IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 April 2024, accepted 17 May 2024, date of publication 23 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404554

RESEARCH ARTICLE

PIXGAN-Drone: 3D Avatar of Human Body Reconstruction From Multi-View 2D Images

ALI SALIM RASHEED[1,2], MARWA JABBERI[2], (Member, IEEE), TAREK M. HAMDANI[2,3], (Senior Member, IEEE), AND ADEL M. ALIMI[2,4], (Senior Member, IEEE)

[1]Department of Media Technology and Communications Engineering, College of Engineering, University of Information Technology and Communications, Baghdad 00964, Iraq
[2]REsearch Groups in Intelligent Machines (ReGIM-Lab), National Engineering School of Sfax (ENIS), University of Sfax, Sfax 3038, Tunisia
[3]Higher Institute of Computer Science Mahdia (ISIMa), University of Monastir, Monastir 5147, Tunisia
[4]Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, Johannesburg 2006, South Africa

Corresponding author: Ali Salim Rasheed (ali.rasheed@uoitc.edu.iq)

This work was supported in part by U.S. Department of Commerce under Grant BS123456, and in part by the Ministry of Higher Education and Scientific Research of Tunisia under Grant LR11ES48.

ABSTRACT Study is being conducted on training Generative Adversarial Networks (GANs) from 2D datasets to generate 3D human body avatars. Numerous applications, such as virtual reality, sports analysis, cinematography, surveillance, and more, have advanced significantly as a result of the promising research in this subject. Aerial photography sensors together with drone active tracking can remove occlusions and enable 3D avatar body reconstruction by avoiding obstacles and generating high-resolution, rich-information multi-view (RGB) photos. Training failures of 3D avatar reconstruction techniques lead to distortions and loss of features in 3D reconstructed models due to several reasons, including limited viewpoint coverage, visible occlusions, and texture disappearance. The recently developed end-to-end trainable deep neural network technique This work presents PIXGAN-Drone, a photo-realistic 3D avatar reconstruction system for the human body from multi-view photos. To create high-resolution 2D models, is predicated on integrating aerial photography sensors (a steady autonomous circular motion system) coupled with active tracking drones into the Pix2Pix GANs training framework. Accurate and realistic 3D models can be created with conditional image-to-image translation and dynamic aerial views. This study used tests on several datasets to show that our approach outperforms state-of-the-art approaches for a variety of metrics (Chamfer, P2S, and CED). Our 3D reconstructed human avatars in RenderPeople were 0.0293, 0.0271, and 0.0232; on People Snapshot (inside), 0.0133, 0.0136, 0.0050; on People Snapshot (outdoor), 0.0154, 0.0101, 0.0063; and on Custom data-drone (collected dataset), 0.0316, 0.0275, 0.0216.

INDEX TERMS 3D human avatar, 3D reconstruction, PIX2PIXGAN, body model rendering, drone active tracking.

I. INTRODUCTION

3D human avatar reconstruction methods from 2D images have achieved a significant role in computer vision and interactive virtual graphic environments, as well as being the focus of attention in the animation industry community [1]. Recently, deep adversarial approaches have often been used for 3D models of human body reconstruction from 2D images. These methods present a challenge in the depth estimation of a captured photographic scene. Several

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

details are disappearing in addition to the experiencing fusion of some angles in the 3D body rendering from 2D images.

For this reason, full immersion in 3D scenes has become a pressing need that accompanies rapid digital progress. It provides a more realistic and comprehensive vision to understand the intricate details that are difficult to visualize. This requires comprehensive photographic scenes with high-quality, non-dim angles. Thus, 2D multi-view images of high-resolution and sharpened surfaces can be captured, resulting in advanced results during 3D reconstruction [2], [3], [4], [5].

Furthermore, image-to-image translation techniques demonstrated an inspiring revolution [6]. They provided remarkable performance in generating images with realistic and high-resolution scenes by training them according to paired datasets, which proved their efficiency in creating 2D models with clear patterns and rough features, free from distortions and occlusions [7].

Camera-equipped drones have the active ability to capture a large number of high-resolution 2D images in a short period. Circular motion tracking paths enable high-resolution scenes without noise or blurry effects. This is a key factor in restoring 3D models saturated with realistic details from scenes collected by this powerful acquisition device [8]. The technical development of drone camera sensors has provided new capabilities for capturing aerial information in the form of high-resolution images from different destinations, angles, and heights [9]. allowing the collection of multiple images of the human scene, from which data sets can be generated that are considered as training inputs for many algorithms and applications.

In this approach, we propose a PIXGAN-Drone model to reconstruct 3D avatars of the human body from 2D multi-view images by employing the technical capabilities of DJI Mavic Mini2 drones. In particular, their ability to capture high-resolution aerial scenes automatically by the (Active Tracking System) [10]. An image-to-image GAN for salient features improvement is used to generate realistic 2D human subjects with a detailed and integrated visual representation of their surface texture [11], [12].

The presented work highlights the limitations of traditional human body reconstruction methods, such as occlusions, fusion of angles, body details, and complexity in composition due to limited viewpoints. Also, a custom-drone dataset is created using drones with some processing steps using computer vision applications to achieve images with ideal details from different points of view.

Our approach also seeks to enhance deep 3D avatars of full-body human reconstruction methods like PIFu [30], and PIFuHD [32] by creating main data collection of high-resolution multi-view 2D images without blur, noise, and distortion effects or damages. We establish a hybridization of automatic circular motion (DJI Active Tracking) to capture quality video [13] with all directions and predicate 2D model by PIX2PIX GAN algorithm.

The remainder of this paper is structured as follows: A related work in Section II on the existing approaches of 3D reconstruction from 2D images, including a brief description of the objectives of our proposed method. In Section III, the proposed approach is presented and we provide an overview of our pipeline with the justification of each step during this research. In Section IV, we exposed the experimental qualitative, quantitative, and graphical results. A comparison with a detailed analysis followed by a deep discussion to improve the efficiency and robustness of our proposed work is carried out. In Section V, we present the main conclusion of the paper.

Overall, the purpose of this study is to provide insights into the PIXGAN-Drone technique and its potential applications in a variety of sectors. The conclusions and discoveries reported in this research provide the basis for future developments in 3D avatars of human body reconstruction and add to the increasing body of knowledge in computer vision, graphics, and drone-based data collection approaches.

II. RELATED WORKS

A. ELEVATING DEEP-LEARNING ANALYSIS: THE IMPACT OF DRONE-GENERATED 2D IMAGERY ON FEATURE EXTRACTION

Unmanned Aerial Vehicles (UAVs) provide technology that can identify and capture targets from more than one side during a particular scene. Recently, drones have been widely used by specialists in the field of object imaging and surveillance, as well as in various fields such as research, military, and medical applications [14], [15], [16], [17]. Automating these intelligent acquisition devices is one of the challenges for programmers using deep learning and computer vision algorithms, as they can execute commands more efficiently and target details that are difficult for traditional imaging devices. Due to the presence of hard obstructions that affect the horizontal field of view when using cameras or traditional photography methods, it is necessary to invest in these technologies to allow UAVs to provide integrated aerial or horizontal images [18].

Le and Aryafar [19], presented a mechanism between object tracking and object detection in which the purpose is running object tracking and detection algorithms simultaneously. Continuous advances in machine learning and the growth of smart algorithm-driven applications have led to an increasing demand for smaller drones that are easier to transport and use, especially those powered by pioneering DJI innovations. Drone design has changed a lot in recent years, especially in terms of the way it flies, the angles of capture lenses in the wild, and the high quality provided for the scenes. A wonderful panoramic aerial photos of scenes are very realistic [20].

Georgiou et al. [21] used drones to film objects at crime scenes and discovered realistic evidence in real-time, which represents a great effort and advanced results. Feeding

computer vision algorithms by videos provided by these drones to analyze these targeted scenes provided instant results. This work shows that there is a great correspondence between the application of machine learning and computer vision combined with UAVs in finding quantitative and real-time solutions that can provide efforts and advances to this field. In [22], we provided a comparative study on feature extraction from 2D images captured by different devices such as smartphones and drones with application to 3D reconstruction. We show how far this can affect the quality of the reconstruction.

B. 3D RECONSTRUCTION FROM 2D IMAGES WITH APPLICATION TO HUMAN BODY AVATARS GENERATION

In the application of 3D reconstruction of human faces or bodies using 2D images, standard conditions must be met to create 2D images of high quality without blurring or other negative effects that can cause distortions in the 3D reconstructed model. There are several approaches with numerous applications with regard to this research scope in the state of the art which implies that this subject is a novelty devoured by several researchers.

Garg et al. [23], performed human motion reconstruction in a real-time environment using two drones with a multi-camera control system. To evaluate the simulated performance of jogging and playing football in the real world, in [24] authors proposed an applied technical contribution to previous deep-fused surface reconstruction frameworks by combining deep-fused surface reconstruction with The system includes an inertial odometry estimator with powerful camera tracking and high-precision 3D scanning capabilities. Mishra et al. [25], utilized OpenCV technologies [26] to reduce noise through background removal, extract features, and process the image results to create a painting robot platform. Xu and Lu [27], developed a simple GUI open-source for object imaging, intending to bridge the gap between machine vision technology and users. The GUI was developed in Python with the use of OpenCV libraries for image collection, transformation, and visualization. In [28], a method is presented using a mobile-captured multi-view RGBD dataset containing high-precision 3D ground-truth annotations of 153 models. The reference camera position for evaluation is then calculated using the camera calibration algorithm from the OpenCV library. In the literature there are two categories: 3D reconstruction from a single 2D image for specific scenarios such as capturing a single image of a criminal which is very challenging and 3D reconstruction from multiple 2D images provided from video frames such as filming a scene. Given that the first category only has one input 2D image, it is clear that the findings could be affected by the lack of details and information loss. Nonetheless, the use of numerous 2D images enhances 3D reconstructed renders. In Tables 1 and 2, we provided a summary of the most widespread methods on two categories of 3D human body avatar reconstruction in the wild.

TABLE 1. 3D human reconstruction from single 2D image.

