3D Scanning Technologies for Human Face Expression Recognition: a Comparison Framework

Luca ULRICH, Sandro MOOS, Federica MARCOLIN, Enrico VEZZETTI
Department of Management and Production Engineering, Politecnico di Torino, Turin, Italy

https://doi.org/10.15221/22.71

Abstract

In the last few years, application fields for face analysis have considerably increased, phenomenon that has been fed by a significant improvement of acquisition technologies. In this work, the focus is pointed on face expression recognition, which aims to recognize user's feelings through the analysis of facial data for different applications that moves from security to marketing domain. Data considered in this research are those acquired from 3D cameras that seem to provide more reliable performances rather than 2D devices, especially in "nonstandard" working conditions. This study aims to create a framework able to support a systematic analysis and comparison of 3D acquisition technologies to support the development of real-time applications, with a specific focus on RGB-D.

Keywords: 3d scanners, RGB-D devices, Face Expression Recognition

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: land surveying [10], architecture [11], archaeology [12] for research and tourism purposes and security. Smart cities, urban areas equipped with interconnected sensors able to collect data to be used to manage products and services [13], aim to benefit of the spreading of face recognition technology and deep learning techniques to solve problems such as quickly finding missing children and identifying criminals [14] or monitoring public places such as airports [15]. Geometry of the surfaces acquired with cameras capable of capturing depth information can be used for a more accurate face reconstruction [16], to build 3D aging models [17], face manipulation [18] and landmarking [19]. 3D techniques require a higher computational cost than 2D methods [20], especially if the 3D face model has to be reconstructed from multi-view images [21] or through 3D morphable models obtained from 2D images and 3D scans or even without 3D data [22]. Nonetheless, the robustness given by the opportunity to operate in critical lighting conditions [23], in presence of occlusions [24] [25] and regardless of the orientation of the subject [26] make a 3D approach preferable. Literature about 3D is varied and fragmented due to lack of a shared methodology for analyzing the field and developing new applications in the face of a growing number of 3D devices on the market. Since the presence on the market of a considerable number of devices and technologies has been observed, a systematic comparison of the available solutions has been considered necessary to provide a baseline for the design of automatic procedures able to perform real-time face expression recognition.

2. Methodology

This comparison framework design has been carried out through the integration of a quantitative and qualitative strategy by the usages of a desk research [27] integrated with the suages of a Quality Function Deployment (QFD) [28] model able to correlate two different orthogonal dimensions, namely the qualitative requirements of the face expression recognition usage scenario and the technical specifications of 3D acquisition technologies, retrieved from the results of the desk research.

2.1. Face Expression Recognition

Face expression recognition aims to understand human emotions by observing different areas of the face. Paul Ekman studies have made the way for this discipline, indeed he studied the nervous system activity response to emotions [29]; he defined a set of six basic emotions, which are fear, anger, joy, sadness, disgust and surprise highlighting that his intent was not to deny the variety of affective phenomena, but to attempt to organize those phenomena [30]; he studied the numerous cross-cultural agreements, but also the differences in the judgments of facial expressions [31] and he defined the Facial Action Coding System (FACS) to map facial muscles movements in Action Units (AUs) and trace back the emotions [32]. Bartelett et al. [33] have presented a system to automatically detect faces, to use Gabor representation and then to process data through a bank of SVM classifiers, Shan et al. [34] have used Local Binary Pattern (LBP) to recognize salient micro-patterns on the face and to develop a low-computational method, further developed using Boosted-LBP features and comparing different machine learning algorithms, among them SVM [35], Kotsia et al. [36] have proposed a methodology using geometrical deformation features and SVM, Guo et al. [37] have studied the problematic case of a small number of images to train classifiers, obtaining that the best results in these conditions are achieved by SVM and their own Feature Selection via Linear Programming (FSLP).

