# A General Survey of Hand Pose Estimation Projects

Morayo Ogunsina
Department of Computer Science.
California State University, Los Angeles, USA.
mogunsi@calstatela.edu

April 12, 2022

## Abstract

This paper reviews some of the existing and relatively new work done in the area of hand pose estimation and prediction, with a focus on sign language translation and hand emoji prediction, in relevance to the ongoing *Emoji-Hand* project. These previous works are centered around certain computer vision, and machine learning tasks consisting of but not limited to 3D hand mesh reconstruction and keypoint mapping, emoji shape matching, prediction of different hand poses, and hand shape recognition. First, hand pose estimation is introduced by highlighting some relevant works done on the topic, and in some cases, the use they have been developed for. The paper will also discuss the methods, approaches, challenges, and limitations of the research. Finally, the paper concludes with a recommendation for some of these methods and approaches for use in the *Emoji-Hand* project.

## I. Introduction

Hand pose estimation research focuses on representing a human hand, including the joint locations, orientation and articulation in space, as a mesh model or a feature outline. It also involves the simulation of hand parts in a 3D space. Hand pose estimation tasks such as 3D hand mesh reconstruction [2] and keypoint mapping, emoji matching and association [8], estimation of different hand poses, and 3D hand shape recognition [10]. Hand mesh reconstruction and hand pose estimation involve estimating and representing the 3D surface and location information of the hand joint and surface in a graph-like format and generating the 3D hand mesh in this graph from associated image features [2] [10]. For hand pose estimation and reconstruction, depth sensors are used to provide some representation of hand joint location for further processing, although more research has been geared towards the estimating hand poses and reconstructing 3D meshes from RGB cameras since they are widely available [2]. Keypoint mapping and estimation involves the determination of the kinematic structure of the joints a hand affords [10]. These motion parameters reflect the degrees of freedom (DoF) of a hand structure, allowing proper representation of the complex motion associated with a hand [10]. Hand emoji prediction involves hand shape recognition and mapping a standard emoji to its outline, as predicted by a deep neural network [4] [5] [8]. These research topics can be considered as important foundations for the development of robust and accessible applications for use in areas like sign language translation [3].

Modern hand pose estimation tasks are made possible through a singular or a combination of several computer vision (CV) and pattern recognition tasks, largely aided by Deep Learning and low cost computation and edge computing. It has become increasingly easy to generate predictive models that solve certain hand pose estimation problems and researchers leveraged this to optimize existing solutions and explore new topics in the field of hand pose estimation. In certain use or research cases, like sign language translation, gesture interaction, and even in virtual reality and hand emoji mapping, natural language processing (NLP) tasks have also been utilized to develop robust and accessible

solutions [6].

## II. Related Works

Classical methods and algorithms in machine learning and CV were some of the first approaches available to hand pose estimation problems and later, as computation became cheaper, faster and powerful, deep learning gained popularity with researchers who utilized them in CV, and NLP tasks relevant to estimation.

Wu *et al.* [10] explored articulated hand motion problems centered on high dimensionality introduced by the complex motion of the hand, and particle degeneracy which is a problem of the sampling process. In the paper, 3 (static constraints, joint correlation constraint, and *purposive constraints* ) constraints which prevents feasible hand articulation from spanning the entire joint angle space were highlighted. The approach was to use a sequential Monte Carlo tracking algorithm as part of a divide and conquer strategy to separate the hand articulations restructure them, thereby reducing the complexity caused by many DoFs[1]. the hand affords. An importance function was tied to the for sampling process for the Monte Carlo algorithm. and thus the representation of the hand articulation is more detailed, as smaller number of points are used for efficient tracking.

Liuhao *et al.*'s [2] approach was different for the same set of problem in hand pose estimation. They first represented the 3D hand mesh as a graph-structured data, taking inspiration from works on Graph CNN such as [9]. Using an RGB image as input, 3D hand mesh vertices are generated in graph form. This approach ensures better representation of high-varied 3D hand pose as well as capture more local details.

