The Impacts of Depth Camera and Multi-Sensors for Face and Fingerprint Recognition in 2D and 3D - A Case Study

Mohamed Abdul-Al*[1], George Kumi Kyeremeh[2], Rami Qahwaji[3] and Raed Abd-Alhameed[4]
{m.abdul.al@bradford.ac.uk*[1], g.k.kyeremeh@bradford.ac.uk[2], r.s.r.qahwaji@bradford.ac.uk[3], r.a.a.abd@bradford.ac.uk[4]}

Faculty of Engineering and Informatics, University of Bradford[1], Bradford, BD7 1DP, UK[1,2,3,4]

Abstract. This paper reviews the multi-modern sensors for multimodal (Face ad Fingerprint) of a biometric system by using a depth camera. The use of face authentication in biometric data allows this innovation to expand and be used in a variety of fields. Recently, attendance monitoring systems depending on biometric identification for higher education are underutilization, presenting an excellent chance to do fascinating experiments. The installation of biometric attendance systems necessitates the use of both hardware and software components. The first deep CNN (Convolutional Neural Network) method for light source-oriented Face Recognition (FR) takes advantage of the more detailed data given in a lens let display technology picture and that has been used in different borders such as airports, seaports, and land ports. In addition, the use of 3D camera technology for measuring medical outcomes in the healthcare marketplace is growing. The Intel® RealSense TM is one of the top 3D thermal imaging cameras systems on the market today, and it is well-suited for usage in a variety of areas such as medical systems, automation, and medical. Advances in in-depth sensor cameras technology have led to a considerable rise in the incorporation of these innovations into moveable systems, implying the great future potential for widespread in clinic and sector medical screening systems. Furthermore, initially, the use of dispersion maps in conjunction with depth maps and Two-Dimensional Red Green Blue (2D-RGB) pictures has been examined in terms of a fusion strategy to increase FR performance. The suggested approach employs the 2D-RGB crucial radiations emitted viewpoint, and depth maps and dispersion derived from the whole collection of radiations emitted pictures connected with a lens let light source. Following that, feature separation was carried out with the use of a VGG- deep sighs description for texturing and individually well their representations for depth maps and dispersion. The collected characteristics are then combined and supplied into a classification model.

Keywords: Face Recognition (FR), Fingerprint, Sensor, Intel RealSense Camera, 2D-RGB, 3D-RGB, Feature Extraction, Depth maps, VGG-Face, and Facial Tracking.

1 Introduction

One most popular accepted, collectible, and ubiquitous biometric modalities are Face Detection or Face Recognition (FD/FR), attracting considerable interest from both the research

IMDC-IST 2021, September 07-09, Sakarya, Turkey

DOI 10.4108/eai.7-9-2021.2314977

world and the industry [1] [2]. Significant gains in FR systems have been achieved in recent decades due to the advent of deep learning technologies and the accessibility before training of deep learning systems [3]. Unsurprisingly, deep CNN (Convolutional Neural Network) approaches lead to the presence of modern technology in FR/FD. Face photos are known to pose hurdles for precise recognition including for individuals, particularly in unsupervised scenarios with substantial variability in age, mood, stance, light, occlusion, and so on [4], and hence even deep learning systems might fail when faced with this obstacles. Concurrent with advancements in feature representation and classification for face identification, great strides have been made in the creation of richer image sensors, such as lens let light field cameras, stereo, depth, NIR (Near Infrared), and thermal. Industrial lens let light field cameras, including Lytro [5], have lately brought light intensity imaging technology to the forefront. These camera sensors not just the brightness of light on a single 2D (Two-Dimensional) level point, as well as the brightness of visible light originating from other directions [6] [7]. Furthermore, lens let light field cameras are gaining popularity in the forensic and biometric fields for facial identification and verification [8] [9] [10] [11] [12] [13].

In the current marketplace, there are a rising number of 3D (Three-Dimensional) camera sensors technologies, resulting in an extensive region of technical uses in a variety of industries. Identifying a one-of-a-kind solution in healthcare research, 3D gaming, scanning, motion capture, wellness applications, and clinical is thus becoming increasingly challenging. One among the most popular notable 3D depth-sensing camera is the Intel RealSense, particularly with the recent revelation that production of the Microsoft Kinect Sensor V2/2.0 has already ended, likely signalling the end of Microsoft's engagement in this equipment field [14]. While Kinect assistance for the dynamic responses will keep going in the meantime, the restricted availability of this innovation for purchasing decisions will most likely herald a shift in the acceptance and usage of this specific objective device for innovation projects.

