IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Human Pose Estimation Using Wireless Signals

Mr. P. Rahate, K. Mane, K. Gadge

Assistant Professor, Student, Student
Department of Computer Engineering,
Datta Meghe College of Engineering, Navi-Mumbai, India

Abstract: In this project, we use Deep Learning to teach wireless devices to sense people's postures and movements even from other side of the wall or any other form of occlusions. While normal humans can't see through walls, this Deep Learning system is trained using images of humans in certain poses and corresponding reflected radio frequencies from their bodies to tell what someone is doing. The system counts on the fact that Wifi Signals are not stopped by solid walls but get reflected off a person's body. By using this combination of visual data and Deep Learning to see through walls, we can enable better scene understanding and smarter environments to live safer and more productive lives. We have used the CSI tool based on Raspberry pi to extract and process Wifi signals.

Keywords **– WiFi Pose Estimation, CSI, FCN, Deep learning, pi.**

I. INTRODUCTION

The most beautiful thing about being engineers is our ability to transform science fiction into reality and today, we are living at a unique time in history where we have the computational and technological tools to achieve this. For decades, we have read and seen science fiction stories about supe powers like seeing the invisible. For example, we have fantasized about seeing through walls, similar to batsman X-ray vision and over the past few years, research on this particular topic is transforming the fantasy of seeing through walls into reality.

Human pose estimation is a significant research topic in computer vision and artificial intelligence [1-4]. Recently, there have been noteworthy advances and significant improvements in human pose estimation. The traditional 2-D human pose estimation methods are mainly based on the Pictorial Structure (PS) model, exploiting the Kinematic Tree to characterize the distribution of human joints. Compared with traditional methods, methods based on deep learning train the networks with large quantity of images, learn features from global space and has good robustness. According to the network structures, these methods are classified as follows:

A. Single CNN methods

Compared with traditional methods designing features in a handcraft way, CNN is better at feature learning. The single CNN methods extract features by exploiting CNN [1-2] and model the configuration of joints using PS model.

B. Multi-stage CNN methods

After training, each layer of a Multi-stage CNN [2-3] network becomes relatively independent and is given a clear meaning in the structure and function. These methods solve the convergence difficulty of a large-scale network, thus gaining better performances compared to single CNN methods

C. Multi-branch CNN methods.

The multibranch CNN methods combine the processing results of CNN branches to obtain rich representations of image information, so as to improve the performance.

D. GAN methods

GAN is based on unsupervised learning and completes the training through massive "confrontation" between the generator and the discriminator. It captures high-order correlations of data in the absence of target class label information. Compared with regression models, GAN fits data faster and more adequately, and generates better samples.

The above methods are some the used methods for pose estimation. For our system we are going to use a fully convolutional Neural (FCN) network [10] Teacher-Student model.

II. BACKGROUND

The basic **Wifi Pose Estimation (WPE)** system consists of the pre-processing system, feature extraction and classifier block as shown in figure. The raw CSI quantities in radio signal after pre-processing stage are given to the feature extraction block. The extracted features are then sent to the classifier to classification section. Finally, the Pose is estimated after classification.

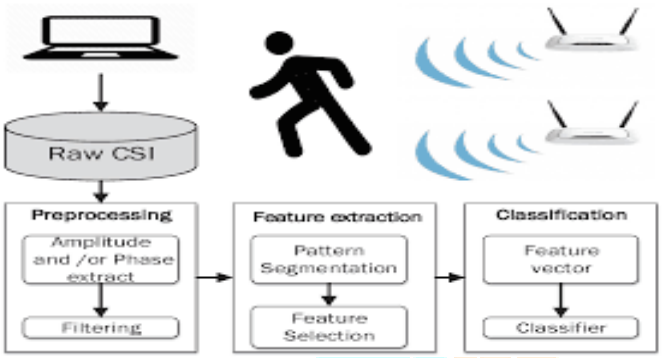

Figure 2.1: Wifi Pose Estimation System

In wireless communications **channel state information** (**CSI**) refers to known channel properties of a communication link. This information describes how a signal propagates from the transmitter to the receiver and represents the combined effect of, for example, scattering, fading, and power decay with distance [5]. The CSI signal is pre-processed before giving it to the feature extraction module so as to improve the efficiency and accuracy of the feature extraction process. The pre- processing stages are Amplitude/Pulse extraction and filtering. Filtering is the process used to reduce the noise in a CSI signal that occurs due to the disturbances in the environment or during the recording of the CSI sample. We also do pattern segmentation and feature selection in filtering stage. These pattern provides the basis for classification.

