9	0	0	0
Accurate at far distance	4	9	1	3	0	9	0	0
Able to detect facial features	5	9	0	3	9	9	1	0
Integrable into a smartphone	4	1	0	3	3	3	3	9
Portable	3	0	0	0	0	0	0	9
Robust to light	4	3	1	3	3	3	3	0
Head pose invariant	4	3	1	3	3	3	0	0
Robust to occlusions	4	3	3	1	1	1	0	0
Relative total score		26%	10%	14%	16%	20%	6%	8%

Key technical specifications used to analyze 3D sensor technologies are strongly linked to accuracy more than the acquisition time. From the datasheet analysis, it has been found that all the considered 3D sensors can provide several FPS when acquiring single shot acquisitions; if all of them can satisfy the real-time requirement, it has been unavoidable to focus on other technical specifications to discriminate between RGB-D cameras and to evaluate 3D acquisition technologies.

4 Conclusions

In this paper a survey to understand which 3D sensor technology can fit better different facial applications has been conducted. Qualitative requirements for the most common face applications and sensors specifications considered in the present survey are the result of a desk research about 3D facial applications and 3D sensor technologies.

A focus group has filled-in four QFDs to identify the main features involved in each application and to understand which the most suitable technology for depth detection is. Results show that passive stereoscopy is the best technology choice for Face Detection applications, structured light is the most suitable sensor technology for Face Authentication applications and active stereo is the most interesting technology for Face Recognition and Face Expression Recognition (Table 10).

Table 11 Technology ranking for facial application

	Face Detection	Face Authentication	Face Identification	Face Expression Recognition
1st	Passive stereoscopy	Structured light	Active stereoscopy	Active stereoscopy
2nd	Active stereoscopy	Active stereoscopy	Passive stereoscopy	Passive stereoscopy
3rd	Structured light	Passive stereoscopy	Structured light	Structured light
4th	ToF	ToF	ToF	ToF

Future work consists in performing an empirical analysis of 3D sensors to proof the theoretical results presented in this survey. Furthermore, a 3D QFD will be presented to further point out a correct technology choice for facial applications presenting a new orthogonal dimension in addition to qualitative requirements and technical specifications.

Funding information Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Abate A, Nappi M, Riccio D, Sabatino G (2007) 2D and 3D face recognition: A survey. Pattern Recogn Lett 28:1885–1906
2. G. Albakri and S. Alghowinem, "The effectiveness of depth data in liveness face authentication using 3D sensor cameras," *Sensors,* vol. 19, no. 8, p. 1928, 2019.
3. Albino V, Berardi U, Dangelico RM (2015) Smart cities: definitions, dimensions, performance, and initiatives. J Urban Technol 22(1):3–21
4. Alexandre GR, Soares JM, Thé GAP (2020) Systematic review of 3D facial expression recognition methods. Pattern Recogn 100:107108
5. Aljohani M, Tanweer A (2017) Real time face detection in ad hoc network of android smart devices. Advances in Computational Intelligence:245–255
6. M. R. Andersen, T. Jensen, P. Lisouski, A. K. Mortensen, M. K. Hansen, T. Gregersen and P. Ahrendt, "Kinect depth sensor evaluation for computer vision applications," *Technical Report Electronics and Computer Engineering,* vol. 1, no. 6, 2012.
7. J. Ashbourn, Biometrics: advanced identity verification: the complete guide, Springer, 2014.
8. "Asus," [Online]. Available: https://www.asus.com/us/3D-Sensor/Xtion_PRO_LIVE/specifications/.
9. M. S. L. G. Bartlett, I. Fasel and J. R. Movellan, "Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003). Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction," in Conference On Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03, 2003.
10. "Baslerweb," [Online]. Available: https://www.baslerweb.com/en/products/cameras/3d-cameras/time-of-flight-camera/.
11. V. Bettadapura, "Face expression recognition and analysis: the state of the art," *arXiv preprint arXiv: 1203.6722,* 2012.
12. R. D. Bock, "Low-cost 3D security camera," Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything, vol. 10643, p. 106430E, 2018.
13. Boffano M, Pellegrino P, Ratto N, Giachino M, Albertini U, Aprato A, Boux E, Collo G, Ferro A, Marone S, Massè A, Piana R (2018) Custom-made 3D-printed pelvic prosthesis: is it a safe option for the limb salvage in tumours and catastrophic total hip arthroplasty failures? *Orthopaedic Proc* 100(SUPP_5):93
14. J. Bouguet, B. Curless, P. Debevec, M. Levoy, S. Nayar and S. Seitz, "overview of active vision techniques," in Procedings of ACM SIGGRAPH Workshop, Course on 3D Photography, 2000.
15. V. Bruce and A. Young, In the eye of the beholder: the science of face perception, Oxford university press, 1998.
16. Calvo MG, Nummenmaa L (2016) Perceptual and affective mechanisms in facial expression recognition. Cognit Emot 30(6):1081–1106
17. Cament LA, Galdames FJ, Bowyer KW, Perez CA (2015) Face recognition under pose variation with local Gabor features enhanced by active shape and statistical models. Pattern Recogn 48(11):3371–3384
18. Cao J, Hu Y, Yu B, He R, Sun Z (2019) 3D aided duet GANs for multi-view face image synthesis. IEEE Trans Inform Forensics Secur 14(8):2028–2042