Author /Year	Dataset	Models	Observation
Georgios Pavlakos et al. [29] (2019)	MoCap, CMU, Human3.6M	Expressive Body Capture	2D features extraction based on SMPLify. 3D models of human body generation based on deep pre-trained CNNs.
Shunsuke Saito et al. [30] (2019)	DeepFashion, BUFF, RenderPeople	PIFu	2D feature extraction based on pixel-aligned implicit function. 3D reconstruction of high-quality human clothes based on 3D and 2D surfaces features fusion.
Alldieck et al. [31] (2019)	People-Snapshot, DeepFashion, 3DPW	Tex2shape	2D individual image translation based on the appearance map.
Shunsuke Saito et al. [32] (2020)	BUFF, RenderPeople, People-Snapshot	PIFuHD	3D high-resolution reconstruction of 3D human parametric models from 2D single image based on training a multi-level structure to get a rougher plane surface and more accurate details.
Tong He et al. [33] (2020)	DeepHuman	Geo-PIFu	3D human mesh recovering from a monocular RGB images based on 3D U-Net.
Zerong Zheng et al. [34] (2020)	BUFF, TWINDOM, DeepFashion, DeepHuman	PaMIR	Parametric Model Conditioned Implicit Representation (PaMIR) to enhance 3D reconstructed models from 2D images based on implicit function combination with the model.
Anjun Chen et al. [35] (2022)	large-scale mmBody	ImmFusion	ImmFusion application based on mmWave radars to reconstruct 3D human mesh from RGB images in bad weather.

C. GENERATIVE MODELS FOR 3D RECONSTRUCTION

Generative models attracted researchers following high feature space generation performance and good predictions. Losses structure of 2D images in pix2pix GAN problems are re-dominantly formulated as per 2D image pixel classification rate or regression formulated as per 2D image pixel classification rate or regression [41], [42], [43], [44]. These formulations present the unstructured output area in the orientation of output pixels, which means each of them is independent of all others in input 2D images. In our context, GAN can find realistic human faces and bodies from unreal art graphics this algorithm is also named conditional image translation.

The image-to-image translation strategy is used comfortably traits with the Conditional GAN (CGAN) that has been image framed to understand the correspondence between a source input and a target image by a training dataset

TABLE 2. 3D human body reconstruction from multiple 2D images.

Author /Year	Dataset	Models	Observation
Muhammed et al. [36] (2020)	AMASS	Vibe	3D body estimation based on SMPL body model parameters using temporal shape regression network.
Hongwen Zhang et al. [37] (2021)	Human 3.6M, 3DPW, LSP, COCO	PyMAF	2D Parameters prediction based on aligned 2D images to improve 3D mesh reconstruction.
Tang et al. [38] (2022)	PiGraphs, PROX	HSIs	Human-scene interactions (HSIs) used for 3D human avatar reconstruction based on video scene captured with an RGB camera.
Erik Gartner et al. [39] (2022)	Diffphy	AIST, Human 3.6M	Diffphy model for explained 3D human motion reconstruction from video frames.
Ruizhi Shao et al. [40] (2022)	DiffuStereo	THuman2.0	Advanced multi-layer stereo network use to handle high-resolution 4k images for depth estimation. Diffusion kernel and additional stereo constraints are provided.

matched set of images [45], [46]. In [47], Choi and Kim introduced a new architectural method for the generator model and discriminator. U-net is used as a generator and the convolutional "PatchGAN" classifier as a discriminator. In [48], authors aimed to build a regression framework between image depth and 3D faces encoded by expression, identity, and pose parameters using an unsupervised network based on (CycleGAN). Regression models learned from synthetic data authentic acquired noisy depth images. Researchers in [49], introduced the MOST-GAN approach that combines the expressiveness and realism of Style-Based GANs with physical disentanglement and flexibility in nonlinear 3D deformable models. However, MOST-GAN enables the photo-realistic processing of portrait images with full-resolution 3D control over their physical characteristics, allowing extreme manipulation of lighting, facial expression, and pose changes, all the way to multi-views.

D. 3D RECONSTRUCTION OF HUMAN AVATARS FROM 2D IMAGES

In PIFuHD [32], authors presented a high-quality 3D reconstruction of the 3D human body using a deep network trained on BUFF dataset [50] to estimate the human pose and body features in 3D representation. 3D reconstruction stage incorporates high-resolution image features (512×512) from higher-resolution input images (1024×1024). while the second module predicts an occupancy probability field using high-resolution embedding. To improve reconstruction quality and accuracy, standard patterns for both the front as well as the rear sides are projected and given as supplementary input using a neural network architecture and multi-layer perception (MLP). More precisely F(X) is considered as the predicted value in a virtual environment(VE) for any 3D point position $X = (X_1, X_2, X_3) \in R^3$, and I is a high-resolution 2D gray image.

F(X,I)= 1 if X is inside the mesh surface.

F(X,I)= 0 otherwise.

The function F is then trained using a neural network architecture, and an image feature embedding is extracted from the projected 2D position at $\Phi(X) = x \in R^2$ which we denote by $\Phi(x, I)$.

Perpendicular projection is used for π, and thus $x = \pi(X) = (X_1, X_2)$. Then, it estimates 3D point X, and thus:

$$\mathrm{f}(\mathbf{X}, \mathbf{I}) = g(\Phi(\mathbf{x}, \mathbf{I}), Z) \tag{1}$$

where Z = X_z is the depth specified by the 2D projection x along the ray.

$$\mathrm{f}^H(\mathbf{X}) = g^H\left(\Phi^H\left(\mathbf{x}_H, \mathbf{I}_H, \mathbf{F}_H, \mathbf{B}_H, \right), \Omega(\mathbf{X})\right) \tag{2}$$

where $\mathrm{I}_(H), F_(H), B_(H)$ are the input image map in the frontal and backside. $(\mathbf{X})$ denotes a 3D human model reconstructed and rendered in a blender virtual environment to display the 3D human reconstruction [51].

One of the primary technical deficiencies found in the referenced works is the 3D reconstruction's resilience and performance. That is, due to the uncertainty that arises from converting RGB pixels from 2D areas into a 3D environment, the majority of the work uses monocular photos. The kinetic structures of the body vanish as a result, and certain intricate junctions in the vicinity fuse together. More specifically, the development process of GAN algorithms is adversely affected by the capture of images of ambiguous datasets, in terms of their primary characteristics and the predominance of blur and distortions on their surface texture [52], [53], [54]. This results in the generation of a 3D model that has missing corners of its 3D physical mesh and obstructions in its tiny details. This led us to utilize the drone's active tracking technology, which records visual scenes automatically and produces stable, vibration-free photos, as we described in one of our works [22]. In fact, substantial detail and high resolution can be found in the source photos' excellent quality and effective acquisition [55], which leads to a reliable and thorough 3D reconstruction of the human body [56], [57].

III. PROPOSED METHOD

Our approach aims to improve the results of 3D avatar of human body reconstruction from multi-view 2D images through three main stages as shown in Figure 1. Also, algorithm 1 shows the main steps.

A. DATA COLLECTION

This phase not only concerns the acquisition of 2D images but also the different pre-processing steps. We enable DJI

Algorithm 1 Overall

```
Begin
[2D_images] ← Data_Acquisition (Device, Data_Folder)
for each image in [2D_images] Do do
  [2D_images_Processed] ← 2D_Images_Processing (2D_images)
  [2D_realistic_images] ← 2D_Images_Generation (2D_images_Processed)
  [3D_Human_avatar] ← PifuHD (2D_realistic_images)
end for
End
```

Active Track which is an automatic tracking technology that captures objects or people from all angles, locating dark places and invisible curves that standard cameras cannot reach. Then we used the OpenCV toolboxes to remove blurry images while maintaining sharp images as they have greater detail, clarity, and quality without blurring. Unblurred images were segmented using the Computer Vision Annotation Tools to create datasets (image sources, segmentations, or depth maps). The main steps of Data collection and 2D images processing are presented in the following algorithm.

Algorithm 2 Data_Collection

```
Begin
[2D_images] ← VideoCapture (Data_Folder)
for each image in [2D_images] Do do
  [2D_images_gray] ← Convert_GrayScale (2D_images)
  [2D_images_smooth] ← Variance_of_Laplacian (2D_images_gray)
  [2D_images_Processed] ← Save (2D_images_Smooth)
end for
End
```

B. 2D MODEL GENERATION

We trained PIX2PIX GAN to predicate new 2D human features from multi-view 2D images using two inputs our main datasets (image source, segmentation, or depth map) and public benchmarks. PIX2PIX GANs approach allows for finding 2D models with more details and high-quality texture. The generator and discriminator are trained simultaneously. In fact, the generator attempts to make pictures realistic enough to mislead the discriminator. Discriminator attempts to accurately determine whether an image is genuine or fake. Algorithm 3 shows the main steps of 2D images generation using Pix2Pix GAN.

C. 3D AVATAR RECONSTRUCTION

We perform this stage by reconstructing a high-resolution 3D avatar from 2D predicted images coming from PIX2PIX GAN. However, the goal of our approach is to reinforce deep 3D avatar reconstruction methods like PIFu [30], and PIFuHD [38] by increasing the occupancy value of $X = (X_x, X_y, X_z) \epsilon R^3$ dense 3D volume F(X) for any 3D features

Algorithm 3 2D_Image_Generation

```
Begin
for each image in [2D_images_Processed] Do do
  img ← Resize (2D_images_Processed, 256 × 256)
  [image_shape] ← img.shape [1:]
end for
Discriminator_model ← Discriminator (image_shape)
Generator_model ← Generator (image_shape)
GAN_model ← GAN (Generator_model, Discriminator_model, image_shape)
data ← Balance_data_merge (2D_image_Processed, target)
[Dataset] ← Preprocess_Data (data)
Train (Discriminator_model, Generator_model, Dataset, epochs, n_batch)
[2D_realistic_images] ← GAN_model.predict (2D_images_Processed)
End
```

position in continuous camera space. The overall framework of the PIXGAN-Drone pipeline will be more detailed later.

D. DATA COLLECTION

Since unmanned aircraft equipped with high-resolution lenses became available for public and daily uses, it has achieved widespread fame due to the high technical capabilities it provides through accuracy in imaging and stability during flight. Because it has a DJI Active Tracking system, it provides integrated scenes that are free from blur and noise, which may have proven to be an important option for capturing films with a higher level of quality [58].

