



“FabDepth I: A Unique Dataset for Efficient Gesture Detection”

Rameez Shamalik^{1,2} · Sanjay Koli^{1,3}

Received: 20 November 2022 / Accepted: 15 May 2023 / Published online: 13 June 2023

© The Author(s), under exclusive licence to Bharati Vidyapeeth’s Institute of Computer Applications and Management 2023

Abstract The recent Pandemic has highlighted the need of gesture detection systems as they become progressively crucial for touch less environment. Available datasets are singular and in general, lack depth and diversity. This paper presents FabDepth I, a unique dataset that not only provides excellent Foreground–Background (FGBG) Separated Images but also Depth map predicted images of 21 hand gestures. Together it forms 2100 images. To add variability and greater comprehension, a scaled version of the identical gestures in the wild is also presented which adds another 2100 images making the total tally of 4200 images.

Keywords Computer vision · Depth estimation · Foreground–background separation · Human–machine interaction

1 Introduction

Gestures have been used as a global language for communication since the dawn of time. Facial gestures aid in comprehending emotions, whereas body movements indicate human action and reaction. Although significant, much work has already been poured into Facial gesture detection [1–4], as

well as Human motion detection [5, 6], particularly in sports, deciphering body language, and transportation surveillance for which various video processing techniques have been utilized [7]. Because of the amount of permutations and combinations of motions performed by fingers and thumb, hand gesture identification is a unique task. It is essential in Human Machine Interaction (HMI) applications such as interactive gaming, augmented reality, virtual reality, smart homes, and industrial automation [8, 9].

Gesture detection is a fundamental component of computer vision applications [10]. The recent pandemic has underlined the need of human–machine association. Sign language has proven to be a lifeline for persons with special needs in communicating with the rest of the world. Many researchers have investigated various strategies for detecting sign language, which naturally incorporate hand motions [11–13]. Gesture detection technology is widely used in smart homes and gadgets. As a result, developing superior gesture detection system becomes imperative.

1.1 Dataset specifications summary

Subject	Artificial intelligence, Computer vision and pattern recognition, Human–computer interaction
Specific subject area	Gesture detection, Human–machine interaction, Depth estimation
Type of data Image	Resized image (480×360) RGB color space, (640×480 pixels) FGBG separated and depth estimated images
How the data were acquired	RGB Camera, ResNet 101, MiDas model, Graphics processing unit (GPU)
Data format	RAW

✉ Rameez Shamalik
shamalik1@gmail.com

Sanjay Koli
sanjay.koli@dypic.in

¹ G H Raisoni College of Engineering and Management, Pune, India

² Bharati Vidyapeeth’s College of Engineering for Women, Pune, India

³ Ajeenkya D Y Patil School of Engineering, Pune, India

Parameters for data collection	FGBG separated and depth estimated image
Description of data collection	4200 images consisting of hand gestures in three different format
Data accessibility	Mendeley Data https://data.mendeley.com/datasets/vvdy2x5vpr/1

2 Literature review

An efficient application of technology is unlocked by an effective dataset. Modern technology and decision making are highly reliant on data. As previously stated, several datasets for diverse Sign languages have been created due to the importance of gestures in Sign language recognition. There are various dataset versions available for hand gesture detection as well [14, 15]. These datasets are mostly focused on single application. Furthermore, the dataset's subject is peculiar. Some exhibit gestures in the wild, while others have a clear background or only display depth images of the gestures. Also, the shape and size of the gestures have never been emphasised. This variation in hand gestures will undoubtedly enhance the accuracy of the gesture detection system. Further said dataset finds its application in building blocks of dynamic hand gesture detection through 3D reconstruction of hand gestures in real time [16].

Gesture detection is the process of transforming an image or sequence of pictures, together into systematic representation. There are several methodologies, datasets, and techniques used for human gesture detection [17], however many of them are bulky [18, 19] as they incorporate videos of gestures, overwhelming the system. One of the most active research areas in machine vision and human-machine communication is hand gesture detection, with several potential applications including video games, healthcare systems, wearable technology and video processing [20, 21]. There are several image processing-based approaches accessible in the literature [22]. As shown in Table 1, multiple forms of datasets on hand gesture detection are present, however their specifications varies

depending on their application. IHMG [23] depicts hand motions from two different viewpoints captured via an RGB-D camera whereas, the HGM-4 dataset [24] depicts letters of Vietnamese sign language in gesture form from four distinct angles. GesHome dataset [25] consists of 18 hand gestures recorded using an embedded accelerometer and gyroscope from 20 subjects of various ages, whilst the Holographic micro-gesture recognition (HoMG) dataset [26] was recorded using a holographic 3D camera and includes only 3 mainstream gestures from 40 participants in various settings and conditions.