The depth sensor generates pictures in the same way as a standard camera does. Rather than colour, the brightness of every image pixel shows the range to the associated point. This is where these devices truly shine in areas such as computer vision technology. Therefore, depth sensors are often referred to as 3D cameras at times. PrimeSense [15] owns the patent for the depth detecting system used in the Kinect sensor system. Depth sensors are used in a multitude of scenarios other than gaming, such as reconstructing the forms of observed things [16] [17], indication language interpretation systems [18] [19], touching systems [20] [21], rehabilitation and ergonomics [16] [17] and interacting with Google Earth [16] [22]. According to M. Andersen et al. [23] investigated many features of the Kinect sensor that are significant in the development of real-world FR systems.

2 Biometric Sensor

To capture an individual's biometric characteristics, biometric sensors are used. The sensors serve two functions in the biometric presence monitoring system. Detectors are employed first to collect biometric information that will be used in models. Following that, each framework is labelled with a named person or roll number. These models, including the participants' data, are then saved in a server as a guide for future contrast by newly gathered biometric information. This stage is just completed once at the beginning of each school semester. Whenever a lecture is held, the sensors' next goal is to collect a fresh set of biometric information from every learner. The identity of each learner must always be memorized to identify the class's involvement. The newly gathered biometric information can be compared to the models to track the proper time, name, and date of the pupils participating in the lecture. The next subtypes will cover the different sensors, as well as the circumstance or situation required to record failure biometric information, also to examining the role and relevance of detectors in the biometric-based presence scheme.

2.1 Fingerprint (FP) Sensor

The fingerprint image is captured using various types of sensors. Essentially, there are different categories of FP image sensors [24]:

a) State of matter
b) Visual
c) Ultrasound

Optical sensors may be used in a range of methods to acquire fingerprint images. The first and most extensively utilized biosensors are dependent on the flawed proposed design of total interior contemplation (FTIR). An optical system dependent on FTIR includes CMOS (Complementary Metal-Oxide-Semiconductor) camera, a CCD (Charge-Coupled Device) or a light source, a glass/plastic prism, and a lens. As the viewer contacted the prism's highest place, the CCD or mirrored illumination from the lenses was caught by the CMOS camera via the concentrating lens. An FP classifies diverse patterns of the valley and ridge characteristics [25]. There are numerous benefits when utilizing an optic detector. Because the device only identifies the finger with a 3D (Three-Dimensional) form, replicated FP magic cannot readily fool it. Furthermore, it can capture multiple pictures resolutions and generate top-notch photos. Although these benefits, the sensor module dependent on FTIR is vulnerable to a wet or dry and warm or cold finger, resulting in a bad impression or a saturated [26]. Furthermore, it cannot be built and tested because of the lack of flexibility of lowering the optical length, which might produce optical error in a picture.

Semiconductor device detectors often described as silicon detectors, are intended to utilise small detectors that constitute pixels of the array. Semiconductor device detectors are classified into 4 categories: i) piezoelectric, ii) capacitive, iii) thermal and iv) electric field [24]. Perhaps the most prevalent is the capacitive kind. A capacitive detector is made up of 2 panels. The first panel is made up of a 2D (Two-Dimensional) array of capacitors positioned under the finger photoactive layer, while another panel is the epithelium of the finger. Semiconductor devises detectors can alleviate issues involved with sensing devices, including a smaller footprint and tuneable electrical parameters, when resolving a wet or dry and warm or cold FP picture. Moreover, because a 3D finger layer is detected based on distance calculation, these detectors cannot be readily replicated by a copy or fraudulent FP picture. While their advantages, Semiconductor devices detectors are susceptible to ESD (Electro-Static Discharges). Furthermore, after several applications of the detector [27], white blobs are apparent in the FP image. Additionally, clean-up is essential to create a nice and clear FP image.

Ultrasound devices are utilized to distinguish between valleys and the sound reflection depths of hills [26]. The detector is made up of a transmitter that creates an audio signal and a receiver that receives the associated revealed or resonance signal from the FP layer. Ultrasound detectors can seizure a high-quality picture deprived of existence influenced by oil, soil, or new impurities on the finger [24]. These detectors, on the other hand, are large, costly, and need an extra period to acquire FP images [25].

When selecting an FP reader, consider the collection area, resulting in picture quality, and structural accuracy [24]. The optimal FP detection region is 1 x 1 inch2 (25 4 x 25 4 mm^2). Nevertheless, the bulk of industrial readers are shorter to decrease cost and size. Furthermore, the lowest photo quality is roughly 500 dpi (pixels per inch). The structural deformation caused by the FP reader is used to determine the correctness of the geometry. Other characteristics to choose include the supporting OS (Operating System) version, I/O connection, FPS (Frames Per Second), automated finger identification, and encryption.