19. M. Carfagni, R. Furferi, L. Governi, C. Santarelli, M. Servi, F. Uccheddu and Y. Volpe, "Metrological and critical characterization of the Intel D415 stereo depth camera," *Sensors,* vol. 19, no. 3, p. 489, 2019.
20. "Carnegie Robotics," [Online]. Available: https://carnegierobotics.com/multisense-s7/.
21. Chae MP, Rozen WM, McMenamin PG, Findlay MW, Spychal RT, Hunter-Smith DJ (2015) Emerging applications of bedside 3D printing in plastic surgery. Frontiers Surg 2:25
22. Chen X, Ma H, Wan J, Li B, Xia T (2017) Multi-View 3D Object Detection Network for Autonomous Driving. *IEEE CVPR* 1(2):3
23. S. Chen, Y. Liu, X. Gao and Z. Han, "Mobilefacenets: efficient cnns for accurate real-time face verification on mobile devices," in *Chinese Conference on Biometric Recognition*, Cham, 2018.
24. Y. Chen, R. Hu, J. Xiao and Z. Wang, "multisource surveillance video coding by exploiting 3d and 2d knolwedge, " in 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.
25. M. Chowdhury, J. Gao and R. Islam, "human detection and localization in secure access control by analysing facial features," in 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), 2016.
26. Chuang PT (2001) Combining the analytic hierarchy process and quality function deployment for a location decision from a requirement perspective. Int J Adv Manuf Technol 18(11):842–849
27. Colombo A, Cusano C, Schettini R (2006) 3D face detection using curvature analysis. Pattern Recogn 39(3):444–455
28. Dagnes N, Vezzetti E, Marcolin F, Tornincasa S (2018) Occlusion detection and restoration techniques for 3D face recognition: a literature review. Mach Vis Appl:1–25
29. Damasio AR, Damasio H, Van Hoesen GW (1982) Prosopagnosia Anatomic basis and behavioral mechanism. Neurology 32(4):331
30. Dawood A, Marti BM, Sauret-Jackson V, Darwood A (2015) 3D printing in dentistry. *British dental journal* 219(11):521
31. Deng J, Roussos A, Chrysos G, Ververas E, Kotsia I, Shen J, Zafeiriou S (2019) The menpo benchmark for multi-pose 2D and 3D facial landmark localisation and tracking. Int J Comput Vis 127(6–7):599–624
32. T. P. Driver, S. Sundaram, G. Khandelwal and M. Sahasrabudhe, "Systems And Methods For Patient Identification Using Mobile Face Recognition". U.S. Patent 11/945, 2009.
33. "DS325 Datasheet," [Online]. Available: https://www.sony-depthsensing.com/Portals/0/Download/WEB_20120907_SK_DS325_Datasheet_V2.1.pdf.
34. "Duo 3D," [Online]. Available: https://duo3d.com/product/duo-minilx-lv1#tab=specs.