Using a DJI Mavic Mini2 drone equipped with a 12-megapixel camera and a resolution of 1920 × 1080 HD pixels moving approximately 5 meters away from the human target model, we capture 2D images (create a custom dataset). Here, details offered by drones during scene capture are very sophisticated compared with those of other portable imaging devices as explained in [22].

The DJI Active Track system is a technology that may allow automatic tracking of persons or targets that are to be monitored using drones like the DJI Mavic Mini2 model. This technology can be found in applications that operate drones. The robot will make use of its vision as well as its frameworks for identifying frames in order to maintain safe flight while also having the potential to achieve the best results possible for photography.

The real-time object detection through the circular tracking technique aims to locate an object in a manner similar to human perception, utilizing bounding boxes as shown in Figure 2. The rectangular surrounding box is used to identify the central region of the image as well as the direction of motion characterized by a significant speed [59].

The stabilization of the drone's circular tracking system as shown in Figure 3, is used to capture the human model from all angles to enable it to locate dark areas and invisible curves that a conventional camera cannot reach. Above all, compared

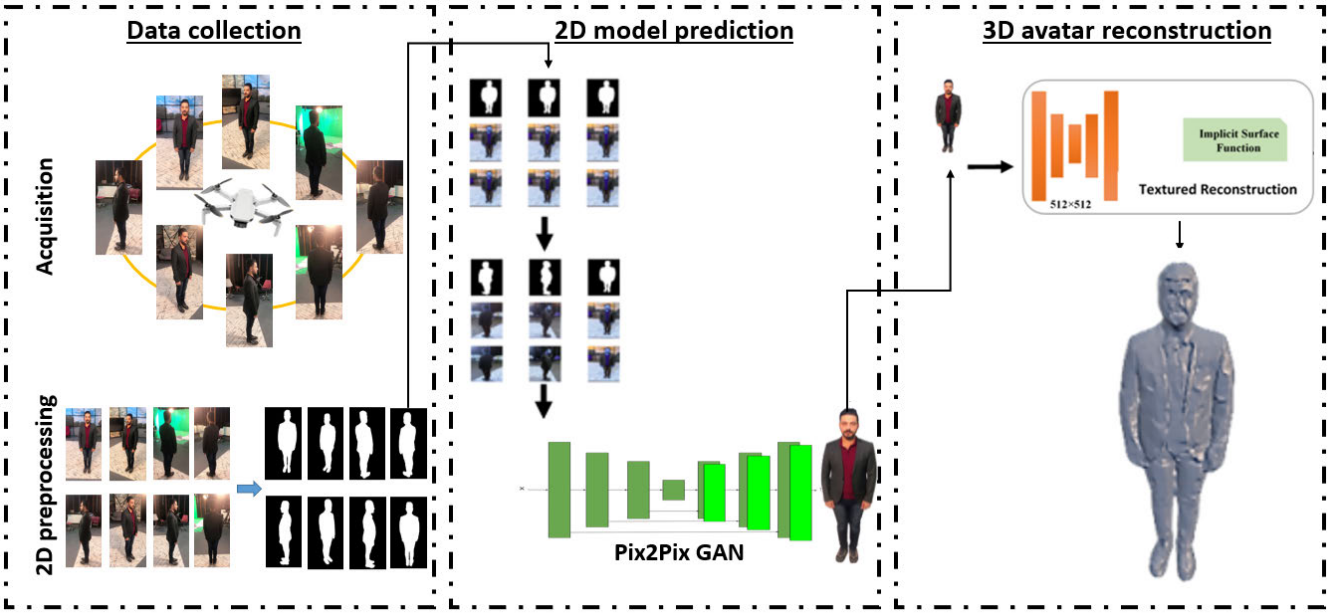


FIGURE 1. Overview of our proposed method.

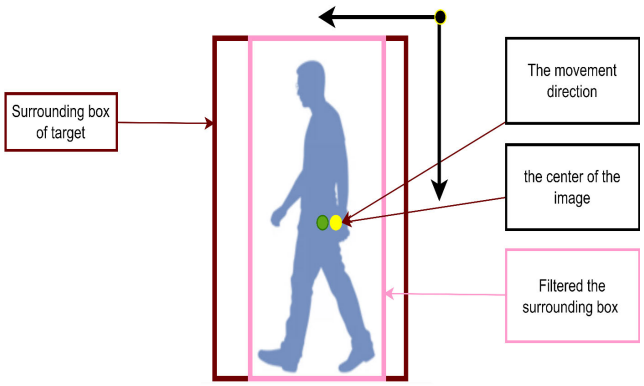


FIGURE 2. Parameters detection in a human body.

FIGURE 3. DJI drone active circular Tracking capture system.

to conventional cameras, the circular tracking of drones is stable, with no sudden movements or trembling, whereas the hand holding the camera is prone to sudden movements and shaking.

Equation (3) explains the control gimbal tracking, considering the center of the image scene is an input. The difference between the filter box and image center in both vertical and horizontal directions is a control produced for gimbal tracking. KP is a proportional coefficient, and Kd is a derivative coefficient [60].

$$u(t) = K_p e(t) + k_d(de(t)/dt) \tag{3}$$

The controller of the DJI Mavic mini drone option provides the possibility of surrounding the target with a rectangle as shown in Figure 4, specifying the target center point (x) for the video scene, and through a stable and balanced circular movement, the target is captured from all angles.

To capture 2D images of the human body by drone flight, the on-orbit trajectory is designed to capture full-frame images with sufficient photo overlap to reconstruct the 3D model using a photogrammetry solution. In summary, in-orbit video images (1920 × 1080 pixel size) were selected for 3D model reconstruction using UAV photogrammetry. Total of 1200 high-resolution still images (dimensions 1080×1920 pixels).

The video captured by circular tracking of the drone's camera must be of high quality and non-blurry in order to achieve these features, OpenCv is applied and that means RGB images will be converted to grayscale images, which are real characteristics of a 2D image that is used during the applications of reconstructing 3D shapes details [61].

FIGURE 4. Bounding Box adjustment to the full body.

The two-dimensional Laplacian sharpening equation is a second-order partial derivative (Laplacian operator) in the orthogonal directions of continuous space to be more suitable for digital image sharpening and less noisy [62]. Laplacian digital filter pixels as shown in Figure 5.

0	1	0
1	-4	1
0	1	0

FIGURE 5. Laplacian operator template.

2D image transformation using the Laplacian operator is expressed:

$$u(t) = \nabla f(x, y) = \partial^2 f/\partial x^2 + \partial^2 f/\partial y^2 \quad (4)$$

1) Drone-captured scene Segmentation into frames: Supports 3D human reconstruction from multi-view 2D images by extracting sharpened features from the target (human) and removing noisy, blurry images using OpenCV python coding threshold=100 to discover useful information.
2) Data organization and preparation: Used in machine learning and other data-driven applications is known as dataset creation. A dataset is a data collection used to train, verify, and test machine learning models.
3) Datasets creation based on the project's specific needs and aims among the most frequent approaches is: Data collection, one technique for collecting a dataset for 3D reconstruction of 2D images would be to acquire the data directly using 3D scanning technologies, such as structured light scanners or laser scanners, by capturing high-resolution 3D images of objects or scenes and storing the resulting point cloud data, or to collect the data using the appropriate tools to capture 2D images.
4) Data extraction from internet sources: such as online repositories of 3D models or online datasets of 2D images and corresponding 3D models, is another method for constructing a dataset for 3D reconstruction of 2D images [63].
5) Data augmentation: to produce new data from existing 2D images and 3D models is a third way of producing a dataset for the 3D reconstruction of 2D images. Rotating, resizing, and cropping images, as well as applying different sorts of noise to the images, are instances of data augmentation techniques [64].
6) Data synthesis: Using data synthesis techniques to produce synthetic 2D images and related 3D models is a fourth way of building a dataset for the 3D reconstruction of 2D images. Using computer graphics software to produce realistic 3D models and rendering them from multiple views to create 2D images might be instances of data synthesis approaches [65].

Before selecting the best method or combination of methods for the specific research topic at hand, it is crucial to comprehensively evaluate the advantages and disadvantages of various approaches to dataset development.

To collect data using the appropriate tools of 2D images capturing in the wild in order to perform 3D reconstruction, machine learning, and deep learning algorithms are used to produce 3D models from 2D images using adversarial techniques for the 3D reconstruction of 2D images. To train and assess the performance of the algorithms, these techniques frequently rely on enormous datasets of 2D images and their matching 3D models [66].

Basically, steps of preprocessing are as follow:

1) Collect images using the drone's camera.
2) Removing blurry images using Open CV documentation while retaining sharpened images that are not blurry due to their higher detail, clarity, and quality.
3) The non-blurry images were segmented using computer vision annotation tools (CAVTs) [67].

Preprocessing steps carried out on 2D images are shown in Figure 6.

In Figure 7, we show the impact of CAVTs when producing data collection (image source, segmentation, or depth map).

E. 2D MODEL GENERATION

This study is going to deal with GANs, and as stated in the introduction, Porkodi et al. [68]. Created generative adversarial networks (GANs), a kind of deep learning model, in 2014. They are made up of a generator and a discriminator,

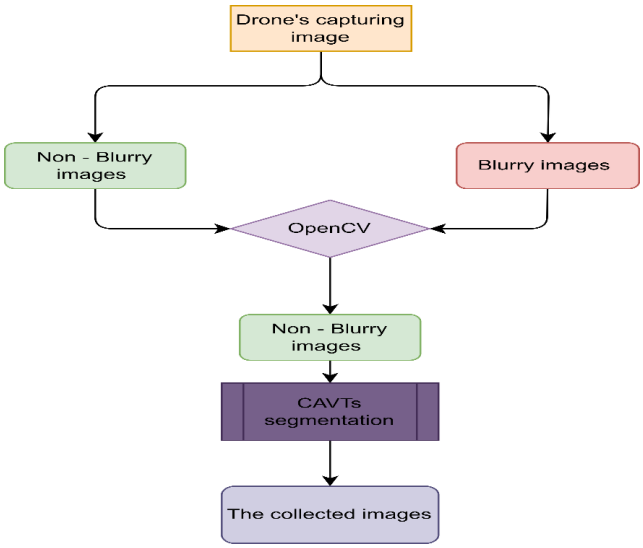


FIGURE 6. Data collection steps.