2.2 Sensor Face

FR capabilities provided or detectors are cameras utilized to record video frames or capture photos. Face information obtained can be 2D intensity picture shape, 3D brightness, and depth data displays, or IR (Infrared) [1]. A camera's crucial elements are lenses and image sensors [28]. Typically, the camera module converts light into electrical ions via the lens of the camera and then to a digital signal. The electrical signal in an input photo is equal to the intensity of the bright, with brighter pictures having more charges than dull ones. CCD and CMOS image sensors are classified as one of two categories. The cumulative energy is equal to the illumination that reaches every pixel, which is the basis for these detectors. For voltage conversion in a CCD detector, the voltage by 1 pixel is transmitted successively to the next pixel till a constant voltage receives. Furthermore, the CCD sensor has an analogy signal. Every pixel in a CMOS sensor converts ions to energy immediately [29].

The photodetector is connected with several attributes including pixel sensitivity, frame rate, and spatial resolution [1]. The film was mainly composed of photos taken in sequential order, commonly defined as frames. As a result, the edge level is described as the number of pictures taken every moment [30]. The sharpness of a picture is similarly proportional to its quality. A video picture's resolution is clear as the whole quantity of pixels in the picture, or by the number of lines and pixels in the vertical and horizontal aspects [31]. Furthermore, pixel sensitivity denotes the sensor's visual disturbances. A back-illuminated detector was created to improve depth perception by utilizing the contralateral silicon layer for a more efficient visible region [28].

3 Face Identification with Modern Image Sensors

New image detectors, including stereo cameras, depth, NIR, and thermal, have been typically adopted for diverse facial biometric purposes [32] [33] [34]. The richer descriptions gained from the environment, as well as their resilience versus specific fluctuations in facial features, inspired the decision to utilise current image detectors. Depth pictures, for instance, could be more resistant to illumination changes [35], but several views of pictures are less susceptible to alterations [36]. Fig 1 summarises a variety of current, comprehensive, and acceptable FR systems dependent on developing image detectors, organised by release date. Fig 1 contains data on the photodetector utilised, the kind of technique for extracting characteristics, a categorizer, and the level of fusion if needed. FR solutions that integrate, it is not uncommon to have many characteristics derived from different image sensors or extracting characteristics algorithms [37]. Learning algorithms control the modern technology in FR based on novel detectors, as shown in Fig 1. In the case of fusion FR solutions, fusion is generally conducted at the different scales since it contains more rich FR data.

Year	Sensor for Imaging	A Categorizer	Technique for Extracting Characteristics	Level of Fusion	References
2014	Kinect and 2D-RGB	NNC	DLBP	Characteristic	(Aissaoui et al. 2014)
2014	Kinect, 2D-RGB and Thermal	WNNC	HAAR, HOG and LBP	Characteristic	(Nikisins et al. 2014)
2015	3D Stereo	SURF	SF, HGORP and AHE	Choice	(Nigam et al. 2015)
2016	Kinect and 2D-RGB	Soft-max, SVM and WNCC	CNN, HOG, HOGOM, LBP and HAAR	Choice and Characteristic	(Simón et al. 2016)
2016	Kinect and 2D-RGB	SVM	CNN	Characteristic	(Lee et al. 2016)
2016	NIR, Kinect and 2D-RGB	Logistic Regression	3D-LBP, LBP, VGG-Face Descriptor, FHOG and PHOW	Characteristic	(Freitas et al. 2016)
2016	Kinect and 2D-RGB	Soft-max	NN	N/A	(Chowdhury et al. 2016)
2016	Thermodynamic	SRC	DSF, WLD, LBP and Gabor jet	Characteristic	(Bi et al. 2016)
2016	NIR and 2D-RGB	DTL	O-CNN	Characteristic	(Liu et al. 2016)
2016	NIR	Soft-max	CNN	N/A	(Reale et al. 2016)
2017	NIR	NNC	Personality in Patches	Characteristic	(Joardar et al. 2017)
2017	Thermodynamic	NNC	SURF, WLD, Gabor, LBP and SIFT	N/A	(Hermosilla et al. 2017)
2017	NIR and 2D-RGB	NNC	CNN	Characteristic	(He et al. 2017)
2017	NIR and 2D-RGB	Soft-max	CNN	Characteristic	(Lezama et al. 2017)
2018	RGB, Depth, Skeleton, Hybrid	RGB-D	CNN, RNN	N/A	(Wang et al. 2018)
2019	RGB, Depth	N/A	Action Representation, Ineraction Recognition	N/A	(Zhang et al. 2019)
2019	RGB, Depth	SVM	CNN, RNN	Characteristic	(Liu et al. 2019)
2019	Mobile, Wearable Sensors, Video	SVM	KNN, ANN, LDC, MKL, GK, NB, DT, HMM	Characteristic	(Nweke et al. 2019)
2020	Video	SVM, SURF	Tb, GM, DM, KNN, RNN, CNN	Characteristic	(Jegham et al. 2020)

