3D Human Pose reconstruction Single-pixel imaging

Carlos A. Osorio Quero*, Daniel Durini, Jose Rangel-Magdaleno, Jose Martinez-Carranza and Ruben Ramos-Garcia
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), 72840, Mexico.

ABSTRACT

Three-dimensional (3D) human pose estimation is a fundamental task in computer vision and has important applications in fields such as gaming, sports analysis, and surveillance. Due to their high mobility and ability to capture images from different angles, the use of drones has rapidly increased in various applications in recent years. However, 3D human pose estimation from drone images remains a challenging task due to the complexity of the problem and the limited resolution of the images captured by the drone. Single pixel imaging (SPI) is a novel imaging technique that has recently gained attention in the field of computer vision due to its ability to capture high-resolution images with low-cost hardware. In SPI, a Hadamard pattern is projected onto the object of interest, and a single pixel detector measures the light that is reflected back from the object. In this paper, we propose a novel approach for 3D human pose estimation from drone images using SPI. The proposed method consists of three main steps: 1) SPI-based image acquisition, 2) feature extraction using deep learning, and 3) 3D pose estimation using a regression-based approach. To evaluate the proposed method, we collected a dataset of drone images of human subjects in different poses. The experimental results demonstrate that our approach achieves state-of-the-art performance in 3D human pose estimation from drone images. Compared to existing methods, our approach is computationally efficient and requires low-cost hardware.

1 INTRODUCTION

Three-dimensional (3D) pose estimation has become an essential tool in many fields, including robotics [1], augmented reality [2], unmanned Aerial Vehicles (UAVs)[3], and virtual reality[4], to name a few. It involves reconstructing an accurate 3D model of a human body from a single 2D image, which is challenging due to the inherent ill-posed nature of the problem [5]. However, advancements in computational techniques and algorithms have paved the way for exciting applications in various fields [6]. In this context, 3D pose estimation has been used in drone technology to improve surveillance [7], emergency response [8], animal conservation [9], and infrastructure inspection [10].

Reconstructing an accurate human shape from imperfect input data, accounting for non-rigid deformations and joint articulations, is a challenging task. However, recent advances in deep learning techniques have made it possible to achieve end-to-end reconstruction of human shape [11]. The Skinned Multi-Person Linear model (SMPL-X) [12] offers a compact representation for 3D human shape, and has been integrated with deep neural networks for 3D human reconstruction from RGB images. This integration involves using deep neural networks to extract powerful image features, followed by direct regression of SMPL-X shape and pose parameters [13]

Different technologies, such as RGB cameras [13], thermal cameras [14], and IR-UWB RADAR [15], exist for estimating human pose. However, traditional cameras have limitations in low-light conditions, making them less useful for nighttime surveillance applications. In contrast, single-pixel camera (SPC) systems offer a promising solution to these limitations [16]. By exploiting the power of deep learning techniques, SPCs can reconstruct high-quality images from sparse measurements, making them an ideal candidate for detecting 3D human pose in nighttime surveillance applications. One key advantage of SPCs over traditional cameras is their ability to capture images in the near-infrared (NIR) spectrum, which provides better visibility in low-light conditions. Combining SPC technology with time-of-flight (TOF) sensing [17], we can obtain 2D/3D images of the environment, providing additional information about the location and movement of objects. The use of SPCs in surveillance applications is not limited to nighttime environments. They can also be used to capture images in harsh environments where traditional cameras may fail [18].

