Measurement 200 (2022) 111643

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A methodology for task-specific metrological characterization of low-cost 3D camera for face analysis

Giacomo Maculotti *, Luca Ulrich, Elena Carlotta Olivetti, Gianfranco Genta, Federica Marcolin, Enrico Vezzetti, Maurizio Galetto

Department of Management and Production Engineering, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy

ARTICLE INFO

Keywords:
Measurement uncertainty
Machine vision
Face analysis
Soft tissue landmarks
Human-machine interaction
Depth cameras

ABSTRACT

Facial appearance is one prominent feature in analyzing several aspects, e.g., aesthetics and expression of emotions, and face analysis is crucial in many fields. Face analysis requires measurements that can be performed by different technologies and typically relies on landmarks identification. Recently, low-cost customer grade 3D cameras have been introduced in the market, enabling an increase of application at affordable cost with nominal adequate performances. Novel cameras require to be thoroughly metrologically characterized to guarantee these performances. Cameras are calibrated following a standard general-purpose procedure. However, the specificity of facial measurements requires a task-based metrological characterization to include typical influence factors. This work outlines a methodology for task-based metrological characterization of low-cost 3D cameras for facial analysis, consisting of: influence factor identification by ANOVA, related uncertainty contribution assessment, uncertainty propagation, landmarking uncertainty estimation. The proposed methodology is then demonstrated on a customer grade state-of-the-art 3D camera available on the market.

1. Introduction

Amongst human characteristics, facial appearance is the most significant in determining aspects ranging from aesthetics to the expression of moods and emotions. Face analysis is essential in several applications, e.g., emotion recognition with facial expressions, biometric authentication and identification, human-computer interaction [1–4], age estimation [5], quantification of facial attractiveness [6–8]. Moreover, several medical, technological, and industrial applications benefit from studying the human face. Concerning the first, i.e., the medical field, oral-maxillo-facial surgery potentially benefits most from face analysis. In fact, the introduction of precise and reliable imaging and reconstruction technologies, such as 3D cone beam computed tomography (3D CBCT), made available a three-dimensional approach for the analysis of hard and soft tissues of the face, which is more sophisticated than traditional bi-dimensional cephalometry [9–13]. Also, surgical planning of bony repositioning can improve its accuracy and reliability [13,14]. For example, it can be achieved by predicting the patients' actual post-operative aesthetic appearance related to facial soft tissue involuntary displacements [15]. Some algorithms have been proposed to forecast the future soft tissue aspect of patients undergoing orthognathic corrective surgeries to that aim [9].

Measuring soft tissue features is an enabling technology for quantitatively evaluating people's interaction and engagement with e-learning systems, allowing to overcome the loss of face-to-face interaction [16,17]. The recent widespread use of distance learning led to the impossibility for educators to observe external factors, e.g., facial features, speech, gestures [18], to evaluate learners' engagement and attention. Computer vision applied to facial expression analysis aids in overcoming this limit [19]. Spivak et al. exploited similar ideas for a different remarkable application, proposing the automation of the marketing decision-making process through a custom facial expression recognition system capable of evaluating product quality [20]. Evaluating customer satisfaction is one of the most powerful tools for retailers and service providers [21]. Traditionally, data are collected through surveys and self-reports, whilst artificial intelligence and, particularly, machine learning leads to automated data collection and process measurement [22]. Facial expression analysis appeared to be the most suitable and the least intrusive method for large scale studies for these applications [23]. In this context, Gonzales-Rodriguez et al. proposed an application of facial emotion recognition to the field of tourism, highlighting the potentiality of such an approach [22].

* Corresponding author.
E-mail address: giacomo.maculotti@polito.it (G. Maculotti).

https://doi.org/10.1016/j.measurement.2022.111643
Received 22 October 2021; Received in revised form 2 May 2022; Accepted 15 July 2022
Available online 19 July 2022

The development of human-computer interaction (HCI) and human-robot interaction (HRI) is preparing the ground for the industry of the future, where humans and machines will share spaces and perform tasks in collaboration [24]. However, along with benefits such as increased productivity and efficiency, there are also emerging risks associated with direct interaction between humans and robots [25,26]. In this sense, the capability of the machines to understand human emotions through face expression recognition (FER) could enable a more effective interaction [27–31], but the gap between humans and machines in FER ability should be addressed.

All the above-mentioned applications can be designed and implemented relying on the colour and depth information provided by RGB-D cameras. Nonetheless, the evaluation of the suitability of a device is usually a qualitative assessment, which could lead not to choose the proper camera for the application. This work aims to present a task-based methodology for the metrological characterization of a device. The considered task is related to face analysis, involving the identification of a subset of facial landmarks, and the adopted device is the Intel RealSense SR300 RGB-D camera. The following subsection motivates this work by describing the main criticalities of facial analysis, the main state-of-the-art measurement techniques' strength and weaknesses and, discussing how current characterization methods for these technologies presents some shortcomings in thoroughly evaluating measurement uncertainty of face analysis tasks.

1.1. State of the art

In the last decades, computer-aided facial analysis in mentioned fields relied upon a 2D landmarking approach, i.e., positioning significant cephalometric points, called landmarks, on 2D data. The spread of new sensors for image acquisition, e.g., CBCT and depth cameras, enabled a three-dimensional approach for landmark positioning, with potentially higher reliability thanks to the information from the three dimensions and the increased robustness to light changes, pose orientations, and occlusions [32]. Regardless of the measurement approach, landmarking is a complex task, subject to errors due to the inter- and intra-operator variability in the location of cephalometric points [33]. Additionally, facial dysmorphism may induce further criticalities. Despite the research in the direction of automatic methods for 3D (or 2.5 D) landmarking [33,34], there are still cases that might benefit from manual positioning by an experienced operator, e.g., maxillo-facial-surgery application, or when there is no need for a real-time landmarking, e.g., in off-line marketing and satisfaction/engagement evaluation. Concerning the mentioned development of FER systems to improve the effectiveness and the safety in the domain of HRI, the construction of a training dataset to instruct the machine to recognize changes in human emotions should be based more on the reliability of the located landmarks, even if positioned manually by an expert operator, rather than on an automated time-saving method.

Several technologies are available for depth acquisition. These include expensive technologies, e.g. laser scanner [35], and customer grade alternatives, which are cheaper but still with interesting metrological performances and allow real-time application. These characteristics have ensured the spread of these devices in the last decade, and this work focuses on them. In particular, the interest in RGB-D cameras has grown considerably since the integration of colour (RGB) and depth (D) information has given impetus to the improvement of applications in a wide variety of fields such as hand gesture spotting and recognition for HCI [36], mobile robot navigation [37], automotive [38], and even agriculture [39].

As mentioned, several alternatives to obtain the depth information have been developed and integrated in RGB-D cameras. Passive stereoscopy has been the first developed technology amongst the consumer market cameras, with the Microsoft Kinect v1 [40]. At least two cameras must be integrated into the same device to gather the depth information since depth is computed through triangulation between the point in the world space and the pixels on camera planes on which the point is projected. Resolution and operative range are excellent compared to the other technologies. Conversely, short-range applications may suffer from occlusions: if neither the cameras can detect a certain point in the space, the distance of that point from the device cannot be computed [41].

Structured-light devices retrieve the depth information projecting a pattern on the scene. Then, the pattern is deformed by the object surface and acquired by a receiver. The data are processed by an application-specific integrated circuit (ASIC). From the difference between the a priori known pattern and the deformed pattern acquired by the receiver, it is possible to compute the distance of the object from the camera. The transmitter and the receiver are placed close to each other to minimize the occlusion problem encountered in passive stereoscopy. Structured-light devices provide state-of-art results in short-range applications for consumer market cameras [42].

The distance of a point in space is computed using time-of-flight technology. An electromagnetic wave is projected on the scene, and the distance is computed by timing the lapse between the moment when the light is emitted and the moment when light is acquired by the receiver after that has been reflected back by the object surface. This technology is becoming adopted widely in the smartphone market [43] and can be adapted to the application by using different wavelengths and varying the power of the emitted light. The resolution is typically poor compared to the other technologies.

The active stereoscopy working principle is similar to the passive stereoscopy one. The main difference lies in the introduction of an emitter that projects a pattern on the scene. This pattern artificially adds some features to the object to solve the correspondence problem, even in critical scenarios. For instance, a flat wall does not have any feature that the two opposite cameras can retrieve. Adding new features allows finding the conjugate points, namely the pixels on the images acquired by the two cameras, which refer to the same point in the space, hence computing the distance of the point from the device. This improvement widened the operative range; nevertheless, the occlusion problem still persists, especially at short-range distances.