Fig 1. A look at some of the developing – anti-field detectors that are being used for FR. The following abbreviations have been utilized in this Figure: AHE stands for (Adaptive Histogram Equalization). DLBP stands for (Depth Local Binary Patterns); DNM stands for (Discriminant Normal Maps); DSF stands for (Down Sampling Feature); FHOG stands for (Felzenszwalb's HOG); HAOG stands for (Histogram of Averaged Oriented Gradients); HOGOM stands for (Histograms of Gabor Ordinal Measures). HPOG stands for (Histograms of Principal Oriented Gradients); HGORP stands for (Horizontal Gradient Ordinal Relationship Pattern); LDA stands for (Linear Discriminant Analysis); LDP stands for (Local Derivative Pattern); O-CNN stands for (Ordinal Convolutional Neural Network); PCA stands for (Principal Component Analysis); PSIFT stands for (Pyramid Scale Invariant Feature Transform). RS-LDA stands for (Random Subspace Linear Discriminant Analysis); SRC is for (Sparse Representation Classifier); SURF is for (Speeded Up Robust Features), and SF stands for (Steerable Filter); WNNC stands for (Weighted Nearest Neighbour Classifier); N/A stands for (Not Applicable).

The suggested method leverages the additional data present in the display technology picture by combining characteristics generated from the 2D-RGB sector directly aspect, and even the depth maps and contrast, and is projected to outperform 2D-RGB+depth and 2D-RGB FR methods. The recommended FR solution necessitates the necessary stages:

a) **Before-processing:** The programme Light Area Toolkit v0.4 [38] has been utilized to generate the sub-hinge several views vector. The facial section is then cut throughout all sub-hinge photos dependent on the features in the dataset, and the cut sub-hinge pictures are shrunk to (24 × 24) pixels**.**

b) **Disproportion map removal:** A dispersion map is recovered from the trimmed sub-hinge array of several views, collecting the geometric data accessible in the display technology picture. The display technology dispersion map was recovered utilizing the approach provided in [39] and [40], which generates the discrepancy map as slopes of unipolar plane images.

c) **Extraction of depth maps:** A feature vector is recovered from the clipped several views of the sub-hinge array, giving feature vectors on the placement and form of the face elements. The feature vector was retrieved utilising the approach provided in [41], which predicts several views audio correlations and then improves them utilizing the analysis to identify.

d) **VGG-Face characteristics removal:** Before training the VGG deep markers, which were originally formed for 2D FR, are well-adjusted individually for depth map and mismatch. The depth characteristics, texture, and disparity are extracted from the 2D-RGB core image, and also the depth maps and disparity, utilizing three deep learning solutions depending on the VGG-Face classifier [42].
e) **Level of Fusion Characteristics:** For every input, the collected characteristics are concatenated into a single category using Fusion of Different Scales.
f) **SVM classification:** The synthesized extracted features are given to the SVM classifier (developed with [43]), which returns the participants' privacy. In this research, the accuracy of a soft-max classifier was also tested, with SVM operating slightly better than soft-max, explaining SVM's selection as the last classifier.

4 Extraction of VGG-Face Features

VGG-Very-Deep-16 has been one of the best CNN designs for numerous image processing applications [44]. Around 2.6 million face photos were trained to create such a before training the VGG-Face model for FR, which includes rich changes in lighting, voice, posture, and occlusion. Since the VGG-deep model is the first training with 2D pictures, it cannot be adequate for FR to specify dispersion and depth data. This research has before trained the VGG-Face model using the VGG-Very-Deep-16 CNN levels, taking depth maps and dispersion into account at the entry and away from the gradient descent impacts. Despite some researcher gathers and space limits, the fine-tuning for both depths maps and dispersion was conducted utilising a total of 30 epochs, LR (Learning Rate = 0.005) and a BS (Batch Size) of 32. The VGG-deep character produced by executing the VGG-Very-Deep-Face CNN deprived of the final 2 attached layers, as established in [42], is utilised to remove characteristics from the 2D-RGB centre representation and also depth maps and dispersion. The outcome is a VGG-Face specification with an entire number of 4,096 characteristics to every entry, described as completely integrated level 6 characteristics. Before training the VGG-Face model was utilised for 2D-RGB pictures, while fine-tuned models were employed for depth maps and dispersion.