Zimmermann *et al.* [11] used a deep network for detecting keypoints from single RGB images, and in effect, generating 3D hand pose. The use of deep neural networks are useful in bypassing occlusion, multiple ambiguities and strong articulation problems. It was one of the first works done on estimating 3D hand images from RGB cameras and encouraged the use of single-depth cameras thereby providing a more robust use coverage, particularly for sign language projects. Zimmermann *et al.* [11] extended the capabilities of its hand pose estimation system with a classifier for gesture recognition, an important precursor to sign language recognition.

Another related work in hand pose estimation that reflects the versatile nature of deep neural networks is found in [1], in which Transfer learning, a variation of Deep Learning was used to classify hand poses. The learning model is trained on a hand washing pose dataset.

Successful applications of hand pose estimation research has made it possible to explore development of accessible technologies, for example in sign language translation and hand emoji association as seen in [6][7] [4] and [8]. These projects either utilize deep learning and to develop an hand emoji-based hand pose predictive system or optimize existing solutions. For example, [4] and [8] made use of CNNs[2] to predict and associate standard emojis to hand signs. In [6], an NLP library which returns recognized hand signs as words,and established the importance of pretraining unlabelled and the development of a cross-lingual transfer from one sign language to another. This is particularly significant due to the concentration of available research in one sign language (American Sign Language). This research opens up transferability of solutions.

## III. Recommendations and Conclusions

For the purpose of the *Emoji-Hand* project, some of the most suitable approaches that can guarantee significant success is the use of deep learning or some variations of deep learning such as Transfer Learning. Deep Learning and Computer Vision libraries have much to offer towards deploying rapid solutions for hand pose estimation tasks. It is also advisable to explore different approaches to dataset acquisition for the purpose of arriving at a robust solution.

In this paper, we highlighted some relevant works in the field of hand pose estimation as also explored certain works that have extended or optimized existing solutions for use in accessible technologies such as sign language recognition and translation systems.

[1] Degrees of Freedom

[2] Convolutional Neural Networks

## References


[1] Rashmi Bakshi. Hand pose classification based on neural networks, 2021.

[2] Liuhao Ge, Zhou Ren, Yuncheng Li, Zehao Xue, Yingying Wang, Jianfei Cai, and Junsong Yuan. 3d hand shape and pose estimation from a single RGB image. *CoRR*, abs/1903.00812, 2019.

[3] Susan Goldin-Meadow and Diane Brentari. Gesture, sign and language: The coming of age of sign language and gesture studies. *The Behavioral and brain sciences*, -1:1–82, 10 2015.

[4] Jung Koh, Josh Cherian, Paul Taele, and Tracy Hammond. Developing a hand gesture recognition system for mapping symbolic hand gestures to analogous emojis in computer-mediated communication. *ACM Transactions on Interactive Intelligent Systems*, 9:1–35, 03 2019.

[5] Jaya Sahoo, Allam Prakash, Paweł Pławiak, and Saunak Samantray. Real-time hand gesture recognition using fine-tuned convolutional neural network. *Sensors*, 22:706, 01 2022.

[6] Prem Selvaraj, Gokul N. C., Pratyush Kumar, and Mitesh M. Khapra. Openhands: Making sign language recognition accessible with pose-based pretrained models across languages. *CoRR*, abs/2110.05877, 2021.

[7] Sharvani Srivastava, Amisha Gangwar, Richa Mishra, and Sudhakar Singh. Sign language recognition system using tensorflow object detection API. *CoRR*, abs/2201.01486, 2022.

[8] Neil Abraham Usha Kiruthika, Mayank Mohan. Hand gesture recognition for emoji prediction. *International Journal for Research in Applied Science and Engineering Technology*, 8:1310–1317, 06 2020.

[9] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, W. Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. *ArXiv*, abs/1804.01654, 2018.

[10] Ying Wu and John Lin. Analyzing and capturing articulated hand motion in image sequences. *IEEE transactions on pattern analysis and machine intelligence*, 27:1910–22, 01 2006.

[11] Christian Zimmermann and Thomas Brox. Learning to estimate 3d hand pose from single rgb images. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4913–4921, 2017.