The increasing interest in face expression recognition is strongly connected to human-computer interaction (HCI) [38] and involves a variety of fields: Small et al. [39] have studied how much emotional photographs in charity advertisements can evoke sympathy to engender giving, a smart solution applied to marketing, Lee et at. [40] have conducted an experiment involving forty users to study the relationship between emotions and choice of contents on smart TVs, McDuff et al. [41] have presented a toolkit to design user interfaces able to adapt to multiple users' expressions, which could be interesting in interactive applications such as videogames, Calvo et al. [42] have studied the connection between facial expressions and expresser's internal feelings in the context of psychiatry, Olivetti et al. [43] have evaluated the user's engagement in a virtual environment, suggesting that the methodology can be applied to other virtual products, while Nonis et al. [44] have evaluated the engagement of users focused on learning 3D modeling and printing using a 3D simulator. Another relevant application field is robotics, since the capability to automatically understand human's mood significantly improve human robot interaction in terms of communication efficiency and safety [45].

Face expression recognition is a critical task since some expressions, especially fear and sadness according to Alexandre et al. [46], are ambiguous and difficult to be recognized even by a human observer. For this reason, it is necessary to retrieve all the possible kind of information from the face on which the emotion must be detected and, on a visual level, geometrical analysis can be considered the basis of face expression recognition, because the usage of the third dimension is essential to detect and describe facial movements in the most accurate way possible. The main usage scenario requirements could be summarized in the Table 1.

Table 1. Face Expression Recognition usage scenario requirements.

Real-time: individual expressions should be recognized whenever an event associated to what they are assisting is triggered
Wide operating range: individuals' expressions should be recognized both if an individual is getting closer to the camera and moving away.
Accurate at close distance: individuals' expressions should be recognized if an individual is close to the camera.
Accurate at far distance: individuals' expressions should be recognized if an individual is far from the camera.
Able to detect facial features: facial landmarks for face analysis must be recognized.
Integrable into a smartphone: sensors should allow to be put into a smartphone, a tablet, or a laptop to perform face expression recognition.
Portable: this requirement suggests having a sensor small enough to be easily carried by the user.
Robust to light: individuals' expressions should be recognized whatever light conditions are (i.e., in the dark, in a sunny day…).
Head pose invariant: individuals' expressions should be recognized whatever the individual relative orientation with respect to the camera is.
Robust to occlusions: individuals' expressions should be recognized in presence of occlusions (i.e., glasses, scarves…).

2.2. 3D Scanners: RGB-D camera technologies

The interest in the applications mentioned above has received a further impulse since the advent of low-cost 3D cameras, i.e., devices able to detect the third dimension. These sensors are also known RGB-D cameras because they provide two types of data: RGB and D (depth). RGB refers to the color model thanks to which every color can be displayed using three primary color red, green and blue; in other words, it identifies the color images. Depth information is retrievable through depth maps, images on which each pixel has a value representing the distance from the camera. This type of data is an advance compared to 2D data in terms of reliability and is suitable for real-time applications, indeed it is possible to analyze the depth map without building a mesh; every 3D object is identified with x, y coordinates and the depth value instead of set of vertices, edges, and faces. The result is a more responsive acquisition system at the cost of accuracy. The Microsoft Kinect release on the market in 2010 is one of the milestones related to the diffusion of these devices. This sensor has been designed and developed for the specific purpose of recognizing human body actions to perform an original type of human-machine interaction aimed at controlling characters, vehicles, or whatever object movements inside a videogame [47]. 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 shown in table 2

Table 2. 3D Scanner Technologies.

Passive Techniques	Active Techniques
Passive Stereoscopy	Structured Light
	Time of Flight
	Active Stereoscopy

The RGB-D cameras key technical specifications, available both in literature and in datasheets, are shown in table 3.

Table 2. 3D Scanner Technologies technical Specifications.

Resolution: horizontal and vertical number of pixels
Frame rate: number of images captured in one second (FPS, Frames Per Second)
Minimum distance: this parameter establishes the lowest gap for camera functioning
Maximum distance: this parameter establishes the greatest gap for camera functioning
Range: difference between minimum distance and maximum distance
Size: camera dimensions

Twenty-nine sensors belonging to the four categories described above have been analyzed to identify technical specifications strengths and weaknesses of each 3D detection technologies. For each parameters listed as technical specification has been identify a range of working conditions as shown in table 3. Providing a first overview of the analyzed technologies some evaluation could be done. Passive stereoscopy provides great resolution and long operational range; structured light provides good performances for short-range applications since it does not need different points of view to acquire the depth information, making it easier to locate the sensor close to the subject/object of interest.