5 Tracking of the Face

The Intel RealSense camera series could be utilized to detect and observe face emotions and actions either with or without hairstyle and spectacles. It can identify approximately 78 face features points in 3D, which could aid in the creation of facial animation, avatars, and emotion identification [45]. Intel RealSense could also identify head position across 3D coordinates for roll, yaw, and pitch. Based on the camera used, tracking up to 4 people with indicated squares for face boundaries is possible. This is especially important in platforms when the environment cannot be handled since it allows the person to be recognized and observed solely. In comparison to several of the recent industry frontrunners, such as the Microsoft Kinect 2.0.0, the Intel RealSense camera series will offer high-quality and sampling rates. The higher the pixel density and sample rate, while monitoring fine or fast movement, the better.

5.1 Intel RealSense Camera F200, R200 and SR300

The Intel RealSense SR300 is a simple, transmitted light 3D camera sensor that has one of the lowest 3D depths and 2D digital cameras presently commercially available. The Intel RealSense SR300 provides customers the option to engage with active sense by integrating depth detection with a 1080p RGB sensor. Hand movement recognition, background segmentation, 3D scanning, and facial recognition are all examples of techniques that can be used. The Intel RealSense SR300 camera is appropriate for reality and wearable technology applications like hand and finger tracking, face tracking and detection, checking and localization, and scene segmentation. This is accomplished by the simultaneous employment of an IR projector and a Light detector with coded emission spectra [46].

A globe USB 3.0 thermal imager that can provide video feeds in IR is the Intel RealSense Camera R200, colour, and depth and may be utilized with a variety of compatible devices including Ultrabook's TM, 2-in-1, All-in-One, or portable systems (Intel 2016). The R200 has an IR depth detecting capability and a full HD colour camera. Its 3 cameras, which comprise RGB and stereoscopic IR, are used to produce depth. The VGG-Face description produced by executing the VGG-Deep-Face CNN omitting the previous two convolution levels, as proved to be successful in [47], is utilized to detect edges from the 2D-RGB centre perspective and also depth maps and dispersion. The outcome is a VGG-Face specification with a maximum number of 4096 characteristics for every entry, described as completely integrated level 6 characteristics. Utilizing ultrasounds, the developer may generate 3D printable things and sections, such as the capacity to create and edit bespoke avatars, by computationally recording individuals or stuff in a 3D space. The formation of 3D avatars or things yields things that might be integrated into real-life fields and applications to aid with visualization [45]. Avatars have also been utilized in rehabilitation programmers where people do not wish to view themselves afterward distress permanently or temporarily bodily damaged.

The Intel RealSense spectrum has continuously advanced in current decades, notably in terms of performance capture accuracy, camera capacity, and area of vision lengths. Fig 2 depicts a comparative evaluation of the F200, R200, and SR300 systems' key characteristics.

The Intel RealSense SDK, SDK elements, and deepness camera operators for the F200, R200, and SR300 models are no further getting upgraded leading to the release of the latest software of the visual range [46], as indicated below. Fig 3 provides an in-depth evaluation of the SDK application's running regions and applications.

The Intel RealSense SDK has 3D human tracking include certain critical qualities that aid in the development of strong and productive apps. This means the capability to offer accurate results in the influence of environmental aberrations like piercings or spectacles over a wide variety of colours. Facial hair could be another factor that can significantly affect object tracking; nevertheless, the Intel RealSense SDK will mitigate this effect. Obstacles could also critically degrade the object pursuing of many camera detection systems; common actions like yawning, wiping the face, and so on can affect detection and information collecting. This is something that the Intel RealSense camera spectrum can consider and decide to pursue properly.