FIGURE 7. The main data collection (image source, segmentation).

two deep networks that compete with one another in a two-player minimax game.

The discriminator is taught to discriminate between actual and produced instances, while the generator is trained to generate samples that are comparable to a collection of real cases. The two networks are trained alternately, with the generator attempting to make samples that the discriminator cannot tell apart from genuine instances, and the discriminator attempting to reliably determine whether each sample is real or produced [69].

Typically, the generator is trained to maximize the probability that the discriminator will categorize its produced samples as genuine, whereas the discriminator is designed to maximize the probability of properly categorizing both actual and generated data. This may be formalized as a minimax game with the following objective function:

$$V(D, G) = Ex \approx Pdata(x)[logD_(x)] + Ez \approx_p z(z)[log(1 - D(Gz))] \quad (5)$$

where:

G: is the generator.

D: is the discriminator.

X: is the real sample from the data distribution. (Pdata(x))

Z: is a noise sample from a noise distribution (Pz(z)).

The generator attempts to generate samples that are similar to real samples by transforming noise samples Z with the function G(z), whereas the discriminator attempts to distinguish between real and generated samples by using the function D(x), which returns a probability that X is a real sample. The GAN training procedure is outlined below:

1) Obtain a set of noise samples Z1,…, Zm from the noise distribution.(Pz(z)).
2) Obtain a set of real samples 1,…, m from the data distribution (Pdata(x)).
3) Using the generator, create a batch of fake samples G(Z1),…, G(Zm).
4) Using the genuine and fake samples, update the discriminator by maximizing the objective function V(D,G) with respect to D.
5) Using the noise samples 1,…, m, update the generator by minimizing the objective function V(D,G) with respect to G.

This procedure is continued until the generator and discriminator find an equivalence, at which time the generator ought to be able to generate samples that are identical to genuine samples.

PIX2PIX GANs, It is a form of generative adversarial network (GAN) developed by UC Berkeley researchers in 2016. It is intended for image-to-image translation problems, with the objective of learning a function that can convert an input image into a corresponding output image in order to reinforce the PIFuHD and get higher accuracy and quality than has been achieved in it when the results are compared between the current study and the PIFuHD according to the criteria that the PIFuHD dealt with, as will be seen in the qualitative, quantitative, and visual evaluation.

A PIX2PIX GAN's, architecture is made up of two basic components: a generator and a discriminator, as mentioned in the elaboration of GANs since it's a type of GAN algorithm. The generator is in charge of creating the output image, while the discriminator is in charge of assessing the realism of the created image and delivering input to the generator [70].

In a two-player minimax game, the generator and discriminator are trained concurrently, with the generator attempting to make pictures realistic enough to mislead the discriminator and the discriminator attempting to accurately determine whether an image is genuine or fake [71]. The discriminator is trained to maximize the loss function, which quantifies the difference between the generated image and the ground truth output image, while the generator is trained to decrease it. The PIX2PIX GAN's, overall loss function is expressed as:

$$L = L(D) + L(G) \quad (6)$$

where: L: is the overall loss. L(D): is the loss for the discriminator where:

$$L(D) = -E[logD_(x, y)] - E[log(1 - D(x, G_(x)))] \quad (7)$$

L(G): is the loss for the generator where:

$$L(D) = -E[log(D(x, G_(x)] \quad (8)$$

The PIX2PIX GANs have been utilized for a number of image-to-image translation applications, including grayscale image-to-color conversion, map-to-satellite image conversion, and drawing-to-photograph conversion. The following Figure 8, and flowchart presented in Figure 9, show how PIX2PIX works.

FIGURE 8. Model training to predict 2D image using PIX2PIX GANs.

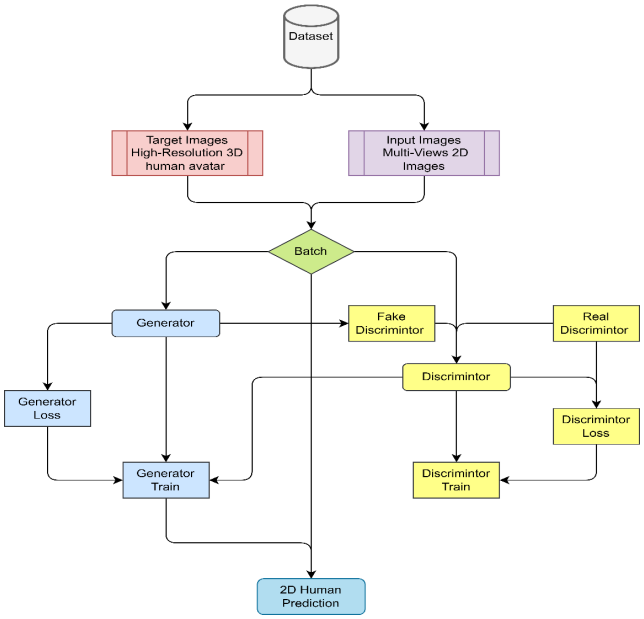


FIGURE 9. Image-to-image translation steps.

F. 3D AVATAR OF HUMAN RECONSTRUCTION

3D avatar reconstruction from a 2D image is a challenging problem that involves generating a 3D representation of a person's face or body from a single 2D image or multiple 2D images. The approach that has been used is PIX2PIX GAN.

The generated dataset by the drone is going to train a PIX2PIX GAN for 3D avatar reconstruction [72]. The GAN would then be trained to learn the relationship between the 2D images and the 3D avatars so that it can generate a 3D avatar that is a reconstruction of the person in a given 2D image. The following figure shows the result that has been achieved for the main dataset and the ''People Snapshot'' dataset [73]:

IV. EVALUATION

Our method (PIXGAN-Drone) is largely compared with various state-of-the-art most widespread methods, namely Vibe [74], Pymaf [75], SMPLbody [76], ICON [77], Pifu [30], and PifuHD [32], using the RenderPeople [78], People snapshot(indoor) People snapshot(outdoor) [79], and Custom data-drone collecting datasets. The approaches described in our study exhibit variations in several aspects, including the data input training, the employed loss functions, the primary architecture, and the utilization of the SMPL body prior, among others. Three evaluation metrics will be applied, as outlined below:

- **Chamfer distance**: measures between the ground-truth scans and reconstructed meshes are reported. Shape comparison is a methodology employed to assess the degree of resemblance or dissimilarity between two given shapes. The algorithm computes the shortest Euclidean distance between points on one geometric shape and the nearest points on the other geometric shape. A decrease in the Chamfer distance indicates a higher degree of similarity between the shapes [80].
- **Point-to-surface "P2S" distance**: is the shortest distance between a point in 3D space and a mesh or model's surface. Computer graphics, vision, and 3D form analysis use it. The P2S distance between the point and the mesh surface will be in the P2S distance variable [81].
- **Cumulative Error Distribution (CED)**: is a statistical metric that can be used to evaluate the precision of 3D form segmentation algorithms. It measures the difference between the algorithm's segmentation bounds and the ground truth segmentation in three-dimensional space. The best CED metric is when the algorithm's segmentation boundaries match the ground truth boundaries, lowering the total error at all distance thresholds to zero and showing a very small error rate at all thresholds [82].

Our method is grounded in the examination of elements that contribute to the attainment of high levels of accuracy in the process of reconstructing three-dimensional (3D) representations from two-dimensional (2D) data. It is imperative to conduct a quantitative assessment of the MLP (Multi-Level-Pixel-aligned-implicit function) for the 3D avatar, which has been reconstructed from high-resolution 2D images. These images were generated through model training using the PIX2PIXGAN technique, which is recognised as a high-resolution input for the PifuHD method employed in the reconstruction of 3D human bodies. In this process, a deep grid is utilised, and random cropping of 512×512 pixels is performed on the 1024×1024 pixels image. This consideration is made to ensure the accuracy of features, thereby enhancing the fundamental basis of our adopted approach.

The quantitative comparison shown in Table 3, was conducted using the RenderPeople dataset as the primary

TABLE 3. Qualitative comparison state-of-the-art on RenderPeople.

Method	Chamfer	P2S	CED
Vibe	0.0309	0.0305	0.0249
Pymaf	0.0310	0.0308	0.0250
SMPLbody	0.0309	0.0309	0.0249
ICON	0.0306	0.0302	0.0283
Pifu	0.0310	0.0306	0.0295
PifuHD	**0.0280**	0.0272	0.0263
Ours	0.0293	**0.0271**	**0.0232**

input for producing 3D human avatars. This was achieved by training our method's model alongside state-of-art approaches. The results indicate that our method exhibited a clear superiority over other approaches. Additionally, competition with the PifuHD method, which was used as a key factor in our approach by improving its performance and results using PIX2PIXGAN to obtain higher accuracy results during the 3D reconstruction process, Quantitative results method, where the Chamfer metric showed a slight superiority of the pifuHD method, a result of 0.0293 with a lower error value compared with our method, a result of 0.0280, which is a better result compared to other methods used in quantitative comparison. As for the P2S and CED metrics, our method achieved a clear superiority over the previous methods, with values of 0.0271 and 0.0232, respectively.

The error slopes of the quantitative measures (Chamfer, P2S, and CED) exhibit a reduced amplitude in our method (PIXGAN-Drone) when compared to other methods using the Renderpeople dataset, as visually depicted in Figure 10. While the PifuHD technique demonstrates a reduced level of error in the Chamfer measure, our approach consistently achieves superior accuracy in the remaining metrics, resulting in a lower percentage of error.