Table 3. 3D Scanner Technologies technical Specifications Values.

Parameter	**Passive Stereoscopy**		**Structured Light**		**Active Stereoscopy**		**Time of flight**	
	Best	Worst	Best	Worst	Best	Worst	Best	Worst
Resolution	2208x1242	640x480	1280x1024	320x240	1280x800	640x480	1024x1024	176x144
Frame rate [FPS]	60	3	60	30	30	30	30	20
Minimum distance [m]	0,23	0,5	0,2	0,87	0,11	0,4	0,15	0,8
Maximum distance [m]	15	1	5	0,49	10	2,8	60	1
Range [m]	14	0,3	4,4	0,23	10	2,25	59,5	0,85
Size [mm]	57x30x14	230x75x84	80x20x20	279x71x66	90x25x25	126x254x345	65x65x68	250x70x45

By the usage of the QFD methodology sensors' technical specifications and face expression recognition requirements have been fused by involving a focus group composed by experts with different skills and perspective to have sufficient heterogeneous a reliable and representative sample of face expression recognition usage scenario.

3. Results and Discussion

Analyzing the data coming from the QFD model implemented for the Face Expression Recognition usage scenario, shown in table 4, it is possible to obtain a systematic overview of the RGB-D technical specifications impact on face expression recognition. Focusing the attention of the relative importance obtained the resolution parameters seems to play a very fundamental role. This parameter has in fact obtained an importance of 26%. This could be justified by the necessity to discriminate between different features on resulting images. Moreover, data should be retrievable both if the subjects is close or far from the camera, that is the reason why Minimum Distance, Maximum Distance and Range parameters immediately follows Resolution. The other following parameters, that show an importance level lower than 10%, so looks on a boundary importance level, do not represent key element in this specific usage domain, as for instance the dimensions parameter, because of sensors should not be always portable.

Table 4. Face Expression Recognition Usage Scenario QFD Model

Requirements	**Weights**	**Technical Specifications**					
		Resolution	***FPS***	***Range***	***Min Distance***	***Max Distance***	***Dimensions***
Real Time	5	9	9	0	0	0	0
Wide operating range	4	0	0	9	1	9	0
Accurate at close distance	4	9	3	1	9	0	0
Accurate at far distance	4	9	1	3	0	9	0
Able to detect facial features	5	9	0	3	9	9	0
Integrable into smartphone	4	1	0	3	3	3	9
Portable	3	0	0	0	0	0	0
Robust to light	4	3	1	3	3	3	0
Head pose invariant	4	3	1	3	3	3	0
Robust to occlusions	4	3	3	1	1	1	0
Relative Total Score		***26%***	***10%***	***14%***	***16%***	***20%***	***8%***

Once the face expression recognition usage scenario requirements overview has been obtained, it is possible to match these results with the technological performances of the 3D devices families analyzed (see figure 1) to support their classification for the specific application domain.