Intel RealSense F200, R200, and SR300 Specifications Comparison

	Intel RealSense SR300	Intel RealSense R200	Intel RealSense F200
Camera with RGB Filter in Pixel	1080p at 30 FPS, 720p at 60 FPS	1080p at 30 FPS	1080p at 30 FPS
Camera with Depth in Pixel	Up to 640 x 480 at 60 FPS (Fast VGA, VGA); HVGA at 110 FPS	640 x 480 resolution at 60 FPS	Up to 640 x 480 at 60 FPS (Fast VGA, VGA); HVGA at 110 FPS
Viewing Angle in RGB Colour	41.5°; 68°; 75.2°	77° x 43° x 70°	43°; 70°; 77°
Area of vision in IR	55° x 71.5° x 88°	70° x 46° x 59°	59°; 73°; 90°
Availability of the SDK	SDK 2.0 Capable and Support Active (GitHub)	Discontinued	Discontinued

3D Camera Specifications

	Intel RealSense SR300	Intel RealSense R200	Intel RealSense F200
Effective Range	0.2 m – 1.2 m	0.4 m – 2.8 m	0.2 m – 1.2 m
Texture Mapping	✓	✓	✓
World Mapping	✓	✓	✓
Person Tracking	X	✓	✓

Fig 2. Intel RealSense F200, R200, and SR300 Characteristically Analysis [46].

The constancy of the monitoring can be harmed by changing light sources, and previously, it has been observed that the detection of feature vectors has been substantially harmed; this was especially apparent in the development of Microsoft Kinect [48]. Vision-based detection systems frequently function best in controlled situations with few changes including shape, skin tone and clothing colour, illumination level, and position level. In addition, peripheral clutter is kept to a minimum. Utilizing this technology in the house for health assessment could be challenging for distant individuals because the temperature will not be appropriate [48]. On the other hand, the Intel RealSense works well during poor-light circumstances, when the other technologies strain and are also troublesome due to large variations in illumination.

Intel RealSense F200, R200, and SR300 Camera Algorithms Operating Systems			
Face Recognition	**Intel RealSense SR300**	**Intel RealSense R200**	**Intel RealSense F200**
Pulse	0.3 m – 0.6 m	0.3 m – 0.7 m	0.3 m – 0.6 m
Pose	0.3 m – 1 m	0.5 m – 1.5 m	0.3 m – 1 m
Landmark	0.3 m – 1 m	0.5 m – 1.5 m	0.3 m – 1 m
Detection	0.3 m – 1 m	0.55 m – 2.5 m	0.3 m – 1 m
Recognition	0.3 m – 1.5 m	0.3 m – 1.5 m	0.3 m – 0.8 m
Expression	0.3 m – 1 m	0.3 m – 1 m	0.3 m – 1 m
Tracking Functions in General			
Person Detection	0.5 m – 2.5 m	0.7 m – 3.5 m	X
Person Tracking	0.5 m – 5 m	0.7 m – 5 m	X
Object Tracking	0.3 m – 1.8 m	X	0.3 m – 1.8 m

Fig 3. Intel RealSense F200, R200, and SR300 Systems Required to operate Environment [46].

5.2 Intel RealSense Camera ZR300, D415 & D435

As the marketplace for the manufacturing of Intel Real-Sense and SDK has grown, so has the requirement for both long - short-term thermal imaging cameras, leading to the creation of the D435, ZR300, and D415 Intel Real-Sense Camera systems.

The ZR300 has a high-quality depth camera with an incorporated powerful motion detection mechanism. The ZR300 offers a wide range of uses because of its high-quality deepness detecting, extremely durable monitoring, long-term assistance, outdoor and indoor usage, and low energy consumption. The camera has an inside variety of around 3.5 m and an outside length capable of capturing far thousands of miles. The recommended option, on the other hand, changes based on the lighting and situation [45]. The Intel RealSense Peripheral Device ZR300 is viewed as an appropriate option for educational videos, mechanical development, and systems engineering, among several other proposed options. The ZR300 can provide 3D spatial awareness by utilizing object tracking, spatial awareness, and region training. Most significantly, the ZR300 provides a full edge level, that allows 3D perusing systems and prompts mechanical actions to measure activity in a variety of environments [45].

Intel just launched a new D400 system set with the most sophisticated information fusion performance to time. The D415 and D435 deepness identifying cameras would cover the newest Intel RealSense vision system and component. Besides that, to the dynamic picture flash and normal angle of vision, the Intel RealSense Deepness Camera D415 provides a critical option with straightforward deepness channel information collection. The D435, on the other hand, can record and transmit depth data from object detection, as well as a global picture shutter and a broader range of vision, resulting in great depth perception accuracy in movement [46]. The optimum situation for the D400 sequence has been significantly enhanced over earlier incarnations of the Intel RealSense Camera line, with information gathering achievable at intervals close to and beyond 600 seconds in both inside and outside scenarios [46].