1.2. Scope of the work

This work proposes an innovative task-based metrological characterization, i.e. identification of influence factors and estimation of measurement uncertainty, for customer grade cameras for face analysis. This section discusses the practical relevance of the metrological characterization, and the shortcomings of currently available methodologies will motivate the need of the hereby proposed task-based approach.

Evaluating the metrological performance of these depth measuring devices is essential, within the complex framework briefly introduced above, including several measuring instruments and algorithms for facial analysis devoted to many applications. In fact, the reliability of the 3D information is core to understand if the depth can be used only to complement the contribution of the RGB data in the development of face analysis algorithms, or even to be used standalone in applications that do not require any other data source but the depth frames. In particular, calibration is necessary to guarantee traceability and estimate accuracy. The evaluation of measurement uncertainty must be carried out to enable the comparison of performances amongst different technologies, evaluate the precision of the measurement result, and verify the compliance with applications precision requirements. Additionally, information on the precision of the measured data may provide valuable input to improve the statistical modelling for machine learning and artificial intelligence-based applications, with definitive advantages in terms of robustness and accuracy [44–46].

Standards and literature agree on calibration and metrological characterization procedures, based on standard artefacts and material measures [47,48]. In particular, calibrated ball bars and spheres are exploited to estimate average errors in measuring specific shapes and

linear distances. Similarly, the measurement of flat surfaces allows for correcting systematic errors of planarity [47,48]. However, the evaluation of measurement uncertainty for face analysis application requires a task-based approach to estimate performances, applicability range and influence factors peculiar to the case at hand. In fact, the peculiarity of the measurand surface in the case at hand, i.e., a face, may introduce specific errors related to the curvature and particular interactions between the measurement instrument and discontinuities of the surface, e.g., occlusions, different types of reflection, presence of facial hair. Consequently, current calibration methods that rely on few specific shapes are likely to neglect some contribution to measurement uncertainty in facial measurements. Moreover, when the task is face analysis, which, as mentioned, often relies foremostly on landmarking, i.e., the identification of specific cephalometric point, average error indicators, as per current standard and state-of-the-art, lack of the required degree of details to thoroughly associate measurement uncertainty to the task result. To the best knowledge of the authors, task-based uncertainty evaluation approaches for face analysis are still lacking in the current literature, also due to the recent introduction of the aforementioned measurement technologies. Accordingly, this work proposes a procedure to evaluate the metrological performances of the depth data acquired for face analysis. In this work, the Intel RealSense SR300, a structured-light RGB-D camera, is considered.

The rest of the paper is organized as follows: Section 2 discusses the considered technology and presents the proposed methodology to metrologically characterize the instrument. Section 3 presents and discusses the results. Finally, Section 4 draws conclusions.

2. Materials and methods

2.1. Measuring instrument

Intel RealSense SR300 is the device considered in this work (see Fig. 1), a coded-light RGB-D camera particularly suitable for facial application, considering its performances in short-range scenarios.

The expression coded-light identifies a subgroup of structured-light depth acquisition methodologies using spatial or temporal codes. The emitted light wavelength is in the infrared (IR) range, around 860 nm, safe and invisible to the human eye. A micro-electro-mechanical system (MEMS) mirror generates a set of predefined, and increasing spatial frequency coded IR vertical bar patterns to be projected on the scene. The IR camera resolution is 640 × 480 pixels, and each pixel is a square with a side of 3.6 μm. Furthermore, the IR camera is equipped with an IR band filter. For completeness, it is worth mentioning that the RGB camera is also equipped with an IR band filter not to be affected by the projected patterns. The depth reconstruction is performed by an ASIC and consists of two main steps: the codeword extraction and the depth generation phase. During the first step, the codeword referring to each pixel is extracted after the scene has been illuminated with different IR patterns. During the second step, the depth map is generated considering all the patterns projected on the scene through a pipeline which consists of codeword decoding and error correction, triangulation using codewords and camera calibration data, and post-processing to minimize the noise introduced by the temporal multiplexing, due to the motion sensitivity of this technique [49].

The measurement results in a depth point cloud, inherently registered to an RGB image, arranged as a matrix of $n_x \times n_y$ pixels. Here, the relevant result is the depth, i.e., the distance between the measured object and the sensor, $z = z(x, y)$, where x and y are the lateral coordinates that are discretized by the pixel matrix. Consequently, the depth map is a set of cartesian triplets $D = \{x, y, z(x, y)\}$. The operating range of the instrument is between 20 cm and 150 cm. The number of measured pixels is constant. Thus, the pixel size changes, and the greater the distance between the measurand and the sensors, the worse the lateral resolution. This also contributes to systematic errors of non-planarity and depth offset. Carfagni et al. showed, by calibration with standard artefacts, that the magnitude of the former depends on the position in the field of view, whilst the latter is quadratically dependent on the distance between the measurand and the sensor [48]. Moreover, they evaluated the metrological characteristics according to the current standard exploiting calibrated ball bars, resulting in a form probing error of 8.3 mm, a size probing error of 1.91 mm, a sphere spacing distance error of 6.05 mm and a flatness error of 6.88 mm.

Given the working principle and the well-known occlusion problem, the interaction between the measurand surface and the projected light pattern is relevant and may result in non-measured points. If not properly managed, these may hamper or distort subsequent analyses [50]. Limiting to face analysis applications, these can be sourced by shadowed areas, e.g., face portions behind nostrils, or due to uneven surface, e.g., due to the presence of beard and facial hair [51].

2.2. Face analysis methodology

The measured frame is then exploited for the landmarking required for the face analysis. Landmarks, particularly the distances between them, are used for several scopes, ranging from face recognition, expression and emotion recognition, with applications in several fields, e.g., healthcare, marketing, industrial safety, manufacturing. Landmarking is carried out either by means of software [52,53] or by highly experienced operators. Ultimately, this operation results in finding the coordinates of a point $L = (L_x, L_y, L_z)$. Manual identification is still relevant for off-line operation and for calibrating and evaluating the accuracy of software for automatic detection and is considered and applied in this work.

Several cephalometric points, i.e., landmarks, are available. In this work, five landmarks are considered, i.e., the two *alae*, the *pronasal*, the *subnasal* and the *nasion*, shown in Fig. 2. The operators relied on the definitions of the cephalometric landmarks given by Swennen et al. [54]. In particular, the *ala* is defined as the most lateral point on each

Fig. 1. Intel RealSense SR300.

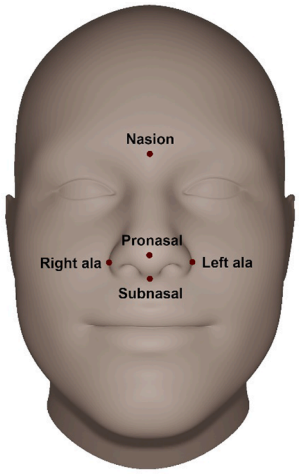

Fig. 2. The five considered landmarks.

alar contour, left and right. The *pronasal* is defined as the most anterior midpoint of the nasal tip; in case of a bifid nose, the *pronasal* is located on the most protruding tip. The *subnasal* is described as the midpoint on the nasolabial soft tissue contour between the columella crest, i.e., the plump external end of the nasal septum and the upper lip. Finally, the *nasion* is defined as the soft tissue midpoint contour of the base of the nasal root at the level of the frontonasal suture. The choice of these five landmarks is suitable for demonstrating in a simplified scenario the methodology, and it is representative of the instrument performances for some face analysis applications. In fact, facial movements due to expressions induce changes in landmark positions. This work considers a landmark subset marginally affected by this variability to demonstrate the proposed methodology for task-based uncertainty evaluation. The results will be directly relevant for biomedical and face recognition applications, although the analysis should be performed on the full considered dataset. Whilst other cases, such as FER, which inherently rely upon landmarks more affected by expression-induced facial movements, would only require the application of the methodology on a different set of landmarks, which can be addressed in future works. The critical positioning of some of the considered landmarks, namely the alae, adds significance to the analysis, catering for the position identification variability. Moreover, the above-mentioned definitions of landmarks allow for robust replication of landmarking amongst different operators.

2.3. Task-based metrological characterization methodology for face analysis applications

The scope of this work is evaluating the metrological performances of RGB-D camera for facial analysis applications. This will allow defining the operating range of the instrument and some good practices for that field.

However, it is first necessary to identify influence factors to measurement uncertainty. These are either due to the instrument working principle, e.g., the projected pattern and distance from the measurand, instrument metrological characteristics, e.g., resolution, noise and flatness deviation, or to the specific application. In the case of facial analysis, the latter includes the morphology of the face, the possible presence of occlusions, and the landmarking operation. Fig. 3 summarizes them through an Ishikawa diagram.

The discussion of Section 2.1 allows identifying several influence factors pertaining to the first category, i.e., the minimum pixel size and the distance from the measurand limit the lateral resolution. As far as the depth measurement is concerned, sensor resolution is relevant. Additionally, the projected pattern introduces some measurement variability, which is carried out on a single frame. Moreover, the depth offset and the flatness deviation introduce systematic errors that require calibration.

