3D Face Analysis vs Applications: Studies and Perspectives

Enrico VEZZETTI, Federica MARCOLIN, Sandro MOOS, Mariagrazia VIOLANTE, Luca ULRICH
Dipartimento di Ingegneria Gestionale e della Produzione, Politecnico di Torino, Turin, Italy

https://doi.org/10.15221/19.230

Abstract

In the present scenario biometry is becoming a very important issue. Starting from the marketing domain where business man are trying to capture people emotions when they buy a product in the supermarket, till arriving to border security, or event security, where companies are trying to propose automatic solutions able to identify critical situations, and be sure that the identification of the person is reliable, also in presence of camouflages, the involvement of biometry is growing up very quickly, involving finger prints, iris recognition, voice recognition but especially human face analysis. In this specific domain many improvements have been done, due to the fact that face has many distinctive issues, and interesting results have been obtained, especially working with 3D models, rather than simple 2D images more fragile for the significant influence due to the working conditions, that are able to provide more reliable results and solve problems connected also with people age. Starting from these challenges, from the more easy availability of low cost 3D sensors, also located on commercial mobile devices, and from the experience acquired in the Politecnico di Torino 3D Lab (www.3dlab.polito.it), this papers wants to presents a review of the main outcomes coming from 3D face morphometric applications, providing a systematic synthesis of the different available methods and approaches in order to guide potential users in simply identifying the most suitable solution and technology in relation to the specific application and working conditions, providing some first issues on strengths and weaknesses.

Keywords: Face Analysis, 3D Modelling, 3D Scanning, Human Body, Morphometry

1. Introduction

In recent years a considerable number of applications have benefited from the usage of the third dimension [1]; there are several research fields in which 3D is currently successfully used: safety, such as for autonomous driving [2]; orthopedics, for both diagnosis and treatment planning [3]; surgery, as 3D models reconstruction gives the possibility of organizing medical equipment [4], attending the surgeon during the intervention and supporting the post-operative evaluation of the results [5]; 3D printing applications [6], including facial prosthesis [7], dental implants [8] and pelvis prosthesis [9].
The ambition of accelerating the evolution process of cities into interconnected communities brings out other application areas as candidates for heavy 3D usage: security, such as the screening monitoring system in public places such as airports [10], land surveying [11], architecture [12] and archaeology [13] for research and tourism purposes.
Despite the idea has come out in '90s, the concept of "smart city" has been extensively examined in last decade [14] [15] [16] [17], thanks to the increasing possibility of being connected to the rest of the world; an example is the Singapore case [18]. It is reasonable to expect an increment of the number of smart cities and the strengthening of the existing ones, such as Amsterdam [19], Dubai [20] [21] and Dohlera [22], taking advantage from the Internet coverage diffusion.
Robustness has been mentioned above to highlight the possibility of operating in critical lighting conditions, since active 3D technologies work with infrared light or other sources of light outside the visible spectrum, in presence of occlusions [23] [24] and regardless of the orientation of the object or, most likely, of the subject [25].
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.
Literature about 3D is varied and fragmented due to lack of a shared methodology for analyzing the field and developing new applications. This scientific survey has been conducted to converge on a unique standard and to provide a baseline for the design of 3D facial applications.

2. 3D Face Analysis Applications Scenario

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” [26]. Human brain has the ability of figuring out characteristics such as identity, age, sex and mood [29] [30] [31], a skill that infants already possess from birth [27] and develops during growth [28].

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 face perception process has come up. Nevertheless, human brain functioning is highly complex and nowadays the possibility of reconstructing a model able to reply 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.

At present the main 3D facial applications could be considered: *Face Detection, Face Recognition*, with the two declinations in *Face Authentication* and *Face Identification*, and *Face Expression Recognition*.

Face Detection aims to detect a face shape inside an image or inside a frame in the case of a video stream. There exists various 2D techniques [32], 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 [33]. 3D techniques [34] [35] [36] benefit of the intrinsic advantages in using the depth information, such as lighting, pose and occlusion independence. This is the family of methods considered in this paper for the sake of robustness, an essential characteristic for real-time video data streams.

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 [37] and it is not uncommon that information obtained with different biometric systems are often fused together to further improve robustness [38]. Fingerprints [39] and iris scans [40] 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 high-quality images to analyse facial features and to minimize false positives and false negatives.

