LPSNet: End-to-End Human Pose and Shape Estimation with Lensless Imaging

Haoyang Ge[1,†], QiaoFeng[1,†], Hailong Jia[1], Xiongzheng Li[1], Xiangjun Yin[1],
You Zhou[2], Jingyu Yang[1], Kun Li[1,*]
[1]Tianjin University, China [2]Nanjing University, China
{ghy0623,fengqiao,jhl,lxz,yinxiangjun,yjy,lik}@tju.edu.cn zhouyou@nju.edu.cn

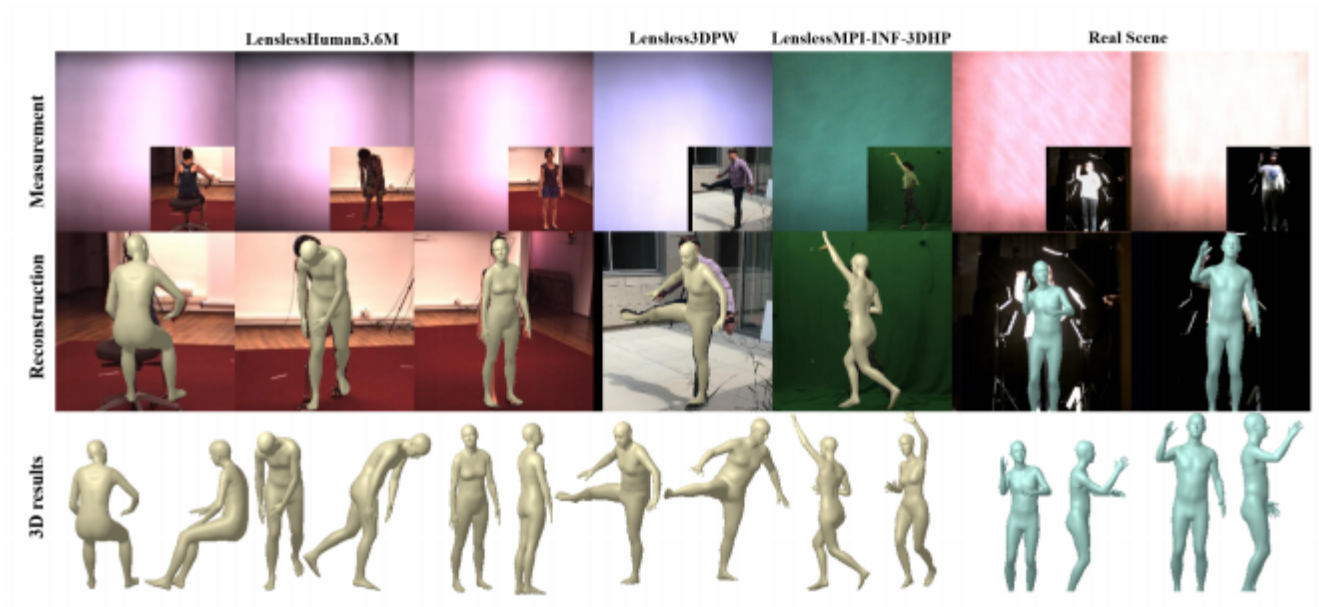

Figure 1. **LPSNet: End-to-End Human Pose and Shape Estimation with Lensless Imaging.** We contribute a framework for estimating human poses and shapes from individual lensless measurements. The first row shows the input measurements acquired by our lensless imaging system, the second row shows the estimated human poses and shapes from lensless measurements, and the bottom row shows the 3D results shown in different views.

Abstract

Human pose and shape (HPS) estimation with lensless imaging is not only beneficial to privacy protection but also can be used in covert surveillance scenarios due to the small size and simple structure of this device. However, this task presents significant challenges due to the inherent ambiguity of the captured measurements and lacks effective methods for directly estimating human pose and shape from lensless data. In this paper, we propose the first end-to-end framework to recover 3D human poses and shapes from lensless measurements to our knowledge. We specifically design a multi-scale lensless feature decoder to decode the lensless measurements through the optically encoded mask

† Equal contribution.

* Corresponding author.

for efficient feature extraction. We also propose a double- head auxiliary supervision mechanism to improve the esti- mation accuracy of human limb ends. Besides, we estab- lisha lensless imaging system and verify the effectiveness of our method on various datasets acquired by our lens- less imaging system. The code and dataset are available at https://cic.tju.edu.cn/faculty/likun/projects/LPSNet.