A. *Collection of Data*

We recruited 8 volunteers and asked them to do casual daily actions in two rooms of the campus, one laboratory room and one classroom. For each volunteer, data of his first 80% recording is used to train the networks, and data of the last 20% recording is used to test the networks.

B. *Database Description*

AlphaPose

AlphaPose is an open-source multi-person pose estimation repository, which is also applicable for single person pose estimation. AlphaPose is a two-step framework, which first detects person bounding boxes by a person detector (YOLOv3) then estimates pose for each detected box by the pose regressor. With the innovative regional multi-person pose estimation framework (RMPE), AlphaPose gains estimation resilience to the inaccurate person detection, which largely facilitates the pose estimation performance. Please refer [6] for more details on AlphaPose and RMPE.

When applied to single person pose estimation, AlphaPose generates n three-element predictions in the format of (xi , yi ; ci), where n is the number of keypoints[6-7] to be estimated, xi and yi are the coordinates of the i-th keypoint, and ci is the confidence of the above coordinates. In this paper, we use the COCO person keypoint setting and n is 18.

COCO Dataset

COCO is a large-scale object detection, segmentation, and captioning dataset. COCO has several features:

- Object segmentation
- Recognition in context
- Superpixel stuff segmentation
- 330K images (>200K labeled)
- 1.5 million object instances
- 80 object categories
- 91 stuff categories
- 5 captions per image
- 250,000 people with keypoints

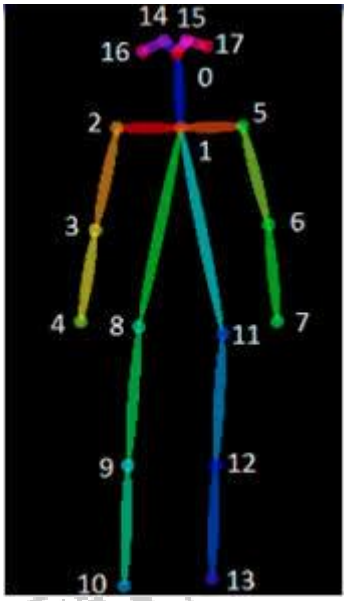

Figure 2.1: Skeletal keypoint orderinng

C. *Channel state information Extraction*

CSI was introduced to capture fine-grained variations in the wireless channel. For CSI extraction we use nexmon-csi extractor [8]. Nexmon extarctor allows you to extract channel state information (CSI) of OFDM-modulated Wi-Fi frames (802.11a/(g)/n/ac) on a per frame basis with up to 80 MHz bandwidth on the Broadcom Wi-Fi Chips. We set WiFi working within a 20MHz band, the CSI of 30 carriers can be obtained through a open-source tool [11] like nexmon.

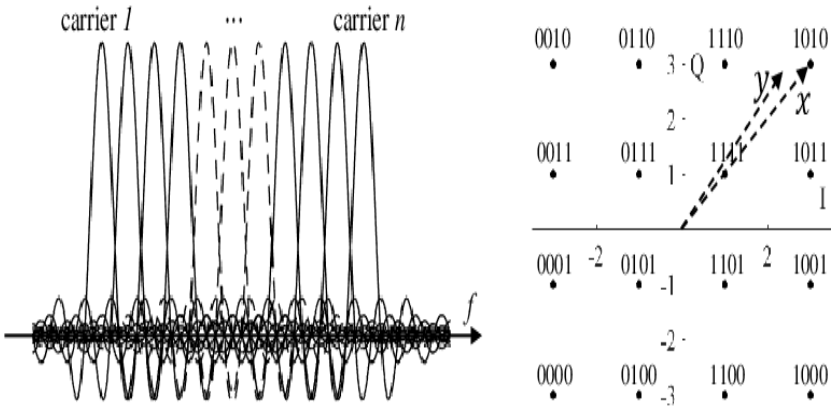

Figure 2.1: Orthogonal Frequency Division Multiplexing (OFDM)