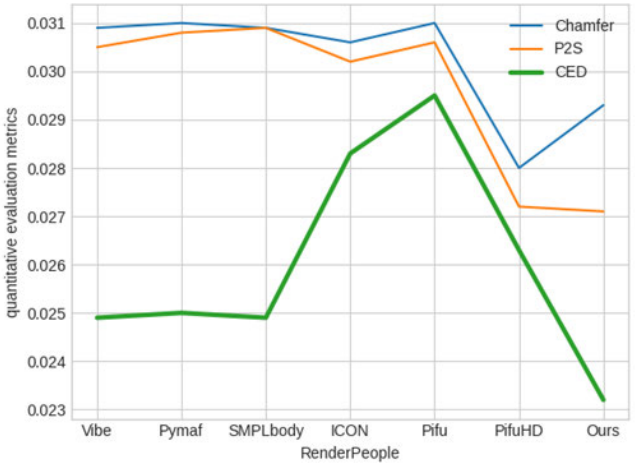


FIGURE 10. Error regression curves of Chamfer, P2S, and CED on Renderpeople.

The visual evaluation of the 3D human avatar reconstructed by our method, as depicted in Figure 11, exhibits a higher degree of realism and clarity in terms of physical details when compared to other methods. Particularly, the avatar's physical curves are more prominent and distinct, devoid of blurring or fusion, thereby enhancing its resemblance to the actual human form. We notice a decline in other approaches for reconstructing legs and feet is evident, but our method demonstrates a noteworthy ability to accurately replicate the surface texture of the reconstructed body parts.

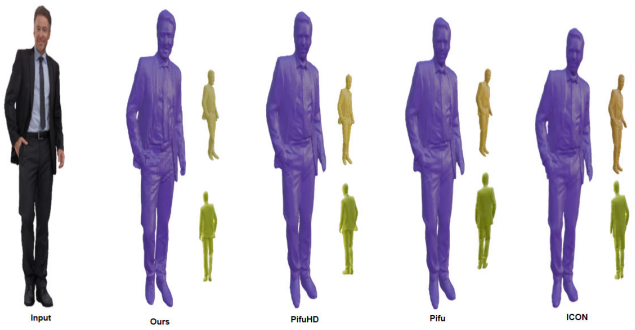


FIGURE 11. Graphical results comparison with state-of-the-art methods on RenderPeople.

In order to improve the quantitative comparison results of our proposed approach (PIXGAN-Drone) in comparison to existing state-of-the-art methods, The People snapshot dataset, encompassing both indoor and outdoor scenes, has been used for training our model as well as other methods. The method employed in our study exhibited better results when compared to all other approaches. The quantitative metrics, namely Chamfer, P2S, and CED findings, were shown in Table 4, with corresponding values of 0.0133, 0.0136, and 0.0050, respectively. Our method shows a reduced level of error in comparison to other methods, particularly the PifuHD method, which is considered competitive with the most widespread approaches in this field.

TABLE 4. Comparison with state-of-the-art on People Snapshot (indoor).

Method	Chamfer	P2S	CED
Vibe	0.0865	0.0839	0.0788
Pymaf	0.0804	0.0781	0.0992
SMPLbody	0.0370	0.0346	0.0332
ICON	0.0315	0.0286	0.0258
Pifu	0.0153	0.0514	0.0464
PifuHD	0.0152	0.0424	0.0071
Ours	**0.0133**	**0.0136**	**0.0050**

The results were as follows: 0.0152, 0.0424, and 0.0071. The quantitative measurements obtained from our method are presented in Table 5, with values of 0.0154, 0.0101, and 0.0063. Therefore, it displays improvements in terms of better quality and reduced errors when compared to the results of state-of-the-art methods, particularly the PifuHD method, which yielded respective results of 0.0492, 0.0470, and 0.0359.

Figures 12 and 13 depicts the error proportion with respect to surface distance, which quantifies the disparity between the produced avatars and the ground truth avatars. The method

TABLE 5. Comparison with state-of-the-art on People Snapshot (outdoor).

Method	Chamfer	P2S	CED
Vibe	0.0773	0.0762	0.0779
Pymaf	0.0826	0.0812	0.0963
SMPLbody	0.0557	0.0538	0.0530
ICON	0.0521	0.0496	0.0446
Pifu	0.0612	0.0595	0.0481
PifuHD	0.0492	0.0470	0.0359
Ours	**0.0154**	**0.0101**	**0.0063**

proposed in this study demonstrates a notable reduction in error percentage when compared to other methods, indicating its capacity to produce avatars with higher precision and more realistic features.

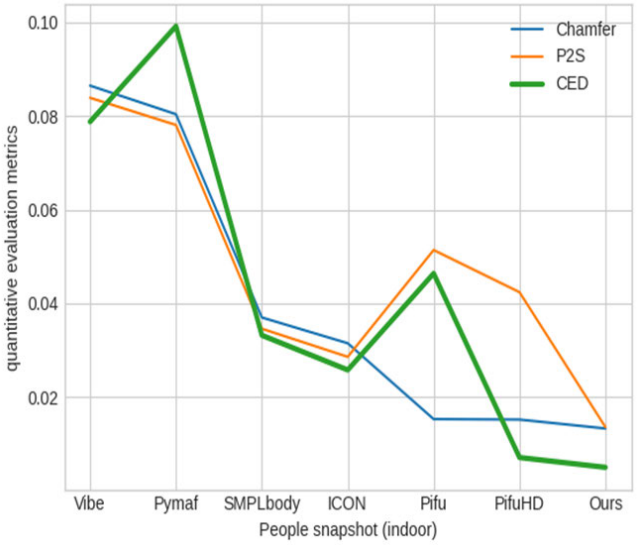


FIGURE 12. Error rates regression on People Snapshot (indoor).

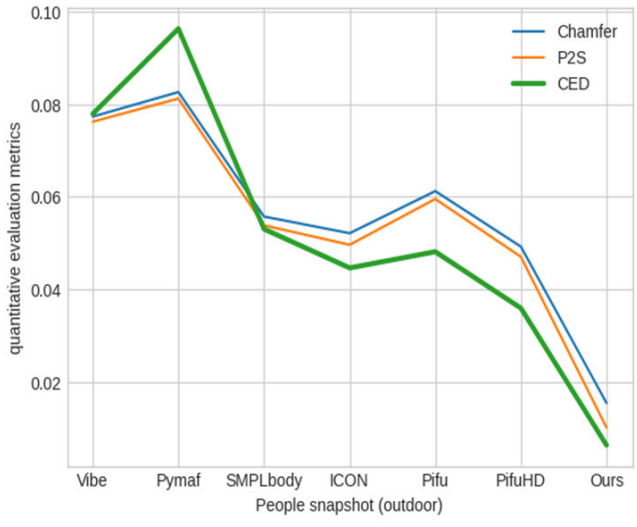


FIGURE 13. Error rates regression on People Snapshot (outdoor).

The visual evaluation is presented in Figures 14 and 15. The avatar produced by the proposed method shows a higher degree of realism and intricacy compared to the avatars generated by other methods. Our method demonstrates the capability to produce avatars that exhibit precise body proportions, seamless skin textures, and authentic muscular detailing. In contrast, the alternative approaches provide avatars characterised by body proportions that are comparatively less realistic. The avatar produced using state of art methods indicates slight variations in body proportions, a somewhat larger cranium and comparatively shorter extremities as compared to our approach (PIXGAN-Drone).

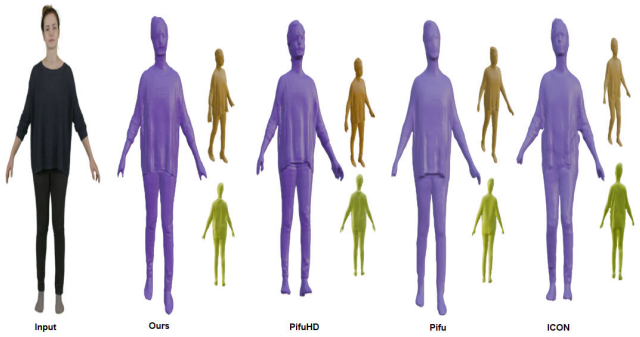


FIGURE 14. Graphical results on People Snapshot (indoor).

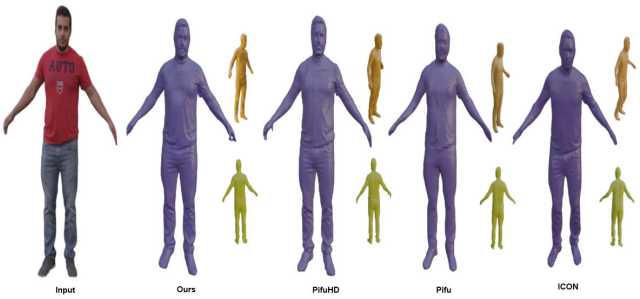


FIGURE 15. Graphical results on People Snapshot (outdoor).

Based on a quantitative comparison using custom data-drone collection, the results presented in Table 6, prove that the PifuHD method surpasses our method in terms of the Chamfer Metric, achieving a value of 0.0192 compared to our method's value of 0.0316. However, it is worth noting that the PifuHD method shows a decline in performance when considering the P2S and CED metrics, with values of 0.0425 and 0.0217, respectively. However, our method (PIXGAN-Drone) demonstrates improvement in these metrics, achieving values of 0.0275 and 0.0216, respectively. These values represent higher percentages of precision when compared to other state-of-the-art methods. Figure 16 illustrates the error rates apparent in the curves of quantitative metrics when comparing our method with other methods employed in the reconstruction of 3D avatars of humans from 2D images. The Chamfer metric curve exhibits a decline in the efficacy of our approach compared to PifuHD. However, we observe improvements in the P2S and CED

metrics across all methods. This outcome is justifiable and admissible in quantitative comparisons, given the minor disparities and competition among the 3D human reconstruction approaches.

TABLE 6. Comparison with state-of-the-art methods on Custom data-drone (collected dataset).