Furthermore, there are influence factors related to the specific task considered. These are the measurand, i.e., a face, which may present some occlusion or hamper interaction with the projected light pattern, e.g., because of facial hair, and the environment. In fact, measurements for facial analysis are hardly carried out in controlled conditions, and illumination may affect the measurement. Last, the landmarking method may induce some further variability.

Performances of the instrument can be evaluated according to several indicators. As mentioned in Section 2.1, a first index of the quality of the measured data set is the ratio between the number of non-measured points and the overall number of pixels:

$$p_{NM} = \frac{\sum_{i=1}^{n_x}\sum_{j=1}^{n_y} I(z(i,j))}{n_x n_y} \tag{1}$$

which is the percentage of non-measured points p_{NM}, where $I(NaN) = 1$ and 0 otherwise, being *NaN* the Not-available Number, i.e., the value of a non-measured point. Secondarily, a quantitative metrological assessment is necessary. Because of how the landmarks are identified and exploited, it is apparent that the standard calibration of RGB-D sensors is not sufficient. In fact, to propagate uncertainty of landmark positioning to their distance, a map describing the expanded uncertainty, U, pixel-by-pixel, i.e., $U = (U_x, U_y, U_z)$, is needed. In the following, a procedure is outlined to evaluate these performance indicators and to assess the statistical significance of the influence factors to the considered instrument.

2.3.1. Face measurement influence factors significance assessment

According to the previous discussion, a design of experiments (DOE) approach has been deployed to evaluate the significance of possible influence factors to facial measurements. Table 1 summarizes the implemented full factorial design.

Measurements are performed in a common room, against a white background, as this is the most common operating condition. The

Table 1
Considered factors and levels to assess the significance of influence factors to facial measurements.

Factor	Levels			
Measured subject	1-Female	2-Bearded Male	3-Male	
Room light	0- Pitch black			1- Artificial Light
Distance from the sensor/cm	25	40	55	70

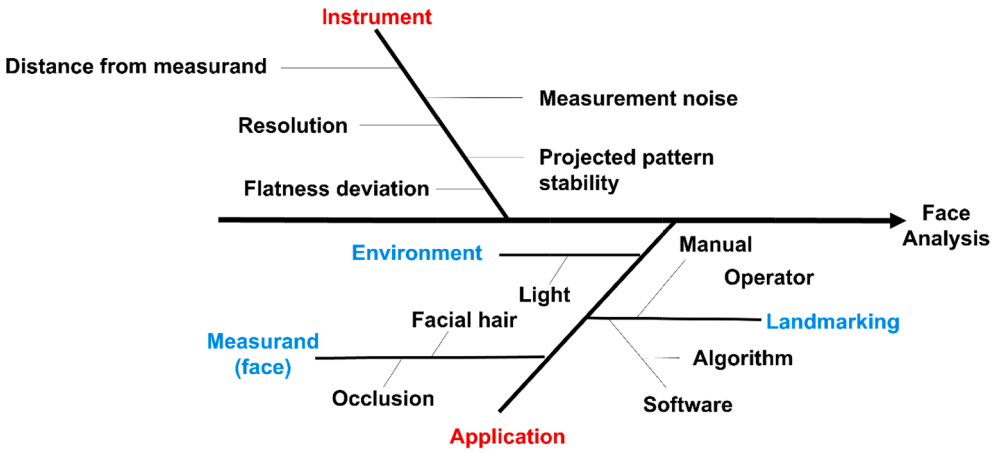

Fig. 3. Influence factors to facial analysis by RGB-D camera based on structured light. Red and cyan describe respectively first and second branches level.

possible effect of the measured subject is catered for measuring a young female and bearded male subject. The male subject was then asked to shave to include an additional level and directly compare the effect of facial hair presence on the same subject. The distance of the measured subject from the sensor is tested considering four equally spaced levels, designed based on experienced operators' knowledge in face analysis, who discourage the use at distances greater than 60 cm. Accordingly, the longer distances were chosen in the neighbourhood of that threshold. The room illumination is a factor included for verification scope. In fact, the considered instrument, i.e., the Intel RealSense SR300, projects a structured IR light pattern. Thus, it should be insensitive to external illumination. The DOE run order is randomized. Thirty replicated measurements per condition were performed to estimate reproducibility so that the DOE results in 720 data. The replicated measurements are obtained measuring a 30″ video at 30 fps.

The outputs that are considered are the p_{NM} and the standard uncertainty of the instrument along each measuring axis, i.e., $u_{instrum} = \left(u_{instrum,x}, u_{instrum,y}, u_{instrum,z}\right)$. However, the nature of those outputs hinders the application of the ANOVA [55,56]. In fact, p_{NM} asymptotically distributes according to a normal distribution $\mathcal{N}(\mu_{p_{NM}}, \sigma_{p_{NM}})$. Correct estimators of the distribution parameters are:

$$\widehat{\mu_{p_{NM}}} = p_{NM} \tag{2.1}$$

$$\widehat{\sigma_{p_{NM}}} = \sqrt{\frac{p_{NM}(1-p_{NM})}{n_x n_y}} \tag{2.2}$$

The DOE results in a set of 30 replicated measurements of p_{NM}, per each of the 24 factor combinations, each of those with their own dispersion, i.e., $P_f = \left\{\left(p_{NMf,r}, \widehat{\sigma_{p_{NM}f,r}}\right)\right\}_f, r = \{1,\cdots,30\}, f = \{1,\cdots,24\}$, $r, f \subset \mathbb{N}$. Consequently, the replications are realizations of the random variable p_{NMf} $\mathcal{N}(\mathbb{E}[p_{NMf}], \sigma_{p_{NM},wf})$. Therefore, the variability within each group of replicated measurements, $\sigma_{p_{NM},wf}$, has to be modelled as a mixture of variables [56]:

$$\widehat{\sigma_{p,w,f}} = \sqrt{\mathbb{V}ar\left[p_{NMf}\right] \bullet 30 + \mathbb{E}\left[\widehat{\sigma_{p_{NM}f}}^2\right]} \tag{3}$$

This modelling prevents straightforward ANOVA analysis application and requires performing pairwise hypothesis *t*-tests to investigate the presence of a systematic difference between the average of different factors combination. In particular, the null hypothesis, H_0, and the alternative hypothesis, H_1, are:

$$H_0 : p_{NMf_i} - p_{NMf_j} = 0 \tag{4.1}$$

$$H_1 : p_{NMf_i} - p_{NMf_j} \neq 0 \tag{4.2}$$

where f_i, $f_j \in f$. The test is carried out at a bilateral symmetric 95% confidence level and H_0 is rejected if.

$$\left(p_{NMf_i} - p_{NMf_j}\right) \notin \left[t_{0.025;29}s\,;\, t_{0.975;29}s\right] \tag{5}$$

where $t_{p,30}$ is the quantile of a Student's *t* distribution with 29 degrees of freedom identifying the cumulated probability p, and $s = u\left(p_{NMf_i} - p_{NMf_j}\right) = \sqrt{\widehat{\sigma_{p,wf_i}}^2 + \widehat{\sigma_{p,wf_j}}^2}$.

As mentioned, the second output is the $u_{instrum}$. This indicator is computed for each factor combination exploiting the replicated measurements. Registration of the 30 replicated measurements to the central one is performed by means of iterative closest point (ICP) algorithm [57,58]. ICP finds the roto-translation matrix to align a pair of surfaces by minimizing the sum of squared distances between matching pairs of points [57–59]. The registration is necessary to correctly match the features and compensate unavoidable head micromovements, otherwise increasing the variance artificially. The result of the registration consists of a set of 30 depth maps, $\boldsymbol{D_f} = z_f(x,y) = \left\{\left(x_{r,f}, y_{r,f}, z_{r,f}\left(x_{r,f}, y_{r,f}\right)\right)\right\}_f$. Although still arranged in a matrix points, the registered surfaces now share the same local coordinate systems, so that the variance of *x*-and *y*-absolute coordinates is not null. The ICP registration is essential because raw measured surfaces have the different local coordinate systems, which makes null the dispersion of the lateral coordinates. Therefore, after the registration has been carried out, per each factor combination, the measured points variance can be evaluated, enabling the mapping of the reproducibility of the depth map, i.e., $Repr_f = \left\{\left(\mathbb{V}ar\left[x_{r,f}\right], \mathbb{V}ar\left[y_{r,f}\right], \mathbb{V}ar\left[z_{r,f}\right]\right)\right\}_f$. This is exploited to evaluate the $u_{instrum}$, for a certain factor combination, as:

$$u_{instrumf} = \left\{u_{instrum_x}, u_{instrum_y}, u_{instrum_z}\right\}_f \tag{6.1}$$

$$u_{instrumf} = \mathbb{E}\left[\left\{\left(\mathbb{V}ar\left[x_{r,f}\right], \mathbb{V}ar\left[y_{r,f}\right], \mathbb{V}ar\left[z_{r,f}\right]\right)\right\}_f\right] + \left(u^2_{Res,x}, u^2_{Res,y}, u^2_{Res,z} + u^2_{Acc,z}\right) \tag{6.2}$$

which combines the reproducibility, the resolution and the accuracy of the instrument through the law of uncertainty propagation [60]. The resolution is modelled assuming a uniform distribution with a range equal to the instrument resolution [60]. The accuracy requires calibration. The results from [48] are exploited in this work, thus considering the accuracy as a type B contribution propagating with a uniform distribution.