In this work, the authors propose a single-pixel imaging (SPI) vision system with active illumination in the NIR wavelength range of 850-1500 nm, which can be employed using single InGaAs photodetectors. As a detection strategy, they use a U^2-net [19] to remove the background of the SPI image and identify the object for segmentation of the area of interest containing the element to detect. They then use Vision Transformers to perform silhouette analysis-based gait recognition for human identification [20]. The information is used to generate a 3D model through the VIBE method [21], which predicts SMPL-X body model parameters using a convolutional neural network pretrained on the AMASS dataset [22], for single-image body pose and shape estimation. This approach can improve the detection and surveillance capabili-

*Email address(es): caoq@inaoep.mx

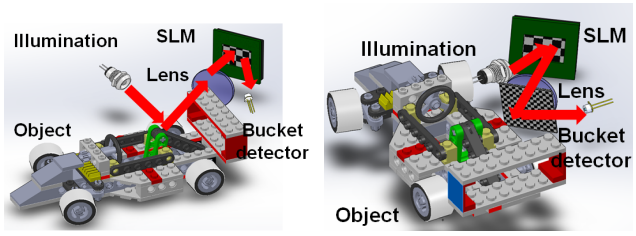
ties of drones, especially in low-light and harsh environments.

Therefore, in this work, we propose the following:

- Investigating the potential of Single-Pixel Imaging in generating 3D human pose from low-resolution 2D images.
- This work focuses on a novel and difficult objective of forecasting the 3D hand pose using a single 2D binary mask acquired through NIR-SPI imaging.

2 SINGLE-PIXEL IMAGE RECONSTRUCTION

The technique known as single-pixel imaging (SPI) [17] reconstructs images by measuring correlated intensity on a detector without spatial resolution. To achieve this, SPI cameras use spatial light modulators (SLMs) like Digital Micro-Mirror Devices (DMD) to create structured light patterns using Hadamard pattern for scene interrogation [17]. There are two architectures in which SPI cameras operate, namely structured detection and structured illumination, as depicted in Figure 1.



(a) (b)

Figure 1: Two different approaches applied to SPI: (a) structured detection, and (b) structured illumination [17].

In structured detection, the object is illuminated with light, and the reflected light is projected onto an SLM, followed by detection using a bucket detector. On the other hand, in structured illumination, the light source is modulated by the SLM Φi and illuminates the object $O(x, y)$, and the reflected light is detected by a bucket detector and converted into an electrical signal Si by Eq.(1) [17].

$$Si = \alpha \sum x = 1^{M} \sum_{y=1}^{N} O\left(x, y\right) \Phi_i \left(x, y\right) \qquad (1)$$

Here, α is a constant factor that depends on the optoelectronic response of the photodetector.

As the light's spatial pattern and the reflected light from the object correlate to an electrical signal, projecting a sequence of spatial patterns produces a sequence of electrical signals that can be used for computational image reconstruction. Therefore, the image $I(x, y)$ is reconstructed from the captured signal S_i and the corresponding pattern Φ_i using Eq.(2) [17].

$$I\left(x, y\right) = \alpha \sum_{x=1}^{M} \sum_{y=1}^{N} Si\Phi i\left(x, y\right) \qquad (2)$$

To generate Hadamard-like patterns Φ_i using active illumination, this work uses an array of 32 x 32 NIR-LEDs that emit radiation with a peak wavelength of 1550 nm. This wavelength is chosen due to reduced scattering by water and reduced water absorption coefficients. The NIR-LED array is placed perpendicular to the lens's focal length to project the light pattern to infinity. However, given the array's size, the patterns are projected up to a distance of 0.3-3 meters. Although the active illumination approach does not fully illuminate the object, the technique of Fast Super Resolution CNN (FSRCNN) can reconstruct high-quality images [23]. The active illumination approach offers several advantages, including operating in different outdoor weather conditions, low-level illumination scenarios, and being less

2.1 Single-pixel camera (SPC)

This work introduces the concept of using structured illumination to enhance the quality of images captured in challenging lighting conditions, such as strong backlight and stray light. To achieve this, the structured illumination is provided by an array of 32 x 32 NIR-LEDs with a peak wavelength of 1550 nm, and a time-of-flight (TOF) system wavelength of 850 nm. The illumination is detected by an InGaAs photodiode (SPD). This active illumination approach has numerous advantages, including its ability to function under various outdoor weather conditions, low-level illumination, and reduced sensitivity to background radiation noise.