6 Conclusion

It is still difficult to select appropriate kinds of computers for designing a biometric attendance monitoring system. The two often used biometrics qualities of the present scheme are the face and fingerprint because of their simplicity and high acceptance ratio. For FP or facial picture capture, an optical system is necessary. CMOS is the most common sensor in virtually most cameras. FR attendance systems save time by eliminating the requirement for individuals to wait and make eye contact with the detector. To communicate attendance records, the communication network is critical. Wi-Fi is an excellent choice since it can carry large volumes of information in a little period of interval. This is beneficial for tracking presence in real-time. Moreover, with the growing usage of smartphones in the IoT (Internet of Things) age, biometric attendance systems are being pushed toward the next level. Currently, new smartphones have built-in hardware and software including a fingerprint sensor, facial recognition, camera, and iris scanner.

The Intel RealSense family is still in its development, but it is rapidly developing thanks to an ever-improving camera and an ever-growing SDK. Because of its 3D scanning, FR, and FP identification features, the Intel RealSense system is regarded as a competitive, if not improved, a competitor to the Microsoft Kinect, and thus seen as a viable tool for research & innovation in the healthcare industry.

As a result, for future study, there is an option to collect attendance simply utilizing cell phones rather than putting up a program with different system components. Utilizing a wireless router on the smartphone, the biometric features may be captured and transferred to the cloud server for verification. Future studies will examine a reorganized photo as additional light field input data to raise the efficiency of an unfocused range of values for different factors located at varying spaces; it may be highly valuable for several faces' identification challenges.

Acknowledgments

This project has received funding from the European Union's Horizon-MSCA-RISE-2019-2023, Marie Skłodowska-Curie, Research, and Innovation Staff Exchange (RISE), titled: Secure and Wireless Multimodal Biometric Scanning Device for Passenger Verification Targeting Land and Sea Border Control.

References

[1] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern recognition letters*, vol. 79, pp. 80-105, 2016.

[2] M. Günther, L. El Shafey, and S. Marcel, "2D face recognition: An experimental and reproducible research survey," Idiap, 2017.

[3] G. Hu *et al.*, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," 2015, pp. 142-150.

[4] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2144-2157, 2014.
[5] M. Rerabek, L. Yuan, L. A. Authier, and T. Ebrahimi, "[iso/iec jtc 1/sc 29/wg1 contribution] epfl light-field image dataset," ISO/IEC JTC 1/SC 29/WG1, 2015.
[6] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," 2005.
[7] M. Levoy and P. Hanrahan, "Light field rendering," 1996, pp. 31-42.
[8] R. Raghavendra, K. B. Raja, and C. Busch, "Exploring the usefulness of light field cameras for biometrics: An empirical study on face and iris recognition," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 922-936, 2015.
[9] R. Raghavendra, B. Yang, K. B. Raja, and C. Busch, "A new perspective—Face recognition with light-field camera," in *2013 International Conference on Biometrics (ICB)*, 2013: IEEE, pp. 1-8.
[10] R. Raghavendra, K. B. Raja, B. Yang, and C. Busch, "Multi-face recognition at a distance using light-field camera," in *2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2013: IEEE, pp. 346-349.
[11] R. Raghavendra, K. B. Raja, B. Yang, and C. Busch, "Comparative evaluation of super-resolution techniques for multi-face recognition using light-field camera," in *2013 18th International Conference on Digital Signal Processing (DSP)*, 2013: IEEE, pp. 1-6.
[12] A. Sepas-Moghaddam, P. L. Correia, and F. Pereira, "Light field local binary patterns description for face recognition," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017: IEEE, pp. 3815-3819.
[13] A. Sepas-Moghaddam, F. Pereira, and P. L. Correia, "Ear recognition in a light field imaging framework: a new perspective," *IET Biometrics*, vol. 7, no. 3, pp. 224-231, 2018.
[14] O. S. Good, "Kinect is officially dead. Really. Officially. It's dead," *Polygon. https://www. polygon. com/2017/10/25/16543192/kinect-discontinued-microsoft-announcement. Accessed*, vol. 20, 2017.
[15] K. Khoshelham, "Accuracy analysis of kinect depth data," in *ISPRS workshop laser scanning*, 2011, vol. 38, no. 1.
[16] M. Zollhöfer, M. Martinek, G. Greiner, M. Stamminger, and J. Süßmuth, "Automatic reconstruction of personalized avatars from 3D face scans," *Computer Animation and Virtual Worlds*, vol. 22, no. 2-3, pp. 195-202, 2011.
[17] Y. Cui and D. Stricker, "3d shape scanning with a kinect," in *ACM SIGGRAPH 2011 Posters*, 2011, pp. 1-1.
[18] K. F. Li, K. Lothrop, E. Gill, and S. Lau, "A web-based sign language translator using 3d video processing," in *2011 14th International Conference on Network-Based Information Systems*, 2011: IEEE, pp. 356-361.
[19] A. D. Wilson, "Using a depth camera as a touch sensor," in *ACM international conference on interactive tabletops and surfaces*, 2010, pp. 69-72.
[20] H. Haggag, M. Hossny, S. Nahavandi, and D. Creighton, "Real time ergonomic assessment for assembly operations using kinect," in *2013 UKSim 15th International Conference on Computer Modelling and Simulation*, 2013: IEEE, pp. 495-500.