35. "e-con Systems," [Online]. Available: https://www.e-consystems.com/3D-USB-stereo-camera.asp.
36. Ekman P (1992) An argument for basic emotions. Cognit Emot 6(3–4):169–200
37. P. Ekman and W. V. Friesen, Unmasking the face: A guide to recognizing emotions from facial clues, Ishk, 2003.
38. L. A. Elrefaei, A. Alharthi, H. Alamoudi, S. Almutairi and F. Al-rammah, "real-time face detection and tracking on mobile phones for criminal detection," in 2017 2nd International Conference on Anti-Cyber Crimes (ICACC), 2017.
39. "Ensenso," [Online]. Available: https://www.ensenso.com/support/modellisting/?id=N35-606-16-BL.
40. Fantz (1961) The origin of form perception. Sci Am 204(5):66–73
41. M. E. Fathy, V. M. Patel and R. Chellappa, "face-based active authentication on mobile devices," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing, 2015.
42. Forte M (2014) 3D archaeology: new perspectives and challenges - the example of Çatalhöyük. J Eastern Mediterranean Archaeol and Heritage Studies 2(1):1–29
43. Galbally J, Marcel S, Fierrez J (2014) Image quality assessment for fake biometric decision: application to iris, fingerprint, and face recognition. IEEE Trans Image Process 23(2):710–724
44. Z. Geng, C. Cao and S. Tulyakov, "3d guided fine-grained face manipulation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
45. S. Giancola, M. Valenti and R. Sala, "metrological qualification of the Intel D400™ active stereoscopy cameras," in A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies, Springer, 2018, pp. 71–85.
46. Heisele B, Serre T, Poggio T (2007) A component-based framework for face detection and identification. Int J Comput Vis 74(2):167–181
47. Henry P, Krainin M, Herbst E, Ren X, Fox D (2012) RGB-D mapping: using Kinect-style depth cameras for dense 3D modeling of indoor environments. Int J Robotics Res 31(5):647–663
48. Horaud R, Hansard M, Evangelidis G, Menier C (2016) An overview of depth cameras and range scanners based on time-of-flight technologies. Mach Vis Appl 27(7):1005–1020
49. Hossain MS, Muhammad G (2015) Cloud-assisted speech and face recognition framework for health monitoring. Mobile Netw Appl 20(3):391–399
50. "Ifm ," [Online]. Available: https://www.ifm.com/us/en/product/O3D303.
51. B Ingxin, L. Yinan and Z. Shuo, "3D Multi-poses Face Expression Recognition Based on Action Units," in Proceedings of the 2019 International Conference on Information Technology and Computer Communications, 2019.