The proposed architecture for this Near-Infrared Single-pixel imaging (NIR-SPI) system is composed of two main parts (See Fig. 5). The first part includes the essential components for generating images, which are an InGaAs photodetector, an array of NIR-LEDs, a TOF system, and an ADC. The second part is responsible for processing the electrical output signal from the SPD module. It accomplishes this by digitizing the signal using the ADC and then using a Graphics Processing Unit (GPU) to process the data. The GPU unit used for this system is the Jetson Xavier NX, which generates Hadamard patterns and processes data from the ADC. It then runs the OMP-GPU Algorithm to generate 2D images [18].

3 HUMAN MODELING

Parametric human models like SMPL-X [12] offer a succinct representation of human shapes by using shape and pose parameters to encode variations [24]. The SMPL-X model offers several benefits, such as the disentanglement of human shape and pose, which allows for independent analysis and control of each [21]-[25]. It also avoids modeling rugged and twisted shapes directly, which can be problematic for neural network-based methods [26]-[27], by using a skinning process to model deformation. Moreover, SMPL-X is differentiable and can be easily integrated with neural networks [25].

Therefore, we have chosen to use SMPL-X as the underlying representation for modeling 3D humans in our research.

The SMPL-X model includes shape parameters β, pose parameters $\theta \in R^{3K}$, and global translation parameters. Body pose is defined by a skeleton rig with K = 24 joints, including the body root,as shown in Fig. 2 are used for shape blending and encode global shape information, while pose parameters are used for pose blending and skinning and encode local information between adjacent joints, except for the root joint's pose parameters, which denote the global rotation of the entire shape. Note that SMPL-X's pose parameters denote the relative rotation from a joint to its parent, which differs from 2D or 3D human pose estimation [28], where the pose refers to joint locations. With β and θ, we can obtain the 3D body mesh $M = f_{SMPL}(\beta, \theta)$, where $M \in R^{Nx3}$ is a triangulated surface with N=6890. The 3D SMPL-X model locations of the body joints X can be predicted with the body mesh using a pre-trained mapping matrix $W \in R^{KxN}$, such that $X \in R^{kx3} = WM$ [29]. To project the body joints from 3D to 2D, we use the perspective camera model. Assuming the camera parameters are $\delta \in R$

4 PROPOSED METHOD

The figure 6 illustrates the steps involved in obtaining a 3D human model using near-infrared single-pixel imaging (NIR-SPI). The process involves multiple computer vision techniques to reconstruct the 3D pose of a human from a single low-resolution image. Each step is explained in detail below:

- Capture a low-resolution image of the human using a single pixel (see Fig. 2). The image is then adjusted for contrast to extract the basic shape of the person without revealing any details. The background of the image is removed using the U^2-net [19] method, a deep learning model that can accurately segment the foreground and background of an image. The resulting image is a silhouette of the person, which only shows the outline without any details of the surface or texture.

- Use the identified poses to regress the SMPL-X human pose (pose θ, shape β, and camera s, R, T) using the Video and Image-based human Body pose Estimation (VIBE) method, a deep learning model that can estimate the 3D pose of a human from a single image or video.

- Create a 3D reconstruction of the human pose using a tool such as SMPL-X, which can fit the estimated pose to a 3D body model (see Fig. 2).

By following these steps, a 3D human model can be obtained from a single low-resolution image using NIR-SPI technology and computer vision techniques.

5 EXPERIMENTAL RESULTS

The results of the experiment demonstrate that the proposed methods can be used to obtain a 3D human model from NIR-SPI imaging at a distance of 1 meter from the SPC camera, and under night-time illumination conditions, after calibrating the SPC camera based on [16]. The process involves multiple steps that utilize various deep learning models, including U^2-net and VIBE, to extract and estimate the 3D pose of a human from a single low-resolution image. Figure 2 depicts the experimental setup and results. To assess the efficacy of the proposed method, the researchers performed experiments on several datasets, including Silhouette-based 3D Human Pose Estimation [20], and the Human Pose SMPL-X dataset, such as the AMASS dataset [22].