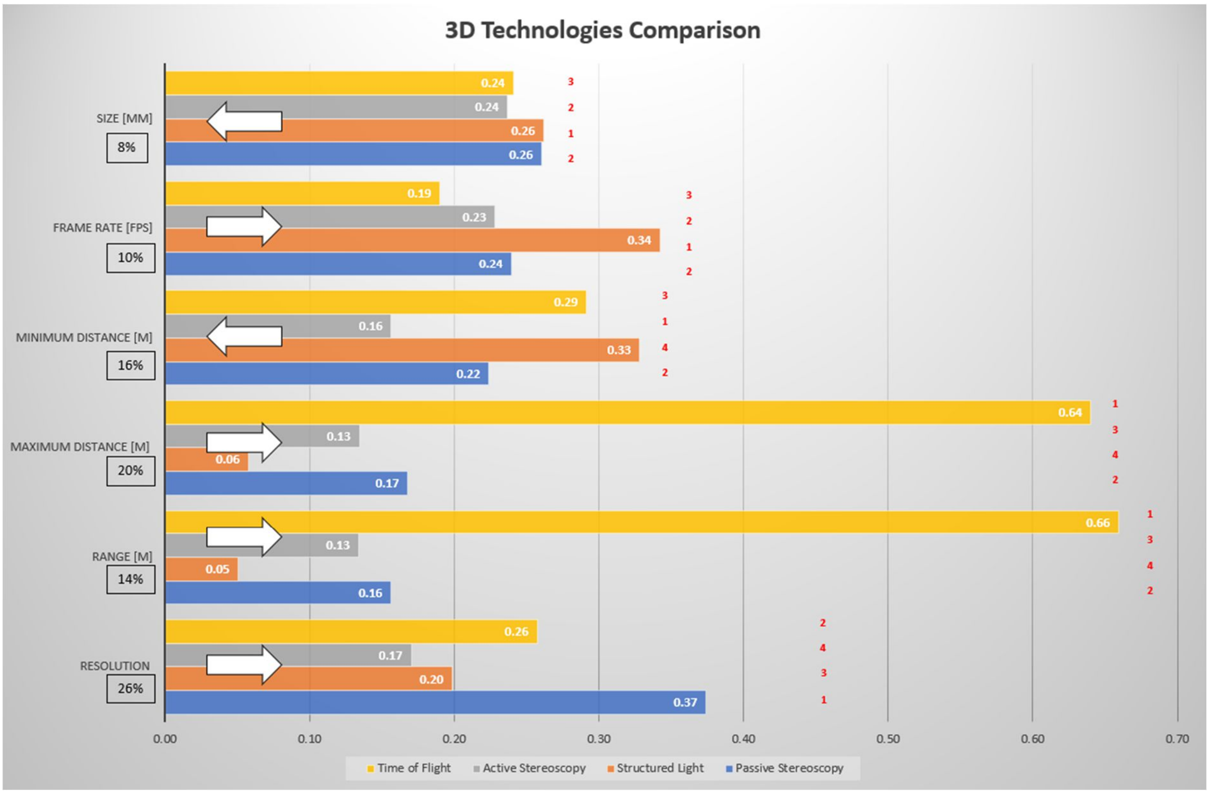

Fig. 1. Normalized 3D technologies comparison (the technical importance got with the QFD is written close to the technical parameter).

Overlapping the results coming from the QFD model with the technical performances of the devices families and focusing the attention on the more critical parameters for the face expression recognition usage scenario time-of-flight seems to be the most suitable solution for face expression recognition usage scenario thanks to its versatility, result confirmed also by the introduction of this kind of technology in most recent smartphone and personal devices. Structured-light appears to be a valid option for short range applications, but it suffers from a rapid performance degradation when the distance of the subject from the camera increases. On the other hand, both stereoscopy technologies benefit from an excellent resolution and operative range, but their strength is the long-range usage. Although in some circumstances the scene reconstruction can be satisfactory at short-distances, even less than 20 cm, occlusions in presence of a complex geometry, such as the human face, make the depth acquisition not optimal and invalid for face expression recognition. Furthermore, the presence of two cameras is difficult to reconcile with the need of small size required by integrating such a type of technology within personal devices such as tablet and, mostly, smartphones.

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] 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.

[11] 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.

[12] 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, 2014.

[13] V. Albino, U. Berardi and R. M. Dangelico, "Smart cities: Definitions, dimensions, performance, and initiatives," Journal of urban technology, vol. 22, no. 1, pp. 3-21, 2015.

[14] H. Wu and H. L. P. Xu, "Design and Implementation of Cloud Service System Based on Face Recognition," in Conference on Complex, Intelligent, and Software Intensive Systems, 2020.

[15] 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.

[16] M. Zollhofer, J. Thies, P. Garrido, D. Bradley, T. Beeler, P. Perez, M. Stamminger, M. Niessner and C. Theobalt, "State of the art on monocular 3D face reconstruction, tracking, and applications," Computer Graphics Forum, vol. 37, no. 2, pp. 523-550, 2018.

[17] S. Riaz, U. Park, J. Choi and P. Natarajan, "Age progression by gender-specific 3D aging model," Machine Vision and Applications, vol. 30, no. 1, pp. 91-109, 2019.