Method	Chamfer	P2S	CED
Vibe	0.0937	0.0919	0.0694
Pymaf	0.0831	0.0811	0.0967
SMPLbody	0.0590	0.0571	0.0480
ICON	0.0519	0.0533	0.0424
Pifu	0.0459	0.0428	0.0351
PifuHD	**0.0192**	0.0425	0.0217
Ours	0.0316	**0.0275**	**0.0216**

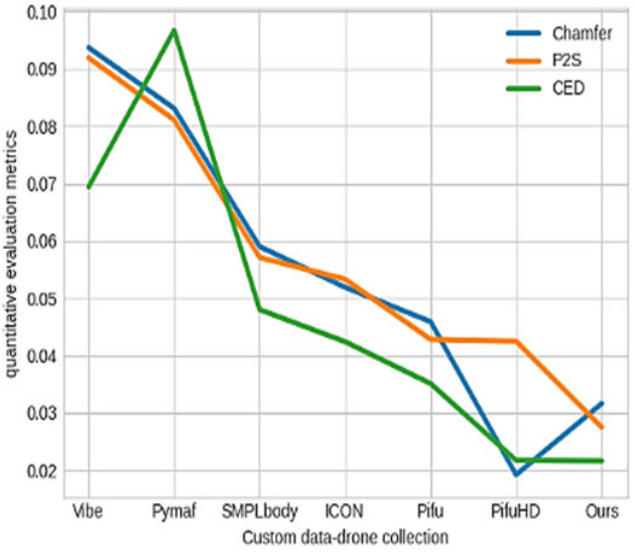


FIGURE 16. Graphical results on Custom data-drone collected.

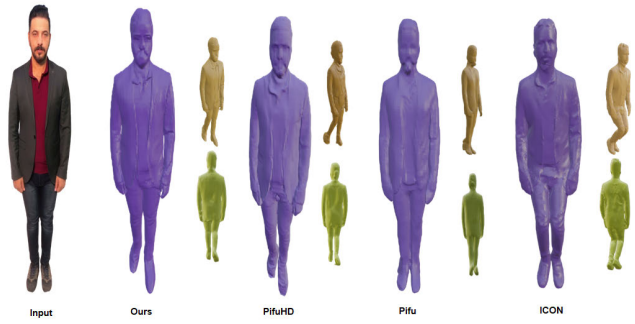


FIGURE 17. Graphical results on Custom data-drone collected.

A visual comparison in Figure 17. of the avatars rendered by the methods used in this approach shows a noticeable competition between our method (PIXGAN-Drone) and the PifuHD method in terms of the closeness of the descriptions and features to the ground truth avatars, as we notice a difference in some features of the face, with real and realistic prominence of body details, and better coordination of the appearance of the feet compared to the other 3D avatars. As for the 3D human avatar that was reconstructed using PifuHD, we notice the appearance of realistic facial details, with the disappearance of body features and confusion in the reconstructed feet. Other methods suffer from obvious failures in the process of 3D reconstruction of human avatars. We also investigated GPU consumption and the time consuming given at the reconstruction step in the convoluted tests. Table 7 presents a comparison with state-of-the-art methods.

TABLE 7. Technical evaluation: Time and GPU consuming during 3D reconstruction step.

Method	Total runtime/Sec.	GPU computation time/Sec.	Data transfer time/Sec.
Vibe	**5.8650**	0.12459	0.0006
Pymaf	5.9366	0.1140	0.0012
SMPLbody	9.5367	0.1190	**0.0005**
ICON	7.7009	0.1331	**0.0002**
Pifu	5.9843	0.1172	0.0012
PifuHD	6.1511	0.1215	0.0011
Ours	**5.8889**	**0.1131**	0.0012

V. CONCLUSION

A systematic approach to reconstructing realistic and detailed 3D avatars of the human body from multi-view 2D images is provided by this system model, which primarily combines the capabilities of the Pix2Pix GAN and active tracking system DJI Mavic Mini2 drone technology to capture high-resolution images from a variety of angles and positions. While the training of Pix2Pix GAN depends primarily on the quality and accuracy of the input images obtained by the drones, this study was able to highlight a number of limitations that should be taken into account. These limitations have an impact on the model's ability to generate a high-resolution 2D representation that may contain occlusions and fused angles. Its performance may be adversely affected by the blurring of its surface roughness and other circumstances, which could lead to a decrease in the effectiveness of the model used to rebuild the 3D human body as well as a limitation in its capacity to merge 2D images. The study's research area suggests a number of future plans wherein depth data from computer vision applications and sophisticated drone photogrammetry techniques can be merged to improve the reconstruction of the 3D avatar of the human body with all of its intricacies. Increasing the number of high-resolution dataset photos also offers potential for high-performance training. This model's simplicity of use for algorithms in real execution time makes it accessible to the general public in the graphics and three-dimensional image-building fields; it doesn't require sophisticated knowledge, and it can be invested in for use in virtual reality applications, cinema animation, applications, and video films that deal with the restoration of celebrities and deceased personalities in applications involving human-computer interaction.

REFERENCES

[1] L. Cui and J. Liu, "Virtual human: A comprehensive survey on academic and applications," *IEEE Access*, vol. 11, pp. 123830–123845, 2023, doi: 10.1109/access.2023.3329573.

[2] S. Sharma and V. Kumar, "3D face reconstruction in deep learning era: A survey," *Arch. Comput. Methods Eng.*, vol. 29, no. 5, pp. 3475–3507, Aug. 2022, doi: 10.1007/s11831-021-09705-4.

[3] J. Gao, J. Liu, and S. Ji, "A general deep learning based framework for 3D reconstruction from multi-view stereo satellite images," *ISPRS J. Photogramm. Remote Sens.*, vol. 195, pp. 446–461, Jan. 2023, doi: 10.1016/j.isprsjprs.2022.12.012.

[4] C. Xu, "Immersive animation scene design in animation language under virtual reality," *Social Netw. Appl. Sci.*, vol. 5, no. 1, p. 42, Dec. 2022, doi: 10.1007/s42452-022-05263-x.

[5] N. Stier, A. Ranjan, A. Colburn, Y. Yan, L. Yang, F. Ma, and B. Angles, "FineRecon: depth-aware feed-forward network for detailed 3D reconstruction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 18377–18386, doi: 10.1109/iccv51070.2023.01689.

[6] J. Lu, J. Öfverstedt, J. Lindblad, and N. Sladoje, "Is image-to-image translation the panacea for multimodal image registration? A comparative study," *PLoS ONE*, vol. 17, no. 11, Nov. 2022, Art. no. e0276196, doi: 10.1371/journal.pone.0276196.

[7] S. Tahri, A. Barateau, C. Cadin, H. Chourak, S. Ribault, F. Nozahic, O. Acosta, J. A. Dowling, P. B. Greer, A. Largent, C. Lafond, R. De Crevoisier, and J. C. Nunes, "A high-performance method of deep learning for prostate MR-only radiotherapy planning using an optimized Pix2Pix architecture," *Phys. Medica*, vol. 103, pp. 108–118, Nov. 2022, doi: 10.1016/j.ejmp.2022.10.003.

[8] M. Y. Arafat, M. M. Alam, and S. Moh, "Vision-based navigation techniques for unmanned aerial vehicles: Review and challenges," *Drones*, vol. 7, no. 2, p. 89, Jan. 2023, doi: 10.3390/drones7020089.

[9] L. Inzerillo, F. Acuto, G. Di Mino, and M. Z. Uddin, "Super-resolution images methodology applied to UAV datasets to road pavement monitoring," *Drones*, vol. 6, no. 7, p. 171, Jul. 2022, doi: 10.3390/drones6070171.

[10] D. Calisi, S. Botta, and A. Cannata, "Integrated surveying, from laser scanning to UAV systems, for detailed documentation of architectural and archeological heritage," *Drones*, vol. 7, no. 9, p. 568, Sep. 2023, doi: 10.3390/drones7090568.

[11] X. Zhong, J. Zhu, W. Liu, C. Hu, Y. Deng, and Z. Wu, "An overview of image generation of industrial surface defects," *Sensors*, vol. 23, no. 19, p. 8160, Sep. 2023, doi: 10.3390/s23198160.

[12] S. Zhou, Y. Wang, W. Jia, M. Wang, Y. Wu, R. Qiao, and Z. Wu, "Automatic responsive-generation of 3D urban morphology coupled with local climate zones using generative adversarial network," *Building Environ.*, vol. 245, Nov. 2023, Art. no. 110855, doi: 10.1016/j.buildenv.2023.110855.

[13] Y. Gao, J. Ji, Q. Wang, R. Jin, Y. Lin, Z. Shang, Y. Cao, S. Shen, C. Xu, and F. Gao, "Adaptive tracking and perching for quadrotor in dynamic scenarios," *IEEE Trans. Robot.*, vol. 40, pp. 499–519, 2024, doi: 10.1109/tro.2023.3335670.

[14] W. Wei, J. Wang, Z. Fang, J. Chen, Y. Ren, and Y. Dong, "3U: Joint design of UAV-USV-UUV networks for cooperative target hunting," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 4085–4090, Mar. 2023, doi: 10.1109/TVT.2022.3220856.

[15] J. M. Robinson, P. A. Harrison, S. Mavoa, and M. F. Breed, "Existing and emerging uses of drones in restoration ecology," *Methods Ecology Evol.*, vol. 13, no. 9, pp. 1899–1911, Sep. 2022, doi: 10.1111/2041-210x.13912.

[16] M. Javaid, A. Haleem, I. H. Khan, R. P. Singh, R. Suman, and S. Mohan, "Significant features and applications of drones for healthcare: An overview," *J. Ind. Integr. Manage.*, Sep. 2022, doi: 10.1142/S2424862222500245.

[17] S. A. H. Mohsan, M. A. Khan, F. Noor, I. Ullah, and M. H. Alsharif, "Towards the unmanned aerial vehicles (UAVs): A comprehensive review," *Drones*, vol. 6, no. 6, p. 147, Jun. 2022, doi: 10.3390/drones6060147.

[18] N. Chen, F. Kong, W. Xu, Y. Cai, H. Li, D. He, Y. Qin, and F. Zhang, "A self-rotating, single-actuated UAV with extended sensor field of view for autonomous navigation," *Sci. Robot.*, vol. 8, no. 76, Mar. 2023, Art. no. eade4538, doi: 10.1126/scirobotics.ade4538.

[19] T. Le, "Real-time object detection and tracking on drones," Portland State Univ., Maseeh College Eng. Comput. Sci., p. 5, Jan. 2018, [Online]. Available: http://archives.pdx.edu/ds/psu/25187

[20] E. Serafinelli and L. A. O'Hangan, "Drone views: A multimodal ethnographic perspective," *Vis. Commun.*, vol. 23, no. 2, p. 93, 2022, doi: 10.1177/14703572211065093.