However, managing the resulting standard uncertainty is critical. Conversely, its square, i.e. the variance of the measurement, distributes according to a Chi-square distribution [56]. However, in any case, the variable distribution hinders the application of ANOVA [55]. Therefore, a hypothesis test based on Fisher distribution is resorted to investigating the presence of systematic differences in the $u^2_{instrum}$ due to the considered factors. In particular, the hypothesis is:

$$H_0 : u^2_{instrum}f_i = u^2_{instrum}f_j \Leftrightarrow \frac{u^2_{instrum}f_i}{u^2_{instrum}f_j} = 1 \tag{7.1}$$

$$H_1 : u^2_{instrum}f_i \neq u^2_{instrum}f_j \Leftrightarrow \frac{u^2_{instrum}f_i}{u^2_{instrum}f_j} \neq 1 \tag{7.2}$$

The test is performed at a bilateral symmetric confidence level of 95%, and the null hypothesis is rejected if.

$$\frac{u^2_{instrum}f_i}{u^2_{instrum}f_j} \notin \left[F_{0.025;29,29}\,;\, F_{0.975;29,29}\right] \tag{8}$$

where $F_{p;29,29}$ is the quantile of the Fisher distribution with 29 degrees of freedom both at numerator and denominator identifying the cumulated probability *p*.

2.3.2. Landmarking influence factors significance assessment

Once the significance of influence factors pertaining to the measurement is evaluated, the face characterization, i.e., the landmarking, is addressed. In this work, as formerly discussed, manual landmarking is considered. Thus, the effect of operators carrying out this task is studied. A DOE with factors shown in Table 2 is considered.

The measured subjects are the same formerly considered. The measurements introduced in Section 2.3.1 are exploited. Only two distances from the sensors are considered to reduce the number of the tests. The

Table 2
Considered factors and levels to assess the significance of operators performing landmarking to facial analysis.

Factor	Levels		
Measured subject	1-Female	2-Bearded Male	3-Male
Distance from the sensor/cm	d_1		d_2
Operator	1	2	3

actual choice depends on the results of the DOE analysis discussed in Section 2.3.1. Three experienced operators are considered and are asked to place landmarks, according to the methodology outlined in Section 2.2, on three different consecutive frames of the measurement, namely the fourteenth, the central one, i.e., the fifteenth, and the sixteenth. Therefore, the DOE results in 18 factors combinations with 3 replication each, hence 54 data. The considered output is the cartesian coordinates of the five placed landmarks $L_l = (L_x, L_y, L_z)_l$, $l \in \{1, \cdots, 5\} \subset \mathbb{N}$. ANOVA is performed (at a 95% confidence level) to investigate if operators introduce systematic differences in the average estimated coordinate.

Additionally, ANOVA variance modelling allows estimating the reproducibility of the landmarking coordinate due to the operator, $Repr_{op}$. If the null hypothesis, i.e., no systematic differences are present, cannot be rejected, the reproducibility of the landmarking is the total variance. Conversely, if the null hypothesis has to be rejected, the variance within the group identified by the operator is considered.

2.3.3. Evaluation of measurement uncertainty

The evaluation of the task-based measurement uncertainty for the landmarking operation can be finally carried out. Per each landmark coordinate, $L_\blacksquare$, the following metrological model [60,61] is considered, in accordance with the highlighted influence factors from Fig. 3:

$$L_\blacksquare = L_\blacksquare + Acc + Res + Repr_{meas} + Repr_{op} \tag{9}$$

In Eq.(9), *Acc* is the instrument accuracy, is calibrated and includes the effect of the flatness deviation; *Res* is the resolution contribution, which is either vertical or lateral, depending on the considered coordinate. Two reproducibility components are included, one due to the facial measurement, which is modelled as $Repr_{meas} = Repr_f$, and the second due to the operator. All the above referenced terms, with the obvious exception of $L_\blacksquare$ which is an average value, are random variables with zero mean and (non-zero) standard deviation that represent the metrological characteristics of the system. The availability of a pixel-by-pixel map of the accuracy and measurement reproducibility avoids the association of average variability contribution evaluated on the whole measured surface [58,62].

Variability contributions, $u_w^2(L_\blacksquare)$, are combined and propagated according to the law of uncertainty propagation [60], which in the case at hand, allows evaluating the standard uncertainty of a landmark coordinate as:

$$u(L_\blacksquare) = \sqrt{\sum_w u_w^2(L_\blacksquare)} = \sqrt{u_{Acc}^2 + u_{Res}^2 + u_{Repr_{meas}}^2 + u_{Repr_{op}}^2} \tag{10}$$

In this work, the accuracy is included as a non-statistically evaluated contribution, i.e., a type B contribution, relying upon the calibration reported in Carfagni et al. [48]; accuracy includes bias and is propagated as a uniform distribution. The resolution is a type B contribution, propagated as a uniform distribution. The reproducibility of the face measurement and analysis represent the task-based effect on the measurement precision and are evaluated as statistical contribution, i.e., type A, as discussed in Section 2.3.1 and 2.3.2, which distribute with a normal distribution.

Expanded uncertainty is then evaluated at a confidence level of 95% by means of the coverage factor *k*, assuming a Student's *t* distribution, as:

$$U(L_\blacksquare) = k \cdot u(L_\blacksquare) \tag{11.1}$$

$$k = t_{0.975;\nu_{L_\blacksquare}} \tag{11.2}$$

where $\nu_{L_\blacksquare}$ are the degrees of freedom [60].

The evaluation of the expanded uncertainty of the landmark coordinates results in $U(L) = (U_{L_x}, U_{L_y}, U_{L_z})$, This result can be exploited to test the adequacy of the measurement with respect to the expected minimum requirements. In fact, the landmarking operation must be carried out with a tolerance of $t = 4$ mm, with respect to the theoretical landmark position [63–68]. The conformity to this specification can be investigated by performing a tolerance verification [69]. This requires comparing the specification interval centered in the landmark, i.e., $I_T = L \pm t = [LL_T, UL_T]$, and the uncertainty interval, i.e., $I_U = L \pm U(L) = [LL_U, UL_U]$. The verification compares the two interval and requires evaluating the acceptance interval:

$$I_A = [LL_A, UL_A] = I_T - 2gu(L_\blacksquare) = [LL_T + gu(L_\blacksquare), UL_T - gu(L_\blacksquare)] \tag{12.1}$$

$$UL_A - LL_A > 0 \tag{12.2}$$

The acceptance interval reduces the specification zone by a quantity proportional to the standard uncertainty of the measurement of a guard band factor *g*. Indeed, the acceptance zone must not be null, so that the condition to verify the compliance to tolerances:

$$UL_A - LL_A > 0 \Leftrightarrow 2t > 2gu(L_\blacksquare) \tag{13}$$

which is intended to be evaluated, independently, along each coordinate axis. Ultimately, the creation of the acceptance interval considers that its extreme values may be affected by uncertainty. Thus, the specification zone must be reduced to be conservative in the verification.

3. Results and discussion

3.1. Face measurement influence factors significance assessment

The first, the DOE introduced in Section 2.3.1 and aimed to identify relevant influence factors to facial measurement is performed. The measurements are performed with a field of view resolution of (640 × 480) pixel. This corresponds to the resolution, which depends on the distance from the sensor (see Section 2.1), on the coordinate axes that is reported in Table 3.

The percentage of non-measured points, p_{NM}, the measurement reproducibility map, $Repr_{meas}$, and the *MSE* are evaluated according to the methodology outlined in Section 2.3.1.

The analysis of the considered factors influences on the p_{NM} is performed by the discussed pairwise *t*-test, which can also be addressed through graphical representation, reported in Fig. 4. At a 95% confidence level, both the measured subject and the distance from the sensor introduce significant systematic differences in the percentage of non-measured points. In particular, the plot in Fig. 4 shows a qualitative quadratic effect on p_{NM} due to the distance. Also, the presence of facial hair in the measured subject 2 significantly increases the p_{NM}. The verification factor, i.e., the room light, does not have a significant effect on the output variable, as expected and consistently with the instrument working principle. The verification factor supports and validates the results of the implemented DOE.