5.1 Discussion: Proposed method

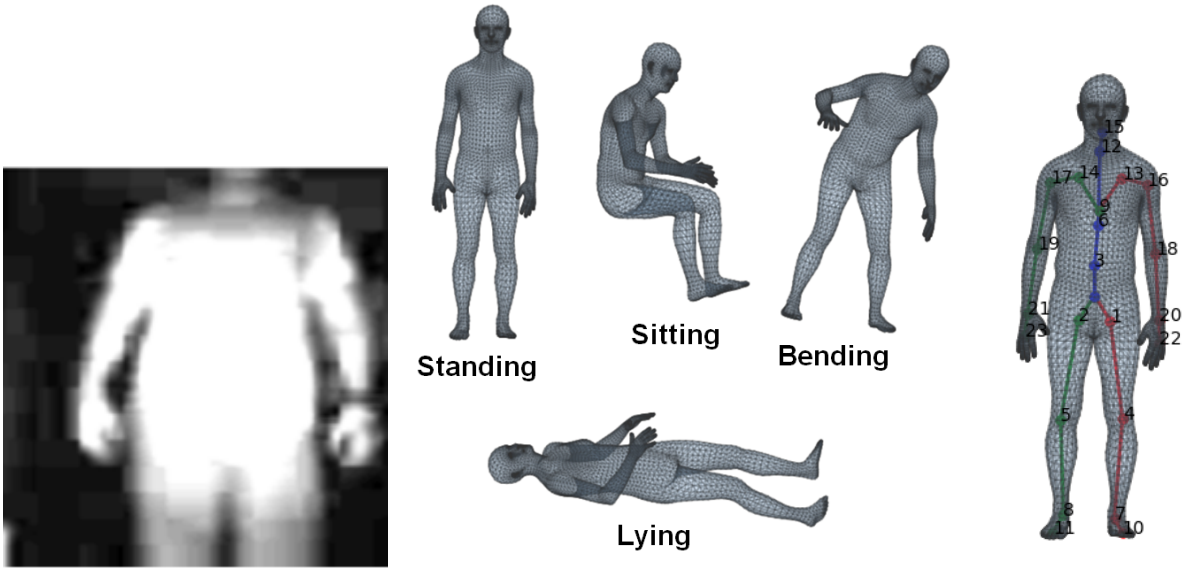
To evaluate the proposed network architecture depicted in Figure 6 (Appendix A), we tested various approaches for reconstructing human positions at night-time from a distance of 1 meter using the SMPL-X model. Our tests took into account the limited field of view of the SPC camera, which spans 74º x 57º. We captured NIR-SPI images of human poses, including sitting, standing, bending, and lying. During our analysis, we observed limitations in hand and body positions relative to the reference image, particularly in the bending position due to the loss of information in the input NIR-SPI image caused by reflection effects and low resolution in the reconstructed NIR-SPI image. Nevertheless, our results showed that for the standing and sitting positions (as depicted in Figure 3), the 3D human reconstruction demonstrated superior accuracy in vertex and joint positions (as presented in Table 1).

Table 1: **The results are Mean vertex to-vertex (V2V) (in mm) and Mean Per-Joint Position Error (MPJPE) (in mm) [30] body for the different human position (Lying, Bending, Sitting , and Standing)**

Human Pose	V2V error ↓	MPJPE error ↓
Lying	57.29	53.2
Bending	49.86	40.19
Sitting	34.2	33.7
Standing	42	41