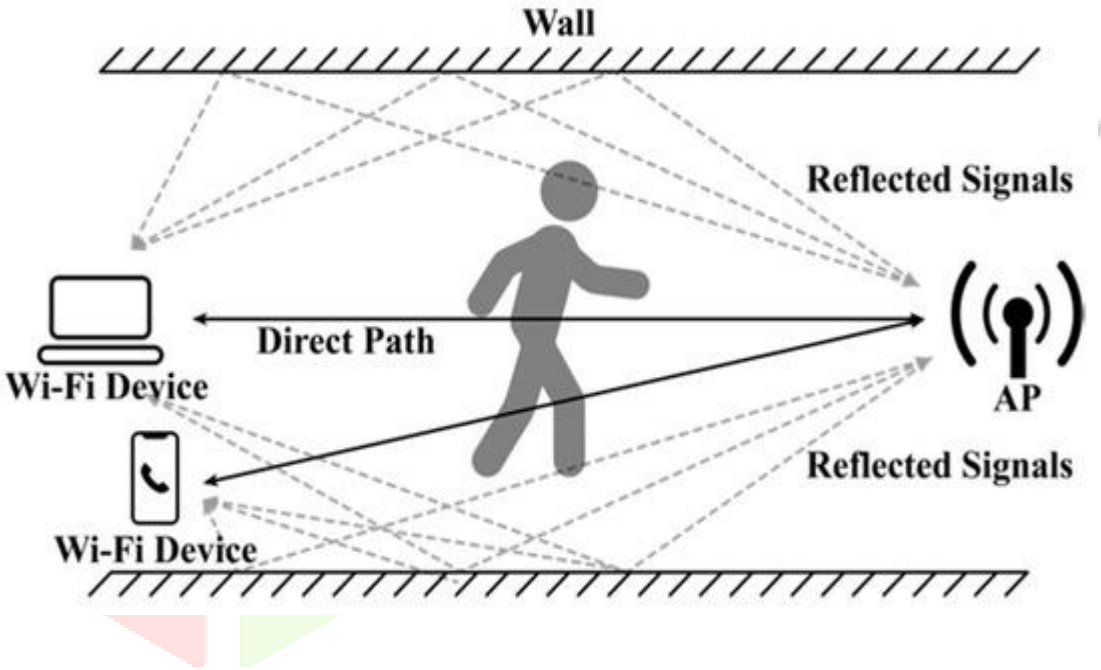

Figure 2.1: Channel State Information Extraction

III. PROPOSED POSE ESTIMATION SYSTEM

The proposed system includes the use of Wifi Signals for detecting human pose. Advantage is that the pose can be detected even in cases of occlusion or darkness. Moreover, our system does not use expensive radar system. Rather, the system is purely based on the presence of Wifi, which are present in the surroundings.

The WiFi CSI recording system is comprised with 2 ends, one 3-antenna sender and one 3-antenna receiver. The sender broadcasts WiFi signals, meanwhile the receiver parses CSI through [9] when receiving the broadcasting WiFi. In our setting, the parsed CSI is a tensor with the size of $n \times 30 \times 3 \times 3$, where the n is for the number of received WiFi packages; 30 is for the subcarrier number; the last two 3s represent the antenna numbers of sender and receiver, respectively. In our data acquisition, we set the sampling rate of WiFi devices and the camera as 100Hz and 20Hz, respectively. Thus we have a paired dataset in which every 5 CSI samples and one image frame are synchronized by their time-stamps.

After sensing the Wifi signals after extracting the Channel State Information, the signals are further processed and trained with a Deep Learning Model, in order to detect the human pose. The main problem with training model was that Wifi signals cannot be directly annotated by humans, as is done normally for most Deep Learning models. Therefore, we use a teacher student model for annotating the Wifi Signals, which is illustrated in the figure 3.2.

In this model, AlphaPose with a camera works as teacher Network and there is a student network with a router working alongside it. In order to create the training model, camera and router are synchronised with each other.

The camera captures the images and the router extracts the csi information at the same time. Since they are both synchronised, this system allots proper annotations to each pattern of wifi signals extracted.