Then, the evaluation of the point-by-point reproducibility ascribed to measurement influence factor, i.e., $Repr_{meas} = Repr_f$, is carried out and allows evaluating the $u_{instrum}$ along the three coordinate axes. The model in Eq. (6) is applied, where the accuracy of the instrument is propagated as a type B contribution distribution with a uniform distribution. Relying upon the calibration of [48], and considering that the measured faces were at the center of the field-of-view, a conservative estimation of a range of 1 mm could be considered. Resolution is also propagated as a type B contribution, associating a uniform distribution with a range

Table 3
Resolution of the coordinate axes at different distances from the sensor.

Distance from sensor/cm	Resolution/mm	
	x, y	z
25	0.548	0.001
40	0.800	
55	1.175	
70	1.714	

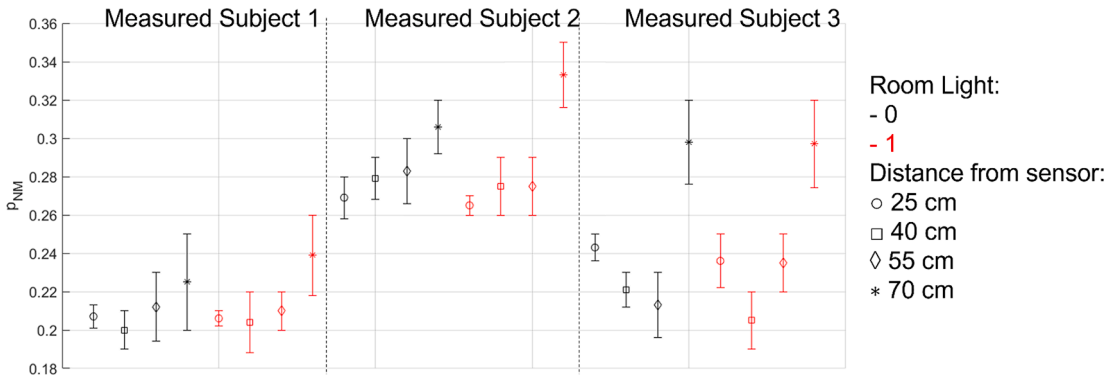

Fig. 4. Results of p_{NM} from the first DOE (see Table 1). The effect of influence factors to face measurement can be appreciated. Error bars represent 95% confidence intervals.

equal to the resolution step, as reported in Table 3.

The resulting $u^2_{instrum}$ is analyzed performing pairwise F-test at 95% confidence level, according to which the distance from the sensor is the only factor introducing significative differences between instrument reproducibility. Fig. 5 shows qualitative trends of the output depending on the considered influence factors, confirming the quantitative result, and suggesting a quadratic effect of the distance on the $u_{instrum}$.

The results show that, for facial measurements, the instrument performs best at short range and significantly worsens both in terms of non-measured points and reproducibility at distances greater than 55 cm, and the measured subject has a significant effect.

3.2. Landmarking influence factors significance assessment

These results are exploited to set up the second DOE, as discussed in Section 2.3.2, to investigate the influence factor to face analysis. In particular, the two investigated distances are 25 cm and 55 cm, being the extremities of the operative range. The five considered landmarks, i.e., the two *alae*, the *pronasal*, the *subnasal* and the *nasion*, are manually placed by three operators, as shown in Fig. 2, on three different frames. ANOVA analysis is performed to investigate the effect of the operators. Because of their statistically significant systematic effect on the face measurement, the measured subject and distance from the sensor are considered blocks of the experimental design, as their effect could overshadow the operators' variability. Consequently, individual ANOVA analyses tackling the sole operator's effect are performed, and Table 4 reports the results, highlighting only the conditions with a *p*-value lower than 5%. As it can be noticed, operators sometimes introduce a systematic difference in landmarking. This is severely significant at greater distances between the face and the sensor.

Accordingly, the variability due to the operator $Repr_{op}$ is estimated, in terms of variance $u^2_{Repr_{op}}$, as the variance within the operator groups, in the case a systematic difference is present, or as the total variance. It is worth remarking that the particular choice of how to analyze the second DOE enabled the evaluation of $u^2_{Repr_{op}}$ so that it does not include other contributions already present in $u^2_{Repr_{instrum}}$.

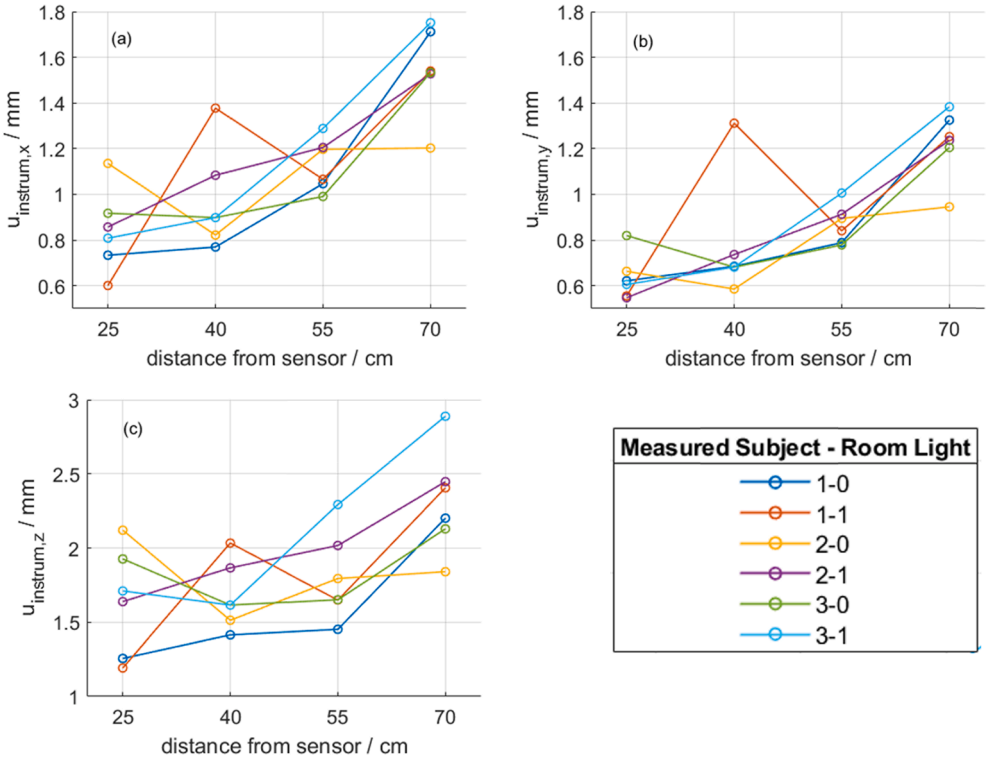

Fig. 5. Results of the instrument reproducibility $u_{instrum}$ along the (a) x-axis, (b) y-axis and (c) z-axis.

Table 4
Effect of operators on landmarking. *p*-value (in percentage) from ANOVA analysis in accordance with Section 2.3.2.

Measured Subject	1						2						3					
Distance from sensor/cm	25			55			25			55			25			55		
Coordinate axis	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
Right Ala					4.8												1.3	
Left Ala										1.8			2.8	3				
Nasion	4.8			4.2	3.3													
Pronasal								4					0.1	2.9				
Subnasal						1.8												

3.3. Evaluation of measurement uncertainty

The uncertainty of the landmarking operation is then propagated according to Eq. (10), and the expanded uncertainty is computed. Table 5 reports the expanded uncertainty of the landmarking operation, evaluated at a 95% confidence level with a coverage factor equal to 2, assuming 30 degrees of freedom in accordance with the P.U.Ma method [70]. The analysis of uncertainty propagation shows that the most relevant contribution is the variability due to the operators.

The task-based uncertainty evaluation allows performance verification. The evaluation of the expanded uncertainty of the landmark coordinates describes an elliptical envelope within which the landmark can be positioned considering the used measuring instrument and the applied face analysis method, as shown in Fig. 6. The elliptical envelope maximizes the uncertainty volume with respect to a right prism [71].

This result is exploited for tolerance verification in compliance with ISO 14253-1:2017 [69]. The results of the verification of Eq. (13) are reported in Table 6.