Face Identification refers to that variety of applications performing Face Recognition without authentication purposes. Some examples can be found in the fields of security, for criminals identification [42], marketing, to target specific customers or at least some of their features such as age and gender [43], 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 [44]. 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, especially those ones in motion, and an adequately wide operating range to accomplish the target if working on a video data flow.

Face Expression Recognition aims to understand human emotion by observing different parts of the face and analyzing the Action Units proposed by Paul Ekman in his works [45] [46]. Face Expression Recognition is a critical task since some expressions are ambiguous and difficult to be recognized even by a human observer. Landmarks analysis is the basis of this application, so input images quality should offer a resolution good enough to identify them and to perform further analysis.

The need of such an application is due to the spread of the concept of human-computer interaction in a variety of fields [47]: marketing [48], affect sensitive televisions [49], videogames [50], psychiatry [51]. One of the most important fields of application is robotics, because of the increasing demand of having robots able to understand human’s mood, so that they can become part of human’s life fully-fledged.

3. 3D Devices Benchmarking

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 projected 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. 1).

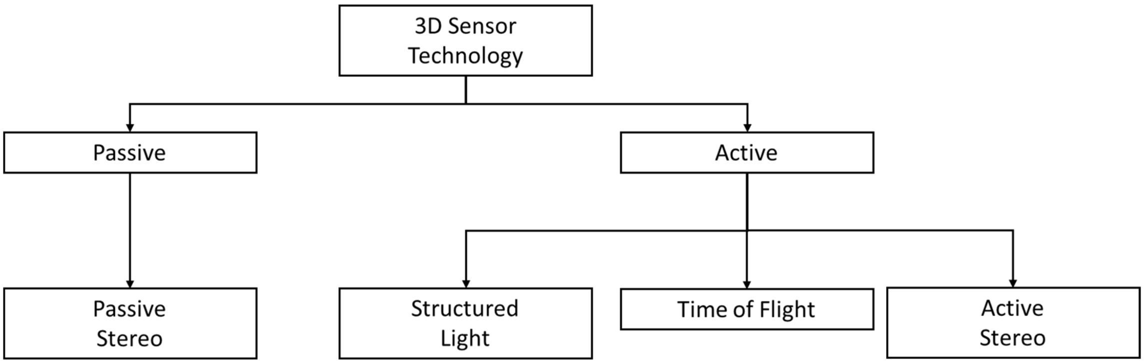

Fig. 1. 3D Sensor Technology

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 [52]. 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 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. Nevertheless, if the scene is featureless, the occlusions will not exist, but a solution for the correspondence problem will not be easy to find, so a suitable scene is required to use properly this technology.

Structured light depth cameras have been studied to overcome the issue of reliability of correspondences [53]. 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 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 transmitter. 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.

Time of Flight (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 framerate, making them suitable for real-time applications [54].

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 [55]. 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.

3. Results and Discussion

A benchmarking among 3D sensors 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
- *Framerate*: 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*: this parameter indicates the part of the scene visible through the sensor
- *Size*: sensors dimensions.

For each Face Analysis application explained in the section 2 some guidelines have been produced in order to provide a starting point for identifying the most suitable solution in relation with the face analysis specific application domain. In Face Detection the analysis accuracy is less important than in the other facial applications since it is easier to detect a face in a scene than to perform identification algorithms. Moreover, flexibility should be a strength point for this application, so that it can work in all range, light, pose and occlusions situations. 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 guarantee a good unblocking mechanism. Face authentication 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 relevant since Face Identification is different from Face Authentication. Face expression recognition shows a scenario very similar to Face Identification in relation to the fact that the working conditions are very closed each other. From the results coming from the analysis it is possible to report that ToF cameras are the best in terms of long range operating functioning, 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 high resolution and remarkable maximum operating functioning distance, followed by the active stereo technology and, in third position, by structured-light cameras, because of their too poor maximum operating functioning distance. 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. 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 each other. Active stereoscopy is the most suitable technology for these applications, 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.

References

[1] P. Henry, M. Krainin, E. Herbst, X. Ren and D. Fox, “RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments,” The International Journal of Robotics Research, vol. 31, no. 5, pp. 647-663, 2012.