[21] A. Georgiou, P. Masters, S. Johnson, and L. Feetham, "UAV-assisted real-time evidence detection in outdoor crime scene investigations," *J. Forensic Sci.*, vol. 67, no. 3, pp. 1221–1232, May 2022, doi: 10.1111/1556-4029.15009.

[22] A. Salim, M. Jabberi, T. M. Hamdani, and A. M. Alimi, "Exploring the potential of high-resolution drone imagery for improved 3D human avatar reconstruction: A comparative study with mobile images," in *Proc. Pacific-Rim Symp. Image Video Technol. (PSIVT)*, Nov. 2023, pp. 167–181.

[23] M. Garg, A. Sehrawat, and P. Savaridassan., "Vehicle lane detection for accident prevention and smart autodrive using OpenCV," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2023, pp. 1–5, doi: 10.1109/ICCCI56745.2023.10128394.

[24] D. Zhang, F. Xu, C.-M. Pun, Y. Yang, R. Lan, L. Wang, Y. Li, and H. Gao, "Virtual reality aided high-quality 3D reconstruction by remote drones," *ACM Trans. Internet Technol.*, vol. 22, no. 1, pp. 1–20, Feb. 2022, doi: 10.1145/3458930.

[25] S. Mishra, V. Verma, N. Akhtar, S. Chaturvedi, and Y. Perwej, "An intelligent motion detection using OpenCV," *Int. J. Sci. Res. Sci. Eng. Technol.*, vol. 9, no. 2, pp. 51–63, 2022, doi: 10.32628/ijsrset22925.

[26] S. Giri, G. Singh, B. Kumar, M. Singh, D. Vashisht, S. Sharma, and P. Jain, "Emotion detection with facial feature recognition using CNN & OpenCV," in *Proc. 2nd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, Apr. 2022, pp. 230–232, doi: 10.1109/ICACITE53722.2022.9823786.

[27] J. Xu and Y. Lu, "OpenWeedGUI: An open-source graphical user interface for weed imaging and detection," in *Proc. SPIE*, 2023, p. 16, doi: 10.1117/12.2664131.

[28] K. Li, J.-W. Bian, R. Castle, P. H. S. Torr, and V. Adrian Prisacariu, "MobileBrick: Building LEGO for 3D reconstruction on mobile devices," 2023, *arXiv:2303.01932*.

[29] X. Zhou, S. Liu, G. Pavlakos, V. Kumar, and K. Daniilidis, "Human motion capture using a drone," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 2027–2033, doi: 10.1109/ICRA.2018.8462830.

[30] S. Saito, Z. Huang, R. Natsume, S. Morishima, H. Li, and A. Kanazawa, "PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2304–2314.

[31] T. Alldieck, G. Pons-Moll, C. Theobalt, and M. Magnor, "Tex2Shape: Detailed full human body geometry from a single image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2293–2303, doi: 10.1109/ICCV.2019.00238.

[32] S. Saito, T. Simon, J. Saragih, and H. Joo, "PIFuHD: Multi-level pixel-aligned implicit function for high-resolution 3D human digitization," 2020, *arXiv:2004.00452*.

[33] T. He, J. P. Collomosse, H. Jin, and Stefano Soatto, "Geo-PIFu: Geometry and pixel aligned implicit functions for single-view human reconstruction," *Arxiv*, vol. 33, pp. 9276–9287, Jun. 2020.

[34] Z. Zheng, T. Yu, Y. Liu, and Q. Dai, "PaMIR: Parametric model-conditioned implicit representation for image-based human reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3170–3184, Jun. 2022, doi: 10.1109/tpami.2021.3050505.

[35] A. Chen, X. Wang, K. Shi, S. Zhu, B. Fang, Y. Chen, J. Chen, Y. Huo, and Q. Ye, "ImmFusion: Robust mmWave-RGB fusion for 3D human body reconstruction in all weather conditions," 2022, *arXiv:2210.01346*.

[36] M. Kocabas, N. Athanasiou, and M. J. Black, "VIBE: Video inference for human body pose and shape estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 5253–5263, doi: 10.1109/cvpr42600.2020.00530.

[37] H. Zhang, Y. Tian, Y. Zhang, M. Li, L. An, Z. Sun, and Y. Liu, "PyMAF-X: Towards well-aligned full-body model regression from monocular images," 2022, *arXiv:2207.06400*.

[38] H. Yi et al., "Human-aware object placement for visual environment reconstruction," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 3959–3970, doi: 10.1109/cvpr52688.2022.00393.

[39] E. Gärtner, M. Andriluka, E. Coumans, and C. Sminchisescu, "Differentiable dynamics for articulated 3D human motion reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, doi: 10.1109/cvpr52688.2022.01284.

[40] R. Shao, Z. Zheng, H. Zhang, J. Sun, and Y. Liu, "DiffuStereo: High quality human reconstruction via diffusion-based stereo using sparse cameras," 2022, *arXiv:2207.08000*.

[41] D. Hölscher, C. Reich, F. Gut, M. Knahl, and N. Clarke, "Pix2Pix hyperparameter optimisation prediction," *Proc. Comput. Sci.*, vol. 225, pp. 1009–1018, 2023, doi: 10.1016/j.procs.2023.10.088.

[42] A. Chahi, M. Kas, I. Kajo, and Y. Ruichek, "MFGAN: Towards a generic multi-kernel filter based adversarial generator for image restoration," *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 3, pp. 1113–1136, Mar. 2024, doi: 10.1007/s13042-023-01959-7.

[43] Z. Guo, M. Shao, and S. Li, "Image-to-image translation using an offset-based multi-scale codes GAN encoder," *Vis. Comput.*, vol. 40, no. 2, pp. 699–715, Feb. 2024, doi: 10.1007/s00371-023-02810-4.

[44] A. Ganjdanesh, S. Gao, H. Alipanah, and H. Huang, "Compressing image-to-image translation GANs using local density structures on their learned manifold," in *Proc. AAAI Conf. Artif. Intell.*, Mar. 2024, vol. 38, no. 11, pp. 12118–12126.

[45] M. S. K. Raghavendra and P. N. Sarappadi, "Transfer learning with Pix2Pix GAN for generating realistic photographs from viewed sketch arts," *J. Southwest Jiaotong Univ.*, vol. 57, no. 4, pp. 197–207, Aug. 2022, doi: 10.35741/issn.0258-2724.57.4.17.

[46] H.-A. Li, M. Zhang, Z. Yu, Z. Li, and N. Li, "An improved pix2pix model based on Gabor filter for robust color image rendering," *Math. Biosciences Eng.*, vol. 19, no. 1, pp. 86–101, 2022, doi: 10.3934/mbe.2022004.

[47] S. Choi and Y. Kim, "Rad-cGAN v1.0: Radar-based precipitation nowcasting model with conditional generative adversarial networks for multiple dam domains," *Geoscientific Model Develop.*, vol. 15, no. 15, pp. 5967–5985, Aug. 2022, doi: 10.5194/gmd-15-5967-2022.

[48] Y. Zhong, Y. Pei, P. Li, Y. Guo, G. Ma, M. Liu, W. Bai, W. Wu, and H. Zha, "Face denoising and 3D reconstruction from a single depth image," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 117–124, doi: 10.1109/FG47880.2020.00005.

[49] S. C. Medin, "MOST-GAN: 3D morphable StyleGAN for disentangled face image manipulation," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 1962–1971, doi: 10.1609/aaai.v36i2.20091.

[50] C. Zhang, S. Pujades, M. Black, and G. Pons-Moll, "Detailed, accurate, human shape estimation from clothed 3D scan sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5484–5493.

[51] A. S. Rasheed, R. H. Finjan, A. A. Hashim, and M. M. Al-Saeedi, "3D face creation via 2D images within blender virtual environment," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 21, no. 1, p. 457, Jan. 2021, doi: 10.11591/ijeecs.v21.i1.pp457-464.

[52] S.-K. Hung, "Image data augmentation from small training datasets using generative adversarial networks (GANs)," M.S. thesis, Faculty Sci. Health, Comput. Sci. Electron. Eng., Univ. Essex, Jan. 2023. [Online]. Available: http://repository.essex.ac.uk/id/eprint/35731

[53] Y. Ge, L. Sun, and D. Wang, "A novel image reconstruction algorithm based on texture aware multiscale GAN for veneer defects," *J. Intell. Fuzzy Syst.*, vol. 45, no. 6, pp. 9753–9769, Dec. 2023, doi: 10.3233/jifs-231692.

[54] C. Han, T. Ma, J. Huyan, Z. Tong, H. Yang, and Y. Yang, "Multi-stage generative adversarial networks for generating pavement crack images," *Eng. Appl. Artif. Intell.*, vol. 131, May 2024, Art. no. 107767, doi: 10.1016/j.engappai.2023.107767.

[55] D. T. Zaidan, A. S. Rasheed, and Z. H. Kareem, "Robust FPV drone vision by artificial intelligence technology," *Al-Kut Univ. Coll. J.*, vols. 780–2616, pp. 208–214, 2023.

[56] G. Qian, "Magic123: One image to high-quality 3D object generation using both 2D and 3D diffusion priors," 2023, *arXiv:2306.17843*.

[57] S. Jiang, H. Jiang, Z. Wang, H. Luo, W. Chen, and L. Xu, "HumanGen: Generating human radiance fields with explicit priors," in *Proc. CVPR*, 2023, pp. 12543–12554, doi: 10.1109/cvpr52729.2023.01207.

[58] K. Zhou, Z. Wang, Y.-Q. Ni, Y. Zhang, and J. Tang, "Unmanned aerial vehicle-based computer vision for structural vibration measurement and condition assessment: A concise survey," *J. Infrastructure Intell. Resilience*, vol. 2, no. 2, Jun. 2023, Art. no. 100031, doi: 10.1016/j.iintel.2023.100031.

[59] M. H. Chamas, S. Amine, E. Gazo Hanna, and O. Mokhiamar, "Control of a novel parallel mechanism for the stabilization of unmanned aerial vehicles," *Appl. Sci.*, vol. 13, no. 15, p. 8740, Jul. 2023, doi: 10.3390/app13158740.