Table 6 indicates clearly that the instrument suitability for facial measurement and analysis depends significantly on the measured subject and the distance from the sensor. In particular, the presence of facial hair (measured subject no. 2) at a larger distance hinders performing precise measurements. Criticalities are present for *alae* *z*-coordinate. This can be explained by the fact that small lateral variation in the alae region may lead to significant changes in the *z*-direction, as it may involve moving from the intersection between the nose and the cheek to the nose side, for example. Consequently, landmarks positioned on highly sloped regions are liable of being estimated with greater uncertainty, as in the *alae* cases. In the case at hand, the operators were the largest contribution to measurement uncertainty, despite not being significantly different in their behavior, as reported in Table 4, result that is consistent with the large variability induced by the inherent difficulty of landmarking positioning.

4. Conclusions

This work developed a task-based evaluation of measurement uncertainty for face analysis based on customer grade 3D cameras. A methodology was proposed to evaluate significant influence factors and propagate uncertainty contribution. The method was applied to a case study to metrologically qualify performances of Intel RealSense SR300 RGB-D camera, a coded-light device typically adopted for short-range applications. Results show that significant effects on the measurements are introduced by the measured subject, the distance from the sensor and the operator performing the landmarking. The latter is relevant when manual landmarking is performed, which at the state-of-the-art is necessary to define training and calibration datasets for automatic analysis software, and for specific tasks such as annotation of cephalometric points in pre-surgery planning. The evaluation of uncertainty formally verifies the adequacy of the considered camera to face analysis. Moreover, the work outlines guidelines and best practices that can help design applications in several fields that require face analysis, i.e., industrial, medical, marketing. In particular optimal performances, considering physical measurement setup and flexibility and metrological properties, are short distances. A suitable best practice may suggest working at 40 cm, provided the evaluated instrument standard uncertainty. The following considerations can explain this trade-off: lower distances may suffer from artefacts due to the vertical bars of the patterns at different spatial frequencies and the variability caused by the temporal multiplexing used to generate the depth map from the codewords. Facial applications imply the presence of humans, whose micromovements in front of the camera are unavoidable and deepen the risk of introducing errors in the depth generation phase. On the other hand, the loss of resolution rapidly worsens when increasing the distance between the camera and the subject in such a way that an acceptable 3D facial reconstruction is no longer possible beyond 55 cm.

The development of consumer-grade depth cameras has made accessible machine vision technologies and methods, speeding up its diffusion in various applications, including face analysis and, particularly, face recognition. The methodology outlined in this study can be applied in future work to compare performances of different measuring instruments and to evaluate performances of automatic landmarking software.

CRediT authorship contribution statement

Giacomo Maculotti: Conceptualization, Methodology, Formal analysis, Software, Investigation, Visualization, Writing – original draft. **Luca Ulrich:** Conceptualization, Methodology, Data curation, Software, Investigation, Visualization, Writing – original draft. **Elena Carlotta Olivetti:** Investigation, Writing – original draft. **Gianfranco Genta:** Validation, Supervision, Writing – review & editing. **Federica Marcolin:** Investigation, Supervision, Writing – review & editing. **Enrico Vezzetti:** Supervision, Resources. **Maurizio Galetto:** Funding

Table 5
Expanded uncertainty of landmarking.

Measured Subject	1						2						3					
Distance from sensor/cm	25			55			25			55			25			55		
$U(L_\bullet)$/mm	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
Right Ala	1.8	1.3	1.6	3.3	2.2	4.3	1.9	1.4	6.6	2.1	4.3	4.7	0.5	1.1	1.4	5.4	4.0	8.1
Left Ala	1.3	2.4	4.2	5.2	4.1	6.6	0.9	2.3	3.4	2.6	2.8	9.9	1.4	2.7	5.0	2.9	1.7	5.2
Nasion	1.2	1.1	1.3	3.0	1.3	1.7	1.5	0.8	2.8	1.9	4.7	5.4	1.7	1.4	1.3	2.6	2.0	1.9
Pronasal	0.9	1.5	1.6	3.3	3.6	2.9	1.5	1.1	3.2	2.0	3.6	6.0	0.5	0.8	1.2	1.6	3.2	2.3
Subnasal	1.6	1.6	2.4	3.7	2.9	1.7	1.3	2.5	3.1	3.6	4.2	9.7	1.3	1.8	3.9	1.3	1.3	2.6

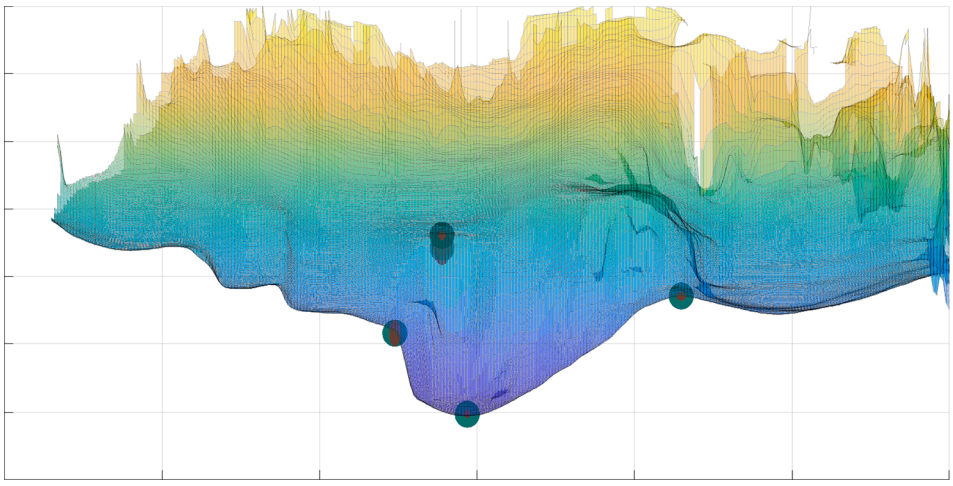

Fig. 6. Uncertainty envelope. Around the positioned landmarks, the tolerance zone (green) and the uncertainty (red) can be seen.

Table 6
Verification of compliance specification. Only coordinate out of the acceptance zone are marked as *NO*.

Measured Subject	1						2						3					
Distance from sensor/cm	25			55			25			55			25			55		
Coordinate axis	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*	*x*	*y*	*z*
Right Ala						NO			NO		NO	NO				NO		NO
Left Ala			NO	NO		NO					NO	NO			NO			NO
Nasion											NO	NO						
Pronasal												NO						
Subnasal											NO	NO						

acquisition, Project administration, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been partially supported by “Ministero dell’Istruzione, dell’Università e della Ricerca”. Award “TESUN- 83486178370409 finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6”. The authors would like to acknowledge the project “3D-Surgeered - Pre-surgery planning by means of 3D modelling” of Department of Management and Production Engineering of Politecnico di Torino. The authors would like to thank Mr. Michele Spadaro for his contribution to data analysis.