1. Introduction

In recent years, lensless imaging technologies [6, 7, 21, 27] have advanced significantly due to their many ad- vantages, such as privacy protection, smaller size, sim- ple structure, and lower cost. 3D human pose and shape (HPS) estimation [11, 16, 36, 37] requires miniaturized and

lightweight imaging system as application scenarios become more diverse. All these advantages, especially privacy, make the lensless imaging system very suitable as an imaging device for human pose and shape estimation. In this work, we propose LPSNet, which aims to estimate 3D human pose and shape from lensless measurements instead of RGB images, achieving cheaper and privacy-protecting 3D human pose and shape estimation.

A thin, lightweight, and potentially cost-effective optical encoder is used in a lensless imaging system instead of traditional cameras with lenses, while others are expensive, rigid, and occupy more space. At this stage, the application of a lensless imaging system is extensive, it is mainly used in the field of microscopic imaging, RGB image reconstruction, and so on. More valuable information can be obtained from lensless measurements due to the special optical encoding method of lensless imaging systems. Directly estimating human pose and shape from lensless measurements is not currently possible. The first step is to reconstruct an RGB image from a lensless measurement and then estimate the human pose and shape from the RGB images. However, experimental findings indicate that the reconstructed RGB images are of suboptimal quality, resulting in incomplete local features and significant deviations in the position of the human body. Combining these factors leads to inaccurate human pose estimation when using lensless measurements to reconstruct RGB images. This approach has computational burden and computational resources, making it very unfavorable for deployment at the endpoint.

In this paper, We aim to advance human pose and shape estimation using a lensless imaging system, which needs to overcome two main challenges. First, how to extract features from lensless measurements for human pose and shape estimation. Secondly, during early experiments, when using features extracted from lensless measurements to estimate human pose and shape, we found poor estimation accuracy of human limbs.

To address these challenges, we introduce LPSNet, the first end-to-end human pose and shape estimation framework with lensless imaging. To extract features from optically encoded lensless measurements, we propose a multi-scale lensless feature decoder (MSFDecoder). Specifically, we introduce a global perception layer to enhance the global decoding capability of MSFDecoder. The global information that has been optically encoded to the global can be efficiently decoded to obtain a feature map that can be used in subsequent processes. To improve pose and shape estimation, we propose a Double-Head Auxiliary Supervision(DHAS) mechanism to be implemented during training. Auxiliary supervision can improve the estimation accuracy of human limbs and correct results with large deviations.

Our main contributions can be summarized as follows:

1) We propose LPSNet, an end-to-end pose and shape esti-

mation network for lensless imaging systems. This is the first work to estimate human poses and shapes directly from lensless measurements.

2) We propose MSFDecoder, a Multi-Scale Lensless Feature Decoder that decodes and extracts features from lensless measurements, which can be receptive to global features in lensless measurements.

3) We propose a Double-Head Auxiliary Supervision mechanism for both pose and shape estimation, which can improve the estimation accuracy of human limbs.

2. Related Work

2.1. Lensless Imaging System

A conventional photographic camera typically comprises a focusing lens, which may consist of one or more optical elements and an image sensor positioned at or near the focal length of the lens. The lens in such a camera directs the projected light from the observed scene onto the sensor, aiming to accurately map specific scene points to individual pixels on the sensor. Conversely, in a lensless imaging system, the absence of a lens defines its configuration. Instead, an optical modulator, such as a coded amplitude mask or a diffuser, is positioned between the scene and the image sensor, often close to the sensor itself. As a result, the recorded data deviates significantly from the expected RGB image during imaging. In this process, the local information of the object transforms overlapping global information through the optically encoded mask.

Within the mask-modulated lensless systems, a fixed optical mask is introduced to create a versatile lensless system that can work for a wide range of object distances and lighting scenarios, whether passive or uncontrolled. The mask modulates the incoming light and generates a measurement that can be decoded through computational methodologies. Mask-modulated lensless imaging systems were used to perform 2D imaging [4, 7, 21], refocusing [7], 3D imaging [2, 7], and microscopic imaging [1, 2, 7].

Generally speaking, amplitude modulators and phase modulators [1, 4] are the two types of optical masks used in lensless imaging systems. Phase modulators can be further sub-categorized into phase gratings [32, 33], diffusers [2], and phase masks [7]. One key characteristic of a mask-modulated lensless system is the pattern the mask produces on the sensor for a point light source in the scene. We call this pattern the Point-Spread Function(PSF), and its properties determine the imaging model of the system. As shown in Fig. 2, we design a simple mask-modulated lensless sys- tem for data acquisition in this experiment. The lensless imaging system designed in our experiments chose a diffuser as the mask.

The main advantages of lensless imaging are as follows: lensless is small in size and can be assembled in a variety of