52. "Intel," [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-D400-Series-Datasheet.pdf.
53. "Intel Euclid ," [Online]. Available: https://click.intel.com/media/productid2100_10052017/335926-001_public.pdf.
54. "Intel RealSense F200 ," [Online]. Available: https://communities.intel.com/docs/DOC-24012.
55. "Intel RealSense R200," [Online]. Available: https://www.intel.it/content/www/it/it/support/articles/000016214/emerging-technologies/intel-realsense-technology.html.
56. "Intel RealSense SR300," [Online]. Available: https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/realsense-sr300-datasheet1-0.pdf.
57. Jain AK, Hong L, Pankanti S, Bolle R (1997) An identity-authentication system using fingerprints. Proc IEEE 85(9):1365–1388
58. Kahraman C, Ertay T, Büyüközkan G (2006) A fuzzy optimization model for QFD planning process using analytic network approach. Eur J Oper Res 171(2):390–411
59. Kedzierski M, Fryskowska A (2014) Terrestrial and aerial laser scanning data integration using wavelet analysis for the purpose of 3D building modeling. Sensors 14(7):12070–12092
60. E. Kirsten, L. Inocencio, M. Veronez, L. Da Silveira, F. Bordin and F. Marson, "3D data acquisition using stereo camera," in GARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, 2018.
61. S. Lamba, N. Nain and H. Chahar, "A robust multi-model approach for face detection in crowd," in 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2016.
62. J.-S. Lee and D.-H. Shin, "Lee, Jong-Sik, and Dong-Hee Shin. "The relationship between human and smart TVs based on emotion recognition in HCI," in International Conference on Computational Science and Its Applications, 2014.
63. C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens and D. Vandermeulen, "feature detection on 3D face surfaces for pose normalisation and recognition," in Fourth IEEE International Conference on Biometrics: Theory Applications and Systems, 2010.
64. D. McDuff, A. Mahmoud, M. T. J. Amr and R. E. Kaliouby, "AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2016.
65. Meadows J (1974) The anatomical basis of prosopagnosia. J Neurol Neurosurg Psychiatry 37(5):489–501
66. Mizuno S, Akao Y (1994) Development history of quality function deployment. Quality Res 90:339
67. M. Mohammed, J. Tatineni, B. Cadd, P. Peart and I. Gibson, "Applications of 3D topography scanning and multi-material additive manufacturing for facial prosthesis development and production," Proceedings of the 27th Annual International Solid Freeform Fabrication Symposium, pp. 1695–1707, 2016.
68. S. Mondal, I. Mukhopadhyay and S. Dutta, "review and comparisons of face detection techniques," in *Proceedings of Interantional Ethical Hacking Conference*, Singapore, 2019.
69. Morton J, Johnson MH CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological review* 98(2):164–181 1191
70. N. Nawana, W. C. Horton, W. J. Frasier, M. O'neil, R. E. Sommerich, J. DiPietro and M. Parsons, "Medical robotics and computer-integrated surgery," Springer handbook of robotics, pp. 1657–1684, 2016.
71. "Nerian," [Online]. Available: https://nerian.com/products/scenescan-stereo-vision/.
72. Nonis F, Olivetti EC, Marcolin F, Violante MG, Vezzetti E, Moos S (2020) Questionnaires or Inner Feelings: Who Measures the Engagement Better? *Applied Sciences* 10:609
73. E. C. Olivetti, M. G. Violante, E. Vezzetti, F. Marcolin and B. Eynard, "Engagement Evaluation in a Virtual Learning Environment via Facial Expression Recognition and Self-Reports: A Preliminary Approach," *Applied Sciences,* vol. 10, no. 1, p. 314, 2020.
74. Olivetti EC, Ferretti J, Cirrincione G, Nonis F, Tornincasa S, Marcolin F (2020) "deep CNN for 3D face recognition," in *International Conference on Design*. The Innovation Exchange, Simulation
75. "Orbecc3d," [Online]. Available: https://orbbec3d.com/astra-mini/.
76. J. V. Patil and P. Bailke, "real time facial expression recognition using RealSense camera and ANN," in 2016 International Conference on Inventive Computation Technologies (ICICT), 2016.
77. Photoneo, [Online]. Available: https://www.photoneo.com/phoxi-3d-scanner/.
78. T. Pribanic, T. Petkovic, M. Donlic, V. Angladon and S. Gasparni, "3D structured light scanner on the smartphone," in International Conference on Image Analysis and Recognition, Cham, 2016.
79. Rabia J, A. H. R. (2009) A survey of face recognition techniques. Jips 5(2):41–68
80. R. Raghavendra, K. B. Raja, A. Pflug, B. Yang and C. Busch, "3d face reconstruction and multimodal person identification from video captured using smartphone camera," in 2013 IEEE International Conference on Technologies for Homeland Security (HST), 2013.
81. A. Rattani, R. Derakhshani and A. Ross, Selfie Biometrics: Advances and Challenges, Springer Nature, 2019.
82. Riaz S, Park U, Choi J, Natarajan P (2019) Age progression by gender-specific 3D aging model. Mach Vis Appl 30(1):91–109