[18] Z. Geng, C. Cao and S. Tulyakov, "3d guided fine-grained face manipulation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[19] J. Deng, A. Roussos, G. Chrysos, E. Ververas, I. Kotsia, J. Shen and S. Zafeiriou, "The menpo benchmark for multi-pose 2D and 3D facial landmark localisation and tracking," International Journal of Computer Vision, vol. 127, no. 6-7, pp. 599-624, 2019.

[20] A. Abate, M. Nappi, D. Riccio and G. Sabatino, "2D and 3D Face Recognition: A Survey," Pattern Recognition Letters, vol. 28, pp. 1885-1906, 2007.

[21] J. Cao, Y. Hu, B. Yu, R. He and Z. Sun, "3D aided duet GANs for multi-view face image synthesis," IEEE Transactions on Information Forensics and Security, vol. 14, no. 8, pp. 2028-2042, 2019.

[22] L. Tran and X. Liu, "On learning 3d face morphable model from in-the-wild images," in IEEE transactions on pattern analysis and machine intelligence, 2019.

[23] S. Zhou and S. Xiao, " 3D face recognition: a survey," Human-centric Computing and Information Sciences, vol. 8, no. 1, p. 35, 2018.

[24] 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.

[25] 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.

[26] E. Vezzetti, S. Moos, F. Marcolin and V. Stola, "A pose-independent method for 3D face landmark formalization," Computer Methods and Programs in Biomedicine, vol. 108, no. 3, pp. 1078-1096, 2012.

[27] P. Verschuren, H. Doorewaard and M. J. Mellion, Designing a research project, The Hague: Eleven International publishing house, 2010.

[28] S. Mizuno and Y. Akao, Development history of quality function deployment, vol. 90, Quality Resources, 1994, p. 339.P. Ekman, R. W. Levenson and W. V. Friesen, "Autonomic nervous system activity distinguishes among emotions," Science, vol. 221, no. 4416, pp. 1208-1210, 1983.

[29] P. Ekman, "An argument for basic emotions," Cognition & emotion, vol. 6, no. 3-4, pp. 169-200, 1992.

[30] P. Ekman, W. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. LeCompte, T. Pitcairn, P. Ricci-Bitti and K. Scherer, "Universals and cultural differences in the judgments of facial expressions of emotion," Journal of personality and social psychology, vol. 53, no. 4, p. 712, 1987.

[31] P. Ekman and W. V. Friesen, Unmasking the face: A guide to recognizing emotions from facial clues, Ishk, 2003.

[32] M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, "Real time face detection and facial expression recognition: development and applications to human computer interaction," in 2003 Conference on computer vision and pattern recognition workshop, 2003.

[33] C. Shan, S. Gong and P. W. McOwan, "Robust facial expression recognition using local binary patterns," in IEEE International Conference on Image Processing 2005, 2005.

[34] C. Shan, S. Gong and McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," Image and vision Computing, vol. 27, no. 6, pp. 803-816, 2009.

[35] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," IEEE transactions on image processing, vol. 16, no. 1, pp. 172-187, 2006.

[36] G. Guo and C. R. Dyer, "Learning from examples in the small sample case: face expression recognition," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 35, no. 3, pp. 477-488, 2005.

[37] M. Matsugu, K. Mori, Y. Mitari and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network," Neural Networks, vol. 16, no. 5-6, pp. 555-559, 2003.

[38] B. Jingxin, L. Yinan and Z. Shuo, "3D Multi-poses Face Expression Recognition Based on Action Units," in Proceedings of the 2019 International Conference on Information Technology and Computer Communications, 2019.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] E. C. Olivetti, M. G. Violante, E. Vezzetti, F. Marcolin and B. Eynard, "Engagement Evaluation in a Virtual Learning Environment via Facial Expression Recognition and Self-Reports: A Preliminary Approach," Applied Sciences, vol. 10, no. 1, p. 314, 2020.

[45] F. Nonis, E. C. Olivetti, F. Marcolin, M. G. Violante, E. Vezzetti and S. Moos, "Questionnaires or Inner Feelings: Who Measures the Engagement Better?," Applied Sciences, vol. 10, no. 2, p. 609, 2020.

[46] Y. Chen, R. Hu, J. Xiao and Z. Wang, "Multisource surveillance video coding by exploiting 3d and 2d knolwedge," in 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019.