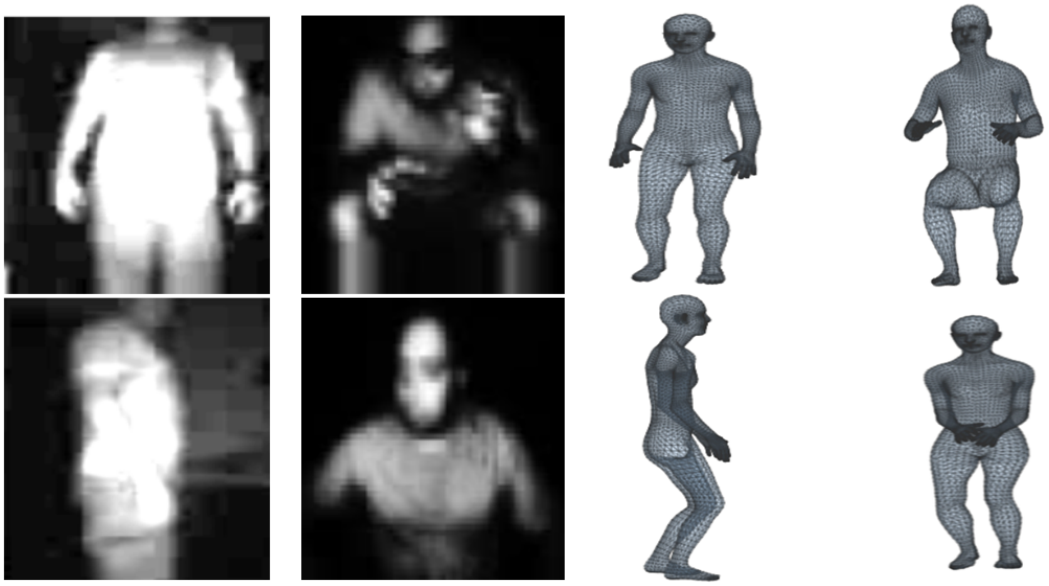
6 NIR-SPI 3D HUMAN POSE RECONSTRUCTION APPLIED TO UAVS AUTONOMOUS APPLICATION RESCUES

The NIR SPI 3D Human Pose reconstruction technology has the potential to revolutionize autonomous rescues carried out by unmanned aerial vehicles (UAVs) (see Fig. 4). Using Near-Infrared Spectroscopy Imaging (NIR SPI), the technology captures the reflective properties of the human body's surface in 3D. This technology can be especially useful in identifying the exact location of individuals who may be



(a) (b) (c)

Figure 2: human poses but with same joint positions generate from NIR-SPI imaging: (a) Test NIR-SPI imaging, (b) SMPL-X model generate base on estimation pose (Standing, Sitting, Bending ,and Lying) , (c) SMPL-X model generate with joints



(a) (b)

Figure 3: Capture human poses imaging at distance of 1 m: (a) Capture NIR-SPI imaging human pose standing, sitting, and bending, and (b) 3D human pose regression based on SMPL-X model.

trapped or injured in hazardous or hard-to-reach areas during UAV rescues. The 3D reconstruction of the human pose can provide invaluable information to rescue teams on the ground, allowing them to plan and execute a more effective rescue operation.

7 CONCLUSION

The study aimed to obtain a 3D model of the human body using NIR-SPI imaging for various poses, including standing, sitting, bending, and lying, with accuracy demonstrated in the V2V and MPJPE error table. Although the approach had limitations in hand positioning due to the low contrast of

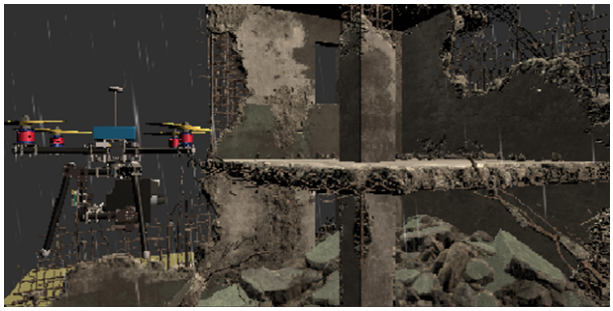


Figure 4: Schematic view to illustrate UAV (drone) environment hard-to-reach areas

the NIR-SPI image, the accurate estimation of the person's 3D pose through qualitative and quantitative evaluations of the level position of the core person detection was shown. These findings demonstrate the potential of the proposed approach for 3D human modeling from a single low-resolution image.