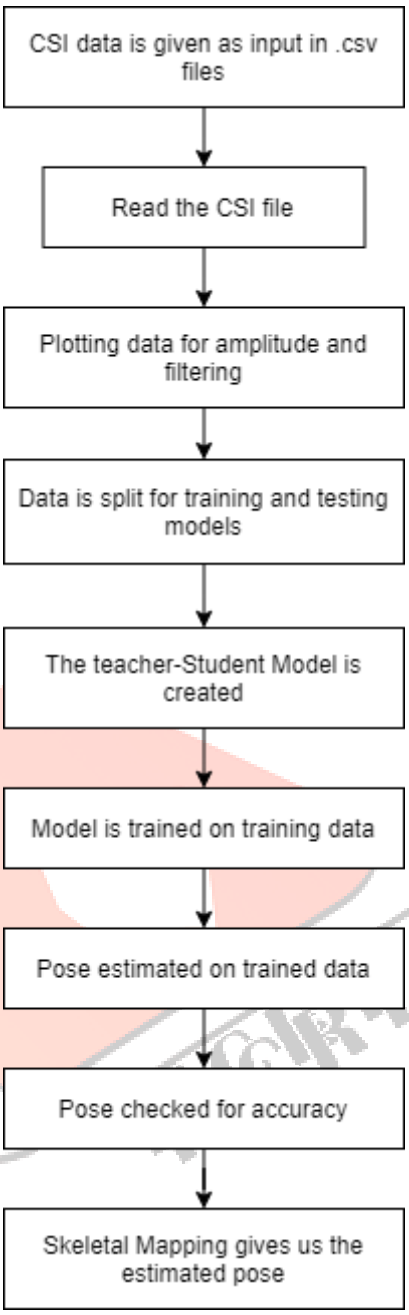

Figure 3.1: Module for WiFi Pose Estimation

BENEFITS OF THE PROPOSED SYSTEM

- It overcomes the limitations of the previous systems.
- Since the previous systems to detect human pose were based only on cameras, they failed when there were instances of detecting some scenarios (for example, fraudulent activities) at night or in darkness.
- Furthermore, the existing systems that see through walls with radio frequency waves are based on radar systems that are very expensive to install.
- The system proposed in this project uses the wifi signals which are readily available all around us.
- Thus, cost required to set up the system is very negligible