83. D. Robertson, D. G. Macfarlane, R. I. Hunter, S. L. Cassidy, N. Llombart, E. Gandini, T. Bryllert, M. Ferndahl, H. Lindstrom, J. Tenhunen, H. Vasama, J. Huopana, T. Selkala and A.-J. Vuotikka, "High resolution, wide field of view, real time 340GHz 3D imaging radar for security screening," Passive and Active Millimeter-Wave Imaging XX, vol. 10189, 2017.
84. "Roboception," [Online]. Available: https://roboception.com/en/rc_visard-en/.
85. Ross A, Jain A (2003) Information fusion in biometrics. Pattern Recogn Lett 24(13):2115–2125
86. Roy S, Podder S (2013) Face detection and its applications. Int J Res Eng Adv Technol 1(2):1–10
87. Salvi J, Pages J, Batlle J (2004) Pattern codification strategies in structured light systems. Pattern Recogn 37(4):827–849
88. A. Sepas-Moghaddam, F. M. Pereira and P. L. Correia, "Face recognition: A novel multi-level taxonomy based survey," IET Biometrics, 2019.
89. "Sick," [Online]. Available: https://www.sick.com/it/it/visione/visione-3d/visionary-t/c/g358152.
90. S. Siebert and J. Teizer, "Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system," Automation in Construction, no. 41, pp. 1–14, 2014.
91. Small DA, Verrochi NM (2009) The face of need: facial emotion expression on charity advertisement. J Mark Res 46(6):777–787
92. "Stackoverflow," [Online]. Available: https://stackoverflow.com/questions/7696436/precision-of-the-kinect-depth-camera.
93. "Stereolabs," [Online]. Available: https://www.stereolabs.com/zed/.
94. Streeter L, Kuang Y (2019) Metrological aspects of time-of-flight range imaging. IEEE Instrument Measurement Magazine 22(2):21–26
95. "Structure," Occipital, [Online]. Available: https://support.structure.io/article/157-what-are-the-structure-sensors-technical-specifications.
96. "Swiss Ranger SR4000," [Online]. Available: http://www.adept.net.au/cameras/Mesa/SR4000.shtml.
97. "Swiss Ranger SR4500," [Online]. Available: http://www.adept.net.au/cameras/Mesa/SR4500.shtml.
98. L. Tran and X. Liu, "on learning 3d face morphable model from in-the-wild images," in IEEE transactions on pattern analysis and machine intelligence, 2019.
99. Valverde I, Gomez G, Gonzalez A, Suarez-Mejias CAA, Coserria JF, Uribe S, Gomez-Cla T, Hosseinpour AR (2015) Three-dimensional patient-specific cardiac model for surgical planning in Nikaidoh procedure. Cardiol Young 25(4):698–704
100. Verschuren P, Doorewaard H, Mellion MJ (2010) Designing a research project. Eleven International publishing house, The Hague
101. Vezzetti E, Marcolin F (2014) Geometry-based 3D face morphology analysis: soft-tissue landmark formalisation. Multimed Tools Appl 68(3):895–929
102. Vezzetti E, Moos S, Marcolin F, Stola V (2012) A pose-independent method for 3D face landmark formalization. Comput Methods Prog Biomed 108(3):1078–1096
103. Vezzetti E, Marcolin F, Tornincasa S, Ulrich L, Dagnes N (2017) 3D geometry-based automatic landmark localization in presence of facial occlusions. Multimed Tools Appl:1–29
104. Wang Z (2020) Robust three-dimensional face reconstruction by one-shot structured light line pattern. Opt Lasers Eng 124:105768
105. H. Wu and H. L. P. Xu, "design and implementation of cloud service system based on face recognition," in *Conference on Complex, Intelligent, and Software Intensive Systems*, 2020.
106. Yarboro S, Richter PH, Kahler DM (2017) The evolution of 3D imaging in orthopaedic trauma care. Unfallchirurg 120(1):5–9
107. Young AW, De Haan E, Bauer R (2008) Face perception: A very special issue. J Neuropsychol 2(1):1–14
108. C. Zhang and Z. Zhang, "A survey of recent advances in face detection," 2010.
109. S. Zhou and S. Xiao, " 3D face recognition: a survey," Human-Centric Comput Inform Sci, vol. 8, no. 1, p. 35, 2018.
110. Zollhofer M, Thies J, Garrido P, Bradley D, Beeler T, Perez P, Stamminger M, Niessner M, Theobalt C (2018) State of the art on monocular 3D face reconstruction, tracking, and applications. Computer Graphics Forum 37(2):523–550
111. A. Neethu and B. Kamal, "people count estimation using hybrid face detection method," in 2016 International Conference on Information Science (ICIS). *IEEE*, 2016.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.