In contrast, the SMPL-X model presented in this study captures the body, face, and hands simultaneously and fits the model to a single NIR-SPI image and 2D joint detections. The study demonstrated that the SMPL-X model can capture bodies, hands, and faces from NIR-SPI images. However, the model had higher errors for bending and lying poses, indicating limitations in the pose parameters θ. Therefore, the study recommends implementing a compensation model in future applications. The authors also suggest future work involving the development of a dataset of SMPL-X fits in real-world scenarios and the direct regression of SMPL-X parameters from NIR-SPI images. Overall, this study is a crucial step towards capturing expressive body, hand, and face movements from an NIR-SPI image, with potential applications in the fields of intelligent automation, unmanned aerial vehicles, and autonomous vehicles.

ACKNOWLEDGEMENTS

The first author is thankful to Consejo Nacional de Ciencia y Tecnología (CONACYT) for his scholarship with No. CVU: 661331.

REFERENCES

[1] Christian Zimmermann, Tim Welschehold, Christian Dornhege, Wolfram Burgard, and Thomas Brox. 3d human pose estimation in rgbd images for robotic task learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1986–1992, 2018.

[2] P. Sudhaman, P. Surendiren, Sherif.S Meeran, and P.C. Kishore Raja. Augmented reality in automation using virtual 3d models. In *2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12)*, pages 1–4, 2012.

[3] Nitin Saini, Elia Bonetto, Eric Price, Aamir Ahmad, and Michael J. Black. Airpose: Multi-view fusion network for aerial 3d human pose and shape estimation. *IEEE Robotics and Automation Letters*, 7:4805–4812, 2022.

[4] Debangana Ram, Bholan3D VRath Roy, and Vaibhav Soni. A review on virtual reality for 3d virtual trial room. In *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*, pages 247–251, 2022.

[5] Yueh-Ling Lin and Mao-Jiun J. Wang. Constructing 3d human model from 2d images. In *2010 IEEE 17Th International Conference on Industrial Engineering and Engineering Management*, pages 1902–1906, 2010.

[6] Wenming Meng, Tao Hu, and Shuai Li. 3d human pose estimation with adversarial learning. In *2019 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 93–99, 2019.

[7] Yap Wooi Hen and Raveendran Paramesran. Single camera 3d human pose estimation: A review of current techniques. In *2009 International Conference for Technical Postgraduates (TECHPOS)*, pages 1–8, 2009.

[8] Savvas Papaioannou, Panayiotis Kolios, Theocharis Theocharides, Christos G. Panayiotou, and Marios M. Polycarpou. Towards automated 3d search planning for emergency response missions. *Journal of Intelligent & Robotic Systems*, 103(1):2, Aug 2021.

[9] C.A. Johnson, J. Seidel, R.E. Carson, W.R. Gandler, A. Sofer, M.V. Green, and M.E. Daube-Witherspoon. Evaluation of 3d reconstruction algorithms for a small animal pet camera. In *1996 IEEE Nuclear Science Symposium. Conference Record*, volume 3, pages 1481–1485 vol.3, 1996.

[10] Christos Papachristos, Kostas Alexis, Luis Rodolfo Garcia Carrillo, and Anthony Tzes. Distributed infrastructure inspection path planning for aerial robotics subject to time constraints. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 406–412, 2016.

[11] Margarita Khokhlova, Cyrille Migniot, and Albert Dipanda. 3d visual-based human motion descriptors: A review. In *2016 12th International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pages 564–572, 2016.

[12] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[13] Dongyue Chen, Yuanyuan Song, Fangzheng Liang, Teng Ma, Xiaoming Zhu, and Tong Jia. 3d human body reconstruction based on smpl model. *The Visual Computer*, Apr 2022.