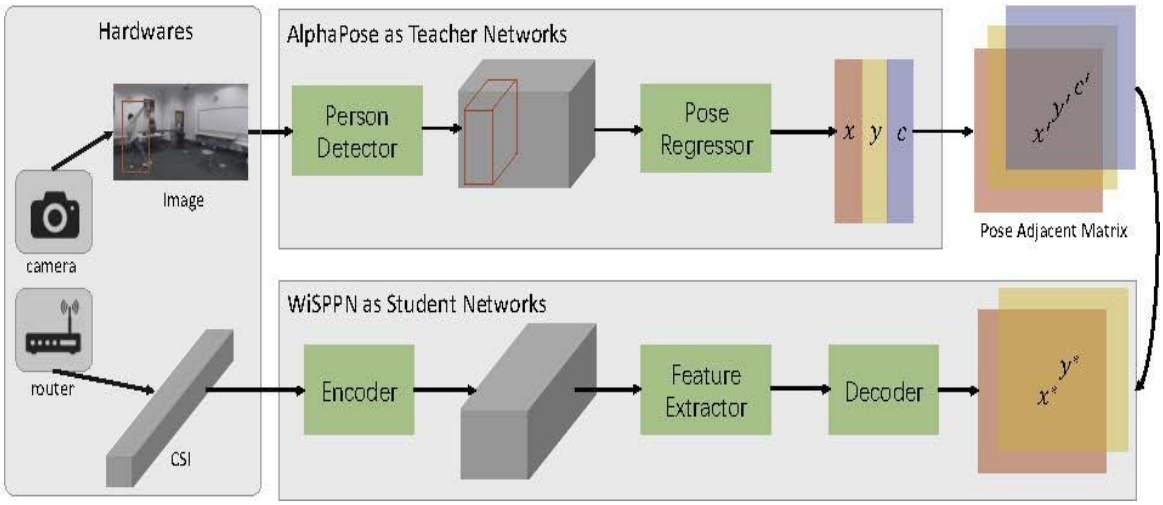

Figure 3.2: Module for WiFi Pose Estimation

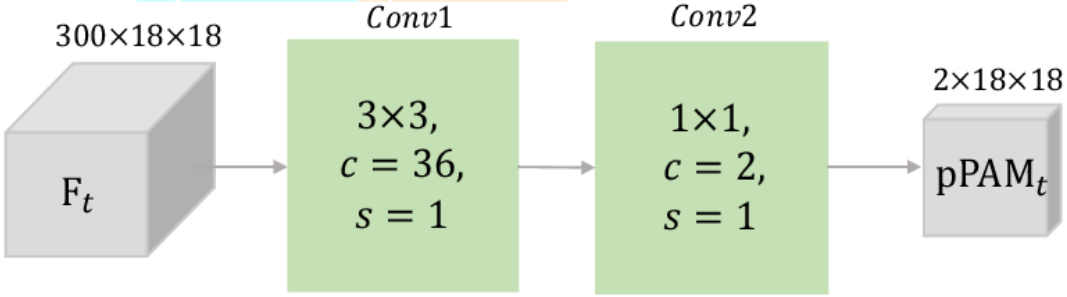

Figure 3.3: Pose Adjacent Matrix of proposed system

Input.	$C_t \in \mathbf{R}^{150\times144\times144}$	
Block name	Output size	Parameters
Block 1	150x144x144	$\begin{bmatrix} 3\times3, c=150, s=1 \\ 3\times3, c=150, s=1 \\ 3\times3, c=150, s=1 \\ 3\times3, c=150, s=1 \end{bmatrix}$
Block 2	150x72x72	$\begin{bmatrix} 3\times3, c=150, s=2 \\ 3\times3, c=150, s=1 \\ 3\times3, c=150, s=1 \\ 3\times3, c=150, s=1 \end{bmatrix}$
Block 3	300x36x36	$\begin{bmatrix} 3\times3, c=300, s=2 \\ 3\times3, c=300, s=1 \\ 3\times3, c=300, s=1 \\ 3\times3, c=300, s=1 \end{bmatrix}$
Block 4	300x18x18	$\begin{bmatrix} 3\times3, c=300, s=2 \\ 3\times3, c=300, s=1 \\ 3\times3, c=300, s=1 \\ 3\times3, c=300, s=1 \end{bmatrix}$
Output.	$F_t \in \mathbf{R}^{300\times18\times18}$	

Figure 3.4: Parameters of feature extractor

Table 1: Results in PCK. ‘R.’ and ‘L.’ are for right and left respectively.

Order	Keypoint	PCK@5	PCK@ 10	PCK@ 20
1	Head	0.0784	0.2222	0.5255
2	Torso	0.0458	0.1712	0.4523
3	R. Arm	0.0536	0.1673	0.4235
4	L. Arm	0.0431	0.1373	0.4131
5	R. Leg	0.0418	0.1373	0.3869
6	L. Leg	0.0353	0.1216	0.3856
Average		0.0496	0.1594	0.4311

The proposed system is carried out using COCO database. Both for the training and testing. The dataset sizes are 79496 and 19931 for training and testing respectively. The table shows the results for PCK (Percentage of Correct Keypoints) for 6 different keypoints. From this we can say that our system estimates well.

The result showed by system suggests that the system is a good pose estimation system.We can get the skeletal output of the person taking the pose with an accuracy of about 90%.

CONCLUSION

While normal humans can't see through walls, this Deep Learning system is trained using images of humans in certain poses and corresponding reflected radio frequencies from their bodies to tell what someone is doing. By using this combination of visual data and Deep Learning to see through walls, we can enable better scene understanding and smarter environments to live safer and more productive lives.

REFERENCES

ASME Standard Journal Paper,

[1] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single wifi access point," in 13th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 16), 2016, pp. 165–178.

[2] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in ACM SIGCOMM Computer Communication Review, vol. 45, no. 4. ACM, 2015, pp. 269–282.

[3] K. Qian, C. Wu, Y. Zhang, G. Zhang, Z. Yang, and Y. Liu, "Widar2. 0: Passive human tracking with a single wi-fi link," in Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services. ACM, 2018, pp. 350–361. 9

[4] X. Li, S. Li, D. Zhang, J. Xiong, Y. Wang, and H. Mei, "Dynamic-music: accurate device-free indoor localization," in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 2016, pp. 196–207.

[5] https://en.wikipedia.org/wiki/Channel_state_information

[6] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2334–2343.

[7] F. Wang, J. Han, S. Zhang, X. He, and D. Huang, "Csi-net: Unified human body characterization and action recognition," arXiv preprint arXiv:1810.03064, 2018.

[8] Matthias Schulz, Daniel Wegemer and Matthias Hollick. Nexmon: The C-based Firmware Patching Framework**.** https://nexmon.org

[9] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," ACM SIGCOMM Computer Communication Review, vol. 41, no. 1, pp. 53–53, 2011

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[11] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," ACM SIGCOMM Computer Communication Review, vol. 41, no. 1, pp. 53–53, 2011.

[12]Article reference and images: arXiv:1904.00277 [cs.CV]