[2] X. Chen, H. Ma, J. Wan, B. Li and T. Xia, “Multi-View 3D Object Detection Network for Autonomous Driving,” IEEE CVPR, vol. 1, no. 2, p. 3, 2017.

[3] S. Yarboro, P. H. Richter and D. M. Kahler, “The evolution of 3D imaging in orthopaedic trauma care,” Der Unfallchirurg, vol. 120, no. 1, pp. 5-9, 2017.

[4] I. Valverde, G. Gomez, A. Gonzalez, C. A. A. Suarez-Mejias, J. F. Coserria, S. Uribe, T. Gomez-Cla and A. R. Hosseinpour, “Three-dimensional patient-specific cardiac model for surgical planning in Nikaidoh procedure,” Cardiology in the Young, vol. 25, no. 4, pp. 698-704, 2015.

[5] 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.

[6] M. P. Chae, W. M. Rozen, P. G. McMenamin, M. W. Findlay, R. T. Spychal and D. J. Hunter-Smith, “Emerging applications of bedside 3D printing in plastic surgery,” Frontiers in surgery, vol. 2, p. 25, 2015.

[7] 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.

[8] A. Dawood, B. M. Marti, V. Sauret-Jackson and A. Darwood, “3D printing in dentistry,” British dental journal, vol. 219, no. 11, p. 521, 2015.

[9] M. Boffano, P. Pellegrino, N. Ratto, M. Giachino, U. Albertini, A. Aprato, E. Boux, G. Collo, A. Ferro, S. Marone, A. Massè and R. Piana, “Custom-made 3D-printed pelvic prosthesis: is it a safe option for the limb salvage in tumours and catastrophic total hip arthroplasty failures?,” Orthopaedic Proceedings, vol. 100, no. SUPP_5, p. 93, 2018.

[10] 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.

[11] 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.

[12] M. Kedzierski and A. Fryskowska, “Terrestrial and aerial laser scanning data integration using wavelet analysis for the purpose of 3D building modeling,” Sensors, vol. 14, no. 7, pp. 12070-12092, 2014.

[13] M. Forte, “3D archaeology: new perspectives and challenges - the example of Çatalhöyük,” Journal of Eastern Mediterranean Archaeology & Heritage Studies, vol. 2, no. 1, pp. 1-29.

[14] R. G. Hollands, “Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?,” City, vol. 12, no. 3, pp. 303-320, 2008.

[15] C. Harrison and I. A. Donnelley, “A theory of smart cities,” Proceedins of the 55th Annual Meeting of the ISSS-2011, Hull, UK, vol. 55, no. 1, 2011.

[16] C. Ratti and A. Townsend, “The social nexus,” Scientific American, vol. 305, no. 3, pp. 42-49, 2011.

[17] M. Angelidou, “Smart city policies: A spatial approach,” Cities, vol. 41, pp. S3-S11, 2014.

[18] A. Mahizhnan, “Smart cities: the Singapore case,” Cities, vol. 16, no. 1, pp. 13-18, 1999.

[19] A. Cocchia, “Smart and digital city; A systematic literature review,” Smart city, pp. 13-43, 2014.

[20] M. S. Khan, M. Woo, K. Nam and P. K. Chathoth, “Smart City and Smart Tourism: A Case of Dubai,” Sustainability, vol. 9, no. 12, p. 2279, 2017.

[21] M. J. Kaur and P. Maheshwari, “Building smart cities applications using IoT and cloud-based architectures,” in International Conference on Industrial Informatics and Computer Systems, 2016.

[22] A. Datta, “New urban utopias of postcolonial India: Entrepreneurial urbanization’in Dholera smart city, Guajarat,” Dialogues in Human Geography, vol. 5, no. 1, pp. 3-22, 2015.

[23] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich and N. Dagnes, “3D geometry-based automatic landmark localization in presence of facial occlusions,” Multimedia Tools and Applications, pp. 1-29, 2017.

[24] N. Dagnes, E. Vezzetti, F. Marcolin and S. Tornincasa, “Occlusion detection and restoration techniques for 3D face recognition: a literature review,” Machine Vision and Applications, pp. 1-25, 2018.

[25] A. W. Young, E. De Haan and R. Bauer, “Face perception: A very special issue,” Journal of neuropsychology, vol. 2, no. 1, pp. 1-14, 2008.