[14] Henry M. Clever, Zackory Erickson, Ariel Kapusta, Greg Turk, C. Karen Liu, and Charles C. Kemp. Bodies at rest: 3d human pose and shape estimation from a pressure image using synthetic data. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6214–6223, 2020.

[15] Gon Woo Kim, Sang Won Lee, Ha Young Son, and Kae Won Choi. A study on 3d human pose estimation using through-wall ir-uwb radar and transformer. *IEEE Access*, 11:15082–15095, 2023.

[16] Carlos Alexander Osorio Quero, Daniel Durini, Jose de Jesus Rangel-Magdaleno, Jose Martinez-Carranza, and Ruben Ramos-Garcia. 2d nir-spi spatial resolution evaluation under scattering condition. In *2022 19th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*, pages 1–6, 2022.

[17] Carlos A. Osorio Quero, Daniel Durini, Jose Rangel-Magdaleno, and Jose Martinez-Carranza. Single-pixel imaging: An overview of different methods to be used for 3d space reconstruction in harsh environments. *Review of Scientific Instruments*, 92(11):111501, 2021.

[18] C. Osorio Quero, D. Durini, J. Rangel-Magdaleno, J. Martinez-Carranza, and R. Ramos-Garcia. Single-pixel near-infrared 3d image reconstruction in outdoor conditions. *Micromachines*, 13(5), 2022.

[19] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognition*, 106:107404, 2020.

[20] Ryosuke Hori. Silhouette-based 3d human pose estimation using a single wrist-mounted 360° camera, 2022.

[21] Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. Vibe: Video inference for human body pose and shape estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[22] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *International Conference on Computer Vision*, pages 5442–5451, October 2019.

[23] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 391–407, Cham, 2016. Springer International Publishing.

[24] Chun-Hao P. Huang, Hongwei Yi, Markus Höschle, Matvey Safroshkin, Tsvetelina Alexiadis, Senya Polikovsky, Daniel Scharstein, and Michael J. Black. Capturing and inferring dense full-body human-scene contact. In *Proceedings IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 13274–13285, June 2022.

[25] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.

[26] Or Litany, Alex Bronstein, Michael Bronstein, and Ameesh Makadia. Deformable shape completion with graph convolutional autoencoders. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1886–1895, 2018.

[27] Gül Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. BodyNet: Volumetric inference of 3D human body shapes. In *ECCV*, 2018.

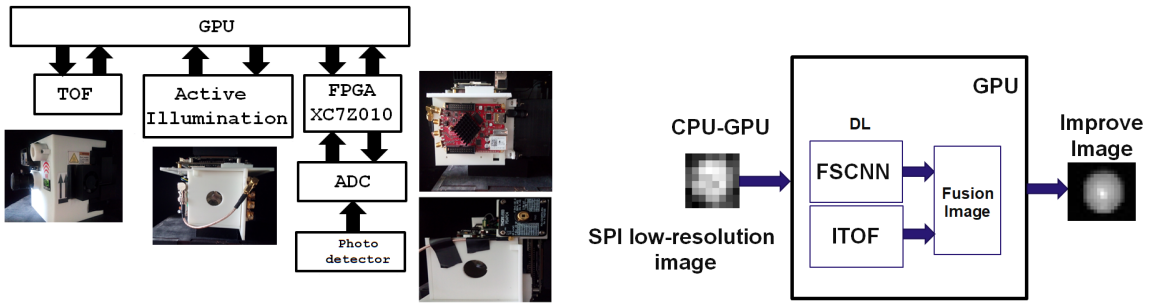
[28] Helge Rhodin, Mathieu Salzmann, and Pascal Fua. Unsupervised geometry-aware representation for 3d human pose estimation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 765–782, Cham, 2018. Springer International Publishing.

[29] Xiangyu Xu, Hao Chen, Francesc Moreno-Noguer, László A. Jeni, and Fernando De la Torre. 3d human shape and pose from a single low-resolution image with self-supervised learning. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, pages 284–300, Cham, 2020. Springer International Publishing.