[60] Q. Ma and X. Huang, "Research on recognizing required items based on OpenCV and machine learning," in *Proc. SHS Web Conf.*, vol. 140, 2022, Art. no. 01016.

[61] K. Singh, N. V. Krishna, A. Shinde, and S. Y. Sawant, "Surface area calculation of asymmetric/axisymmetric shapes utilising simple image processing and OpenCV," in *Proc. 7th Int. Conf. Comput., Commun., Control Autom. (ICCUBEA)*, Aug. 2023, pp. 1–8, doi: 10.1109/iccubea58933.2023.10392058.

[62] L. L. Doskolovich, A. I. Kashapov, E. A. Bezus, N. V. Golovastikov, and D. A. Bykov, "Optical computation of the Laplace operator at oblique incidence using a multilayer metal-dielectric structure," *Opt. Exp.*, vol. 31, no. 10, p. 17050, 2023, doi: 10.1364/oe.489750.

[63] R. Cai, J. Tung, Q. Wang, H. Averbuch-Elor, B. Hariharan, and N. Snavely, "Doppelgangers: Learning to disambiguate images of similar structures," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 34–44, doi: 10.1109/iccv51070.2023.00010.

[64] F. B. Mofrad and G. Valizadeh, "DenseNet-based transfer learning for LV shape classification: Introducing a novel information fusion and data augmentation using statistical shape/color modeling," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119261, doi: 10.1016/j.eswa.2022.119261.

[65] G. Paulin and M. Ivasic-Kos, "Review and analysis of synthetic dataset generation methods and techniques for application in computer vision," *Artif. Intell. Rev.*, vol. 56, no. 9, pp. 9221–9265, Sep. 2023, doi: 10.1007/s10462-022-10358-3.

[66] M. Z. Yousif, L. Yu, S. Hoyas, R. Vinuesa, and H. Lim, "A deep-learning approach for reconstructing 3D turbulent flows from 2D observation data," *Sci. Rep.*, vol. 13, no. 1, p. 2529, Feb. 2023, doi: 10.1038/s41598-023-29525-9.

[67] O. Selnes, T. Bjørsum-Meyer, A. Histace, G. Baatrup, and A. Koulaouzidis, "Annotation tools in gastrointestinal polyp annotation," *Diagnostics*, vol. 12, no. 10, p. 2324, Sep. 2022.

[68] S. P. Porkodi, V. Sarada, V. Maik, and K. Gurushankar, "Generic image application using GANs (generative adversarial networks): A review," *Evolving Syst.*, vol. 14, no. 5, pp. 903–917, Oct. 2023, doi: 10.1007/s12530-022-09464-y.

[69] J. Kim, H. Yang, and K. Min, "24-GAN: Portrait generation with composite attributes," *Mathematics*, vol. 10, no. 20, p. 3887, Oct. 2022.

[70] L. E. Christovam, M. H. Shimabukuro, M. D. L. B. T. Galo, and E. Honkavaara, "Pix2pix conditional generative adversarial network with MLP loss function for cloud removal in a cropland time series," *Remote Sens.*, vol. 14, no. 1, p. 144, Dec. 2021.

[71] R. K. Senapati, R. Satvika, A. Anmandla, G. A. Reddy, and C. A. Kumar, "Image-to-image translation using Pix2Pix GAN and cycle GAN," in *Proc. Int. Conf. Data Intell. Cogn. Inform.*, 2024, pp. 573–586, doi: 10.1007/978-981-99-7962-2_42.

[72] D.-P. Nguyen, T.-N. Nguyen, S. Dakpé, M.-C. H. B. Tho, and T.-T. Dao, "Fast 3D face re-construction from a single image using different deep learning approaches for facial palsy patients," *Bioengineering*, vol. 9, no. 11, p. 619, 2022.

[73] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-Moll, "Video based reconstruction of 3D people models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8387–8397.

[74] M. Kocabas, N. Athanasiou, and M. J. Black, "VIBE: Video inference for human body pose and shape estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5252–5262, doi: 10.1109/CVPR42600.2020.00530.

[75] H. Zhang et al., "PyMAF-X: Towards well-aligned full-body model regression from monocular images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 10, pp. 1–16, Jan. 2023, doi: 10.1109/tpami.2023.3271691.

[76] S. Zhang et al., "EgoBody: Human body shape and motion of interacting people from head-mounted devices," in *Proc. Eur. Conf. Comput. Vis.* (Lecture Notes in Computer Science), vol. 13666. Cham, Switzerland: Springer, Nov. 2022, pp. 180–200, doi: 10.1007/978-3-031-20068-7_11.

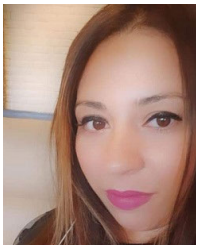
[77] Y. Xiu, J. Yang, D. Tzionas, and M. J. Black, "ICON: Implicit clothed humans obtained from normals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13286–13296, doi: 10.1109/CVPR52688.2022.01294.

[78] G. Sharma et al., "MvDeCor: Multi-view dense correspondence learning for fine-grained 3D segmentation," in *Proc. Eur. Conf. Comput. Vis.*, vol. 13662. Cham, Switzerland: Springer, Nov. 2022, pp. 550–567, doi: 10.1007/978-3-031-20086-1_32.

[79] S. Peng, Y. Zhang, Y. Xu, Q. Wang, Q. Shuai, H. Bao, and X. Zhou, "Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9050–9059, doi: 10.1109/CVPR46437.2021.00894.

[80] F. Lin, Y. Yue, S. Hou, X. Yu, Y. Xu, K. D. Yamada, and Z. Zhang, "Hyperbolic chamfer distance for point cloud completion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 14549–14560, doi: 10.1109/iccv51070.2023.01342.
[81] X. Yang, Y. Luo, Y. Xiu, W. Wang, H. Xu, and Z. Fan, "D-IF: Uncertainty-aware human digitization via implicit distribution field," 2023, *arXiv:2308.08857*.
[82] X. Li, X. Li, and J. Deng, "Disentangled representation transformer network for 3D face reconstruction and robust dense alignment," *Vis. Comput., Int. J. Comput. Graph.*, pp. 1–18, Dec. 2023, doi: 10.1007/s00371-023-03202-4.

ALI SALIM RASHEED was born in Baghdad, Iraq, in 1982. He received the master's degree in computer engineering artificial intelligence and robotics from the Faculty of Computer Engineering, College of Engineering, Ferdowsi University of Mashhad, in 2018. Currently, he is pursuing the Ph.D. degree in computer systems engineering with the National Engineering School of Sfax, University of Sfax, Tunisia. He is carrying out more research studies in the REsearch Groups in Intelligent Machines. His research interests include applying intelligent methods (deep neural networks, generative adversarial networks, 3D avatar reconstruction, 3D archaeology objects reconstruction, deep learning, and texture rendering) to deep machine learning, computer vision, intelligence drone applications, and virtual reality.



MARWA JABBERI (Member, IEEE) was born in Bizerte, Tunisia, in 1990. She received the master's degree in automatic reasoning systems from the Faculty of Sciences of Monastir, Tunisia, in 2014, and the Ph.D. degree in computer science from the Higher Institute of Computer Science and Communication Technologies of Sousse, in 2023. She is currently carrying out more research studies in the REsearch Groups in Intelligent Machines (ReGIM-Lab), ENIS. Her research interests include applying intelligent methods (deep neural networks, feature extraction, and recognition algorithms) to deep machine learning, computer vision, face image processing, intelligent pattern recognition, and analysis of large-scale complex systems.

TAREK M. HAMDANI (Senior Member, IEEE) was born in Tunis, Tunisia, in 1979. He received the M.S. and Ph.D. degrees in computer science engineering from the University of Sfax, Sfax, Tunisia, in 2003 and 2011, respectively. He is currently pursuing the Ph.D. degree in computer science engineering with the National Engineering School of Sfax, Sfax. Since 2003, he has been a Teacher with the University of Sfax and the University of Monastir, Monastir, Tunisia. Since 2012, he has been a Teacher of computer science with Taibah University, Medina, Saudi Arabia. His current research interests include intelligent pattern recognition, learning, analysis of large-scale complex systems, applications of intelligent methods, neural networks, fuzzy logic, and genetic algorithms to pattern recognition. He is a member of IEEE CIS and SMC Societies. He is a Reviewer of *Pattern Recognition Letters* and *Soft Computing*.

ADEL M. ALIMI (Senior Member, IEEE) was born in Sfax, Tunisia, in 1966. He received the degree in electrical engineering, in 1990, and the Ph.D. and H.D.R. degrees in electrical and computer engineering, in 1995 and 2000, respectively. He is currently a Professor of electrical and computer engineering with the University of Sfax. His research interests include the applications of intelligent methods (neural networks, fuzzy logic, and evolutionary algorithms) to pattern recognition, robotic systems, vision systems, and industrial processes. His research focuses on intelligent pattern recognition, learning, analysis, and intelligent control of large scale complex systems. He was the IEEE Sfax Subsection Chair, in 2011, the IEEE ENIS Student Branch Counselor, in 2011, the IEEE Systems, Man, and Cybernetics Society Tunisia Chapter Chair, in 2011, and the IEEE Computer Society Tunisia Chapter Chair, in 2011. He was the General Chairperson of the International Conference on Machine Intelligence ACIDCA-ICMI-2005 and ACIDCA-ICMI-2000. He is the Founder and the Chair of many IEEE Chapter in Tunisia Section. He is an Expert Evaluator of European Agency for Research. He was a Guest Editor of several special issues of international journals, such as *Fuzzy Sets and Systems*, *Soft Computing*, *Journal of Decision Systems*, *Integrated Computer-Aided Engineering*, and *Systems Analysis Modelling Simulation*. He is an Associate Editor and a member of the editorial board of many international scientific journals, such as IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Neurocomputing*, *Neural Processing Letters*, *International Journal of Image and Graphics*, *Neural Computing and Applications*, *International Journal of Robotics and Automation*, and *International Journal of Systems Science*.