[26] J. Morton and M. H. Johnson, “CONSPEC and CONLERN: a two-process theory of infant face recognition,” Psychological review, vol. 98, no. 2, pp. 164-181, 1191.

[27] Fantz, “The origin of form perception,” Scientific American, vol. 204, no. 5, pp. 66-73, 1961.

[28] V. Bruce and A. Young, In the eye of the beholder: the science of face perception, Oxford university press, 1998.

[29] J. Meadows, “The anatomical basis of prosopagnosia,” Journal of Neurology, Neurosurgery & Psychiatry, vol. 37, no. 5, pp. 489-501, 1974.

[30] A. R. Damasio, H. Damasio and G. W. Van Hoesen, “Prosopagnosia Anatomic basis and behavioral mechanism,” Neurology, vol. 32, no. 4, p. 331, 1982.
[31] C. Zhang and Z. Zhang, “A survey of recent advances in face detection,” 2010.
[32] S. Roy and S. Podder, “Face detection and its applications,” International Journal of Research in Engineering & Advanced Technology.
[33] A. Colombo, C. Cusano and R. Schettini, “3D face detection using curvature analysis,” Pattern recognition, vol. 39, no. 3, pp. 444-455, 2006.
[34] B. Heisele, T. Serre and T. Poggio, “A component-based framework for face detection and identification,” International Journal of Computer Vision, vol. 74, no. 2, pp. 167-181, 2007.
[35] 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.
[36] J. Ashbourn, Biometrics: Advanced identity verification: The complete guide, Springer, 2014.
[37] A. Ross and A. Jain, “Information fusion in biometrics,” Pattern recognition letters, vol. 24, no. 13, pp. 2115-2125, 2033.
[38] A. K. Jain, L. Hong, S. Pankanti and R. Bolle, “An identity-authentication system using fingerprints,” Proceedings of the IEEE, vol. 85, no. 9, pp. 1365-1388, 1997.
[39] J. Galbally, S. Marcel and J. Fierrez, “Image quality assessment for fake biometric decision: Application to iris, fingerprint, and face recognition,” IEEE transactions on image processing, vol. 23, no. 2, pp. 710-724, 2014.
[40] 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.
[41] J. Rabia and A. H. R., “A survey of face recognition techniques,” Jips, vol. 5, no. 2, pp. 41-68, 2009.
[42] L. A. Cament, F. J. Galdames, K. W. Bowyer and C. A. Perez, “Face recognition under pose variation with local Gabor features enhanced by active shape and statistical models,” Pattern recognition, vol. 48, no. 11, pp. 3371-3384, 2015.
[43] M. S. Hossain and G. Muhammad, “Cloud-assisted speech and face recognition framework for health monitoring,” Mobile Networks and Applications, vol. 20, no. 3, pp. 391-399, 2015.
[44] P. Ekman, “An argument for basic emotions,” Cognition & emotion, vol. 6, no. 3-4, pp. 169-200, 1992.
[45] P. Ekman and W. V. Friesen, Unmasking the face: A guide to recognizing emotions from facial clues, Ishk, 2003.
[46] M. S. L. G. Bartlett, I. Fasel and J. R. Movellan, “Bartlett, M. S., Littlewort, G., Fasel, I., & Movellan, J. R. (2003, June). 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.
[47] D. A. Small and N. M. Verrochi, “The face of need: Facial emotion expression on charity advertisement,” Journal of Marketing Research, vol. 46, no. 6, pp. 777-787, 2009.
[48] 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.
[49] 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.
[50] M. G. Calvo and L. Nummenmaa, “Perceptual and affective mechanisms in facial expression recognition,” Cognition and Emotion, vol. 30, no. 6, pp. 1081-1106, 2016.
[51] V. Bettadapura, “Face expression recognition and analysis: the state of the art,” arXiv preprint arXiv:1203.6722, 2012.
[52] “Stereolabs,” [Online]. Available: https://www.stereolabs.com/zed/.
[53] E. Vezzetti and F. Marcolin, “Geometry-based 3D face morphology analysis: soft-tissue landmark formalisation,” Multimedia Tools and Applications, vol. 68, no. 3, pp. 895-929, 2014.
[54] J. Salvi, J. Pages and J. Batlle, “Pattern codification strategies in structured light systems,” *Pattern recognition,* vol. 37, no. 4, pp. 827-849, 2004
[55] 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.