[30] Gyeongsik Moon, Juyong Chang, and Kyoung Mu Lee. V2v-posenet: Voxel-to-voxel prediction network for accurate 3d hand and human pose estimation from a single depth map. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

APPENDIX A: NIR-SPI SYSTEM VISION AND MODEL PROPOSED

NIR-SPI architecture comprises modulate in the figure 5 show an overall block diagram. The model proposed based in the feature extraction the NIR-SPI image (background remove and generation silhoutte pose), the 3D pose estimation using a regression-based approach using the information the pose (shape parameters β and pose parameters θ) see Figure 6.



(a) (b)

Figure 5: Overall block diagram: (a) Proposed vision system dimension 11 x 11 x 14 cm, focal length 20 cm, weight 1.2 kg, power consumption 45 W, first stage module photodiode, active illumination source, photodetector diode InGaAs FGA015, TOF system, second stage GPU unit and ADC, (b) diagram of the processing algorithm used by the proposed NIR-SPI vision system from low-resolution SPI image applying FSCNN network and fusion image with information captured way TOF.

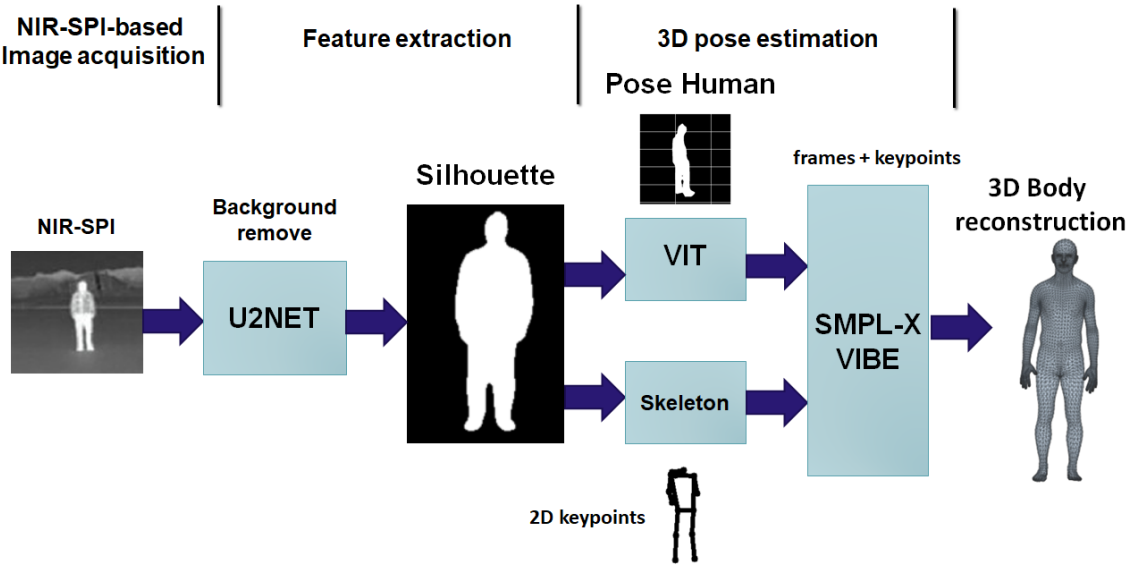


Figure 6: Overview of the proposed network architecture, which takes NIR single-pixel imaging input and outputs 3D body reconstruction based on SMPL-X shape and pose parameters. The entire network consists of three main modules: (i) NIR-SPI-based image acquisition, (ii) Feature extraction using deep learning : To extract the background, the NIR-SPI image is used to obtain the silhouette, (iii) 3D pose estimation using a regression-based approach: The silhouette image is used to obtain the GAIT features (shape estimation), which are then used to pose the human using Vision Transformers (VIT) and skeleton joint features. These features are used to pre-define the pose SMPL-X model. From the pre-defined parameters (pose θ, shape β and camera s, R, T), the SMPL-X model is fed to the off-the-shelf SMPL-X model to obtain the reconstructed 3D human mesh.