References

[1] I. Adjabi, A. Ouahabi, A. Benzaoui, A. Taleb-Ahmed, Past, present, and future of face recognition: A review, Electron. 9 (2020) 1–53, https://doi.org/10.3390/electronics9081188.
[2] J. Wang, Y. Liu, Y. Hu, H. Shi, T. Mei, FaceX-Zoo: A PyTorch Toolbox for Face Recognition, (2021). http://arxiv.org/abs/2101.04407.
[3] V. Albiero, K.S. Krishnapriya, K. Vangara, K. Zhang, M.C. King, K.W. Bowyer, Analysis of Gender Inequality in Face Recognition Accuracy, Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vis. Work. WACVW 2020. (2020) 81–89. https://doi.org/10.1109/WACVW50321.2020.9096947.
[4] D. Dupré, E.G. Krumhuber, D. Küster, G.J. McKeown, S. D’Mello, A performance comparison of eight commercially available automatic classifiers for facial affect recognition, PLoS One. 15 (4) (2020) e0231968, https://doi.org/10.1371/journal.pone.0231968.
[5] X. Liu, Y. Zou, H. Kuang, X. Ma, Face image age estimation based on data augmentation and lightweight convolutional neural network, Symmetry (Basel). 12 (1) (2020) 146.
[6] L. Ulrich, J.-L. Dugelay, E. Vezzetti, S. Moos, F. Marcolin, Perspective morphometric criteria for facial beauty and proportion assessment, Appl. Sci. 10 (1) (2020) 8, https://doi.org/10.3390/app10010008.
[7] A. Jones, B. Jaeger, Biological bases of beauty revisited: The effect of symmetry, averageness, and sexual dimorphism on female facial attractiveness, Symmetry (Basel). 11 (2) (2019) 279.
[8] K.S. Kaya, B. Türk, M. Cankaya, N. Seyhun, B.U. Coşkun, Assessment of facial analysis measurements by golden proportion, Braz. J. Otorhinolaryngol. 85 (4) (2019) 494–501, https://doi.org/10.1016/j.bjorl.2018.07.009.
[9] E.C. Olivetti, S. Nicotera, F. Marcolin, E. Vezzetti, J.P. Jacqueline, E. Zavattero, G. Ramieri, 3D Soft-tissue prediction methodologies for orthognathic surgery-a literature review, Appl. Sci. 9 (2019) 1–23, https://doi.org/10.3390/app9214550.
[10] A. Sam, K. Currie, H. Oh, C. Flores-Mir, M. Lagravére-Vich, Reliability of different three-dimensional cephalometric landmarks in cone-beam computed tomography: A systematic review, Angle Orthod. 89 (2019) 317–332, https://doi.org/10.2319/042018-302.1.
[11] I. Nasseh, W. Al-Rawi, Cone Beam Computed Tomography, Dent. Clin. North Am. 62 (3) (2018) 361–391, https://doi.org/10.1016/j.cden.2018.03.002.
[12] H. Gaêta-Araujo, T. Alzoubi, K. de Faria Vasconcelos, K. Orhan, R. Pauwels, J. W. Casselman, R. Jacobs, Cone beam computed tomography in dentomaxillofacial radiology: A two-decade overview, Dentomaxillofacial Radiol. 49 (2020) 1–20, https://doi.org/10.1259/DMFR.20200145.
[13] S.L. Assouline, C. Meyer, E. Weber, B. Chatelain, A. Barrabe, N. Sigaux, A. Louvrier, How useful is intraoperative cone beam computed tomography in maxillofacial surgery? An overview of the current literature, Int. J. Oral Maxillofac. Surg. 50 (2) (2021) 198–204, https://doi.org/10.1016/j.ijom.2020.05.006.
[14] M. Pons, J.-C. Lutz, B. Chatelain, E. Weber, A. Barrabé, C. Meyer, N. Sigaux, A. Louvrier, Impact of intraoperative cone beam computed tomography in the management of zygomatic fractures, J. Stomatol. Oral Maxillofac. Surg. 122 (4) (2021) 349–354, https://doi.org/10.1016/j.jormas.2020.11.005.
[15] E. Robotti, R.K. Daniel, F. Leone, Cone-Beam Computed Tomography: A User-Friendly, Practical Roadmap to the Planning and Execution of Every Rhinoplasty - A 5-Year Review, Plast. Reconstr. Surg. (2021) 749E–762E, https://doi.org/10.1097/PRS.0000000000007900.
[16] D.T. van der Haar, Student Emotion Recognition in Computer Science Education: A Blessing or Curse?, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 11590 LNCS (2019) 301–311. https://doi.org/10.1007/978-3-030-21814-0_23.
[17] G. Tonguç, B. Ozaydın Ozkara, Automatic recognition of student emotions from facial expressions during a lecture, Comput. Educ. 148 (2020) 103797, https://doi.org/10.1016/j.compedu.2019.103797.
[18] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, J.R. Movellan, The faces of engagement: Automatic recognition of student engagement from facial