[21] M. N. K. Boulos, B. J. Blanchard, C. Walker, J. Montero, A. Tripathy, and R. Gutierrez-Osuna, "Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation," ed: BioMed Central, 2011.
[22] S. Izadi *et al.*, "KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 559-568.
[23] M. R. Andersen *et al.*, "Kinect depth sensor evaluation for computer vision applications," *Aarhus University,* pp. 1-37, 2012.
[24] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.
[25] D. Maltoni, "A tutorial on fingerprint recognition," in *Advanced Studies in Biometrics*: Springer, 2005, pp. 43-68.
[26] N. K. Ratha, A. Senior, and R. M. Bolle, "Automated biometrics," in *International Conference on Advances in Pattern Recognition*, 2001: Springer, pp. 447-455.
[27] A. Ross and A. Jain, "Biometric sensor interoperability: A case study in fingerprints," 2004: Springer, pp. 134-145.
[28] S. C. Hoo and H. Ibrahim, "Biometric-Based Attendance Tracking System for Education Sectors: A Literature Survey on Hardware Requirements," *Journal of Sensors*, vol. 2019, 2019.
[29] D. Litwiller, "CCD vs. CMOS: Facts and Fiction. Phothonics Spectra," ed: Laurin Publishing Co. Inc, 2001.
[30] I. D. Aronson, *DV filmmaking: from start to finish*. " O'Reilly Media, Inc.", 2006.
[31] H. Zettl, *Television production handbook*. Nelson Education, 2014.
[32] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C. C. Loy, and X. Wang, "A survey on heterogeneous face recognition: Sketch, infra-red, 3D and low-resolution," *Image and Vision Computing*, vol. 56, pp. 28-48, 2016.
[33] T. D. S. Freitas, "3D face recognition under unconstrained settings using low-cost sensors," 2016.
[34] Z. Cai, "Feature learning for RGB-D data," University of Sheffield, 2017.
[35] B. Y. Li, M. Xue, A. Mian, W. Liu, and A. Krishna, "Robust RGB-D face recognition using Kinect sensor," *Neurocomputing*, vol. 214, pp. 93-108, 2016.
[36] M. Du, A. C. Sankaranarayanan, and R. Chellappa, "Robust face recognition from multi-view videos," *IEEE transactions on image processing*, vol. 23, no. 3, pp. 1105-1117, 2014.
[37] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters,* vol. 24, no. 13, pp. 2115-2125, 2003.
[38] D. Dansereau, "Light field toolbox v0. 4," *MathWorks R File Exhange. MathWorks R,* 2015.
[39] S. G. Marto, N. B. Monteiro, J. P. Barreto, and J. A. Gaspar, "Structure from plenoptic imaging," in *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2017: IEEE, pp. 338-343.
[40] N. B. Monteiro, S. Marto, J. P. Barreto, and J. Gaspar, "Depth range accuracy for plenoptic cameras," *Computer Vision and Image Understanding*, vol. 168, pp. 104-117, 2018.
[41] H.-G. Jeon *et al.*, "Accurate depth map estimation from a lenslet light field camera," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1547-1555.
[42] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[43] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 1-27, 2011.
[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[45] Intel. "Introducing the Intel® RealSense™ R200 Camera (world facing)." https://software.intel.com/en-us/articles/realsenser200-camera (accessed 30 November 2020.
[46] Intel. "A Comparison of Intel® RealSense™ Front-Facing Camera SR300 and F200." https://software.intel.com/en-us/articles/a-comparison-of-intel-realsensetm-front-facing-camera-sr300-andf200. (accessed 30 November 2020.
[47] L. Luna-Oliva *et al.*, "Kinect Xbox 360 as a therapeutic modality for children with cerebral palsy in a school environment: a preliminary study," *NeuroRehabilitation*, vol. 33, no. 4, pp. 513-521, 2013.
[48] L. Shires, S. Battersby, J. Lewis, D. Brown, N. Sherkat, and P. Standen, "Enhancing the tracking capabilities of the Microsoft Kinect for stroke rehabilitation," in *2013 IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH)*, 2013: IEEE, pp. 1-8.