expressions, IEEE Trans. Affect. Comput. 5 (1) (2014) 86–98, https://doi.org/10.1109/TAFFC.2014.2316163.
[19] S.N. Karimah, S. Hasegawa, Automatic Engagement Recognition for Distance Learning Systems : A Literature Study of Engagement Datasets and Methods, Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-78114-9.
[20] I. Spivak, S. Krepych, V. Faifura, S. Spivak, Methods and tools of face recognition for the marketing decision making, in: 2019 IEEE Int. Sci. Conf. Probl. Infocommunications Sci. Technol. PIC S T 2019 - Proc. (2019) 212–216. https://doi.org/10.1109/PICST47496.2019.9061229.
[21] E. Pantano, Non-verbal evaluation of retail service encounters through consumers' facial expressions, Comput. Human Behav. 111 (2020) 106448, https://doi.org/10.1016/j.chb.2020.106448.
[22] M.R. González-Rodríguez, M.C. Díaz-Fernández, C. Pacheco Gómez, Facial-expression recognition: An emergent approach to the measurement of tourist satisfaction through emotions, Telemat. Informatics. 51 (2020) 101404, https://doi.org/10.1016/j.tele.2020.101404.
[23] E. Carlotta Olivetti, M.G. Violante, E. Vezzetti, F. Marcolin, B. Eynard, Engagement evaluation in a virtual learning environment via facial expression recognition and self-reports: A preliminary approach, Appl. Sci. 10 (1) (2020) 314, https://doi.org/10.3390/app10010314.
[24] I. Maurtua, A. Ibarguren, J. Kildal, L. Susperregi, B. Sierra, Human-robot collaboration in industrial applications: Safety, interaction and trust, Int. J. Adv. Robot. Syst. 14 (2017) 1–10, https://doi.org/10.1177/1729881417716010.
[25] R. Inam, K. Raizer, A. Hata, R. Souza, E. Forsman, E. Cao, S. Wang, Risk Assessment for Human-Robot Collaboration in an automated warehouse scenario, IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA. (2018-Septe (2018)) 743–751, https://doi.org/10.1109/ETFA.2018.8502466.
[26] M. Zlatanski, P. Sommer, F. Zurfluh, S.G. Zadeh, A. Faraone, N. Perera, Machine Perception Platform for Safe Human-Robot Collaboration, Proc. IEEE Sensors. (2019), https://doi.org/10.1109/SENSORS43011.2019.8956547.
[27] J. Deng, G. Pang, Z. Zhang, Z. Pang, H. Yang, G. Yang, CGAN Based Facial Expression Recognition for Human-Robot Interaction, IEEE Access. 7 (2019) 9848–9859, https://doi.org/10.1109/ACCESS.2019.2891668.
[28] D. McColl, A. Hong, N. Hatakeyama, G. Nejat, B. Benhabib, A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI, J. Intell. Robot. Syst. Theory Appl. 82 (1) (2016) 101–133, https://doi.org/10.1007/s10846-015-0259-2.
[29] R. Gervasi, F.N. Digiaro, Mastrogiacomo L, D.A. Maisano, F. Franceschini, Comparing quality profiles in human-robot collaboration: Empirical evidence in the automotive sector, in: 4th Int. Conf. Qual. Eng. Manag. ICQEM 2020, Braga, 2020: pp. 79–104. https://www-scopus-com.ezproxy.biblio.polito.it/record/display.uri?eid=2-s2.0-85093842708&origin=resultslist&sort=plf-f&src=s&sid=9e055edefcaf9a2c8e1674a761250bca&sot=b&sdt=b&sl=115&s=TITLE-ABS-KEY%28Comparing+quality+profiles+in+human-robot+collaboration%3A+Empirical+evidence+in+the+automotive+sector%29&relpos=0&citeCnt=0&searchTerm= (accessed April 2, 2021).
[30] R. Gervasi, L. Mastrogiacomo, F. Franceschini, A conceptual framework to evaluate human-robot collaboration, Int. J. Adv. Manuf. Technol. 108 (3) (2020) 841–865, https://doi.org/10.1007/s00170-020-05363-1.
[31] M.E. Reyes, I.V. Meza, L.A. Pineda, Robotics facial expression of anger in collaborative human–robot interaction, Int. J. Adv. Robot. Syst. 16 (2019) 1–13, https://doi.org/10.1177/1729881418817972.
[32] S. Tornincasa, E. Vezzetti, S. Moos, M. Violante, F. Marcolin, N. Dagnes, L. Ulrich, G. Tregnaghi, 3D facial action units and expression recognition using a crisp logic, Comput. Aided. Des. Appl. 16 (2) (2019) 256–268.
[33] J.J. Bannister, S.R. Crites, J.D. Aponte, D.C. Katz, M. Wilms, O.D. Klein, F.P. J. Bernier, R.A. Spritz, B. Hallgrímsson, N.D. Forkert, Fully automatic landmarking of syndromic 3D facial surface scans using 2D images, Sensors (Switzerland). 20 (2020) 1–14, https://doi.org/10.3390/s20113171.
[34] M.A. De Jong, P. Hysi, T. Spector, W. Niessen, M.J. Koudstaal, E.B. Wolvius, M. Kayser, S. Böhringer, Ensemble landmarking of 3D facial surface scans, Sci. Rep. 8 (2018) 1–11, https://doi.org/10.1038/s41598-017-18294-x.
[35] M. Safeea, P. Neto, Minimum distance calculation using laser scanner and IMUs for safe human-robot interaction, Robot. Comput. Integr. Manuf. 58 (2019) 33–42, https://doi.org/10.1016/j.rcim.2019.01.008.
[36] D.S. Tran, N.H. Ho, H.J. Yang, E.T. Baek, S.H. Kim, G. Lee, Real-time hand gesture spotting and recognition using RGB-D Camera and 3D convolutional neural network, Appl. Sci. 10 (2020) 722, https://doi.org/10.3390/app10020722.
[37] S. Gatesichapakorn, J. Takamatsu, M. Ruchanurucks, ROS based Autonomous Mobile Robot Navigation using 2D LiDAR and RGB-D Camera, in: 2019 1st Int. Symp. Instrumentation, Control. Artif. Intell. Robot. ICA-SYMP 2019, Institute of Electrical and Electronics Engineers Inc., 2019: pp. 151–154. https://doi.org/10.1109/ICA-SYMP.2019.8645984.
[38] W. Liu, F. Li, C. Jing, Y. Wan, B. Su, M. Helali, Recognition and location of typical automotive parts based on the RGB-D camera, Complex Intell. Syst. 7 (4) (2021) 1759–1765, https://doi.org/10.1007/s40747-020-00182-z.
[39] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, Q. Zhang, Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review, Comput. Electron. Agric. 177 (2020) 105687, https://doi.org/10.1016/j.compag.2020.105687.
[40] Z. Zhang, Microsoft kinect sensor and its effect, IEEE Multimed. 19 (2) (2012) 4–10, https://doi.org/10.1109/MMUL.2012.24.
[41] L. Ulrich, E. Vezzetti, S. Moos, F. Marcolin, Analysis of RGB-D camera technologies for supporting different facial usage scenarios, Multimed. Tools Appl. 79 (39-40) (2020) 29375–29398, https://doi.org/10.1007/s11042-020-09479-0.
[42] V.L. Tran, H.Y. Lin, Accurate RGB-D camera based on structured light techniques, in: Proc. - 2017 Int. Conf. Syst. Sci. Eng. ICSSE 2017, Institute of Electrical and Electronics Engineers Inc., 2017: pp. 235–238. https://doi.org/10.1109/ICSSE.2017.8030872.
[43] H.M. Kim, M.S. Kim, G.J. Lee, H.J. Jang, Y.M. Song, Miniaturized 3D depth sensing-based smartphone light field camera, Sensors (Switzerland). 20 (2020) 2129, https://doi.org/10.3390/s20072129.
[44] Y. Bengio, I. Goodfellow, A. Courville, Deep learning, MIT Press, Cambridge, MA, 2016.
[45] K.P. Murphy, Machine Learning: A Probabilistic Perspective, The MIT Press, Cambridge, MA (2012), https://doi.org/10.1007/978-94-011-3532-0_2.
[46] D. Barber, Machine Learning A Probabilistic Approach (2006).
[47] A. Corti, S. Giancola, G. Mainetti, R. Sala, A metrological characterization of the Kinect V2 time-of-flight camera, Rob. Auton. Syst. 75 (2016) 584–594, https://doi.org/10.1016/j.robot.2015.09.024.
[48] M. Carfagni, R. Furferi, L. Governi, M. Servi, F. Uccheddu, Y. Volpe, On the Performance of the Intel SR300 Depth Camera: Metrological and Critical Characterization, IEEE Sens. J. 17 (14) (2017) 4508–4519, https://doi.org/10.1109/JSEN.2017.2703829.
[49] A. Zabatani, V. Surazhsky, E. Sperling, S. Ben Moshe, O. Menashe, D.H. Silver, Z. Karni, A.M. Bronstein, M.M. Bronstein, R. Kimmel, Intel® RealSense™ SR300 Coded Light Depth Camera, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2020) 2333–2345, https://doi.org/10.1109/TPAMI.2019.2915841.
[50] K. He, C. Sui, C. Lyu, Z. Wang, Y. Liu, 3D reconstruction of objects with occlusion and surface reflection using a dual monocular structured light system, Appl. Opt. 59 (29) (2020) 9259, https://doi.org/10.1364/AO.402146.
[51] H. Ding, P. Zhou, R. Chellappa, Occlusion-adaptive deep network for robust facial expression recognition, IJCB 2020 - IEEE/IAPR Int. Jt. Conf. Biometrics. (2020), https://doi.org/10.1109/IJCB48548.2020.9304923.
[52] E. Vezzetti, F. Marcolin, G. Fracastoro, 3D face recognition: An automatic strategy based on geometrical descriptors and landmarks, Rob. Auton. Syst. 62 (12) (2014) 1768–1776, https://doi.org/10.1016/j.robot.2014.07.009.
[53] N. Dagnes, F. Marcolin, F. Nonis, S. Tornincasa, E. Vezzetti, 3D geometry-based face recognition in presence of eye and mouth occlusions, Int. J. Interact. Des. Manuf. 13 (4) (2019) 1617–1635, https://doi.org/10.1007/s12008-019-00582-7.
[54] G.R.J. Swennen, F. Schutyser, J.-E. Hausamen (Eds.), Three-Dimensional Cephalometry, Springer, Berlin, Heidelberg, 2006.
[55] D.C. Montgomery, Design and Analysis of Experiments, 5th ed., John Wiley & Sons Inc., New York, 2013. https://doi.org/10.1007/978-3-030-58292-0_130690.
[56] D. Montgomery, G. Runger, N. Hubele, Engineering statistics, John Wiley & Sons Inc., New York, 2010.
[57] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[58] G. Genta, G. Maculotti, Uncertainty evaluation of small wear measurements on complex technological surfaces by machine vision-aided topographical methods, CIRP Ann. - Manuf. Technol. 70 (1) (2021) 451–454.
[59] G. Maculotti, N. Senin, O. Oyelola, M. Galetto, A. Clare, R. Leach, Multi-sensor data fusion for the characterisation of laser cladded cermet coatings, in: Eur. Soc. Precis. Eng. Nanotechnology, Conf. Proc. - 19th Int. Conf. Exhib. EUSPEN 2019. (2019) 260–263.
[60] JCGM100:, Evaluation of measurement data — Guide to the expression of uncertainty in measurement (GUM), JCGM. (2008) Sèvres, France. https://doi.org/10.1373/clinchem.2003.030528.
[61] M. Galetto, L. Mastrogiacomo, D. Maisano, F. Franceschini, Cooperative fusion of distributed multi-sensor LVM (Large Volume Metrology) systems, CIRP Ann. - Manuf. Technol. 64 (1) (2015) 483–486, https://doi.org/10.1016/j.cirp.2015.04.003.
[62] L. Carli, G. Genta, A. Cantatore, G. Barbato, L. De Chiffre, R. Levi, Uncertainty evaluation for three-dimensional scanning electron microscope reconstructions based on the stereo-pair technique, Meas. Sci. Technol. 22 (3) (2011) 035103, https://doi.org/10.1088/0957-0233/22/3/035103.
[63] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich, N. Dagnes, 3D geometry-based automatic landmark localization in presence of facial occlusions, Multimed. Tools Appl. 77 (11) (2018) 14177–14205, https://doi.org/10.1007/s11042-017-5025-y.
[64] J. Lei, X. You, M. Abdel-Mottaleb, Automatic Ear Landmark Localization, Segmentation, and Pose Classification in Range Images, IEEE Trans. Syst. Man, Cybern. Syst. 46 (2) (2016) 165–176, https://doi.org/10.1109/TSMC.2015.2452892.
[65] J.M.H. Noothout, B.D. De Vos, J.M. Wolterink, E.M. Postma, P.A.M. Smeets, R.A. P. Takx, T. Leiner, M.A. Viergever, I. Isgum, Deep Learning-Based Regression and Classification for Automatic Landmark Localization in Medical Images, IEEE Trans. Med. Imaging. 39 (12) (2020) 4011–4022, https://doi.org/10.1109/TMI.2020.3009002.
[66] X. Fan, Q. Jia, K. Huyan, X. Gu, Z. Luo, 3D facial landmark localization using texture regression via conformal mapping, Pattern Recognit. Lett. 83 (2016) 395–402, https://doi.org/10.1016/j.patrec.2016.07.005.
[67] E. Vezzetti, F. Marcolin, 3D landmarking in multiexpression face analysis: A preliminary study on eyebrows and mouth, Aesthetic Plast. Surg. 38 (4) (2014) 796–811, https://doi.org/10.1007/s00266-014-0334-2.
[68] M. Urschler, T. Ebner, D. Štern, Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization, Med. Image Anal. 43 (2018) 23–36, https://doi.org/10.1016/j.media.2017.09.003.
[69] ISO 14253:2017, Geometrical product specifications (GPS) — Inspection by measurement of workpieces and measuring equipment - Part 1: Decision rules for

verifying conformity or non conformity with specifications, Int. Organ. Stand. (2017).
[70] ISO 14253-2:2011, Geometrical product specifications (GPS) – Inspection by measurement of workpieces and measuring equipment – Part 2: Guidance for the estimation of uncertainty in GPS measurement, in calibration of measuring equipment and in product verification, Int. Organ. Stand. (2011) 71. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53631.
[71] ISO 1101:2017, Geometrical Product Specifications (GPS) - Geometrical Toelrancing - Tolerances of form, orientation, location and run-out, (2017) 158. https://aka.ihsmarkit.com/8ae4c908-b244-4bea-8491-673243373338.