When it comes to high resolution image data, current datasets are rather large at the expense of data quality and clarity. Few datasets, such as the Leapgesture dataset [27], address this issue by including depth images, which enhances data accuracy and validity. Because of depth perception, data size may be considerably decreased without sacrificing dataset transparency. FabDepth I not only gives depth estimated images of 21 distinct gestures for use in Human-Machine interaction, but it also produces efficient FGBG separated images of the same gestures. In addition to these features a separate resized folder has scaled and normalized images having low light and noisy background.

3 Methodology

The experimental setup necessitates the use of a basic monocular RGB camera as well as a computer with a high-speed broadband internet connection. A minimum of 8 GB of RAM is necessary, as well as the most recent GPU. A GPU is specialized to image acquisition and processing, which speeds up the Dataset building process. The usage of an expensive stereo camera is avoided, reducing the system's cost and assuring exceptional repeatability. The size of the data plays a vital role in its application. Valid and clean but low sized data helps researchers to augment data whenever necessary, hence the resolution of the Depth estimated and FGBG separated images is kept at 640×480 while the resized gestures resolution is at 480×360 pixels.

Table 1 Comparison between various datasets

Dataset name	Number of gestures	Images per gesture	Depth images	Total images	Maximum resolution	Available publicly
IHMG [23]	8	–	Yes	836	640×480	Yes
HGM-4 [24]	26	40	No	4000	1280×700	Yes
GesHome [25]	18	50	No	18,000	–	Yes
HoMGR [26]	3	–	Yes	30,635	1920×1080	Yes
LeapGesture [27]	10	200	Yes	20,000	540×240	Yes
FabDepth I	21	50–100	Yes	4200	640×480	Yes

Figure 1 depicts the dataset generation procedure. For consistency, the images are acquired from a live video and downsized to 75% of their original size. A folder for scaled images is created, including 100 images of the same gesture. For FGBG separation of acquired and scaled gesture images, the Semantic Segmentation methodology is adopted. Deeplab ResNet 101 [28] algorithm is highly effective for semantic segmentation where the background is maintained black, highlighting maximum pixels for a gesture. Accordingly, a pre-trained ResNet 101 is employed to get FGBG separated images in real time. The FGBG split pictures are then fed into Intel's online MiDas model [29], which allows accurate Depth estimate of various gestures, thus having 3 different folders for all 21 gestures. This dataset is publicly made available on Mendeley Repository [30]. Researchers can either individually use or combine the three different formats of 21 gestures for training and testing as per their requirements.

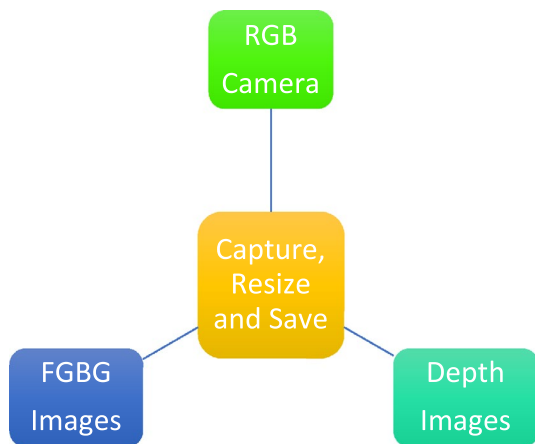


Fig. 1 Hierarchical process of inception of dataset

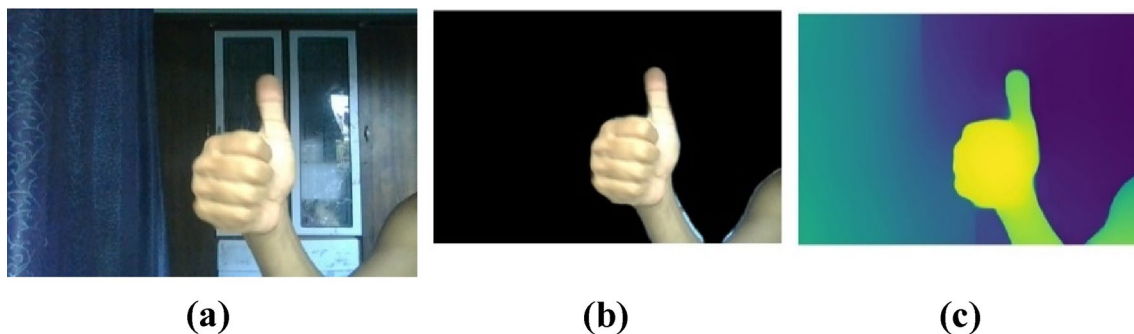


Fig. 2 **a** A low-light, noisy background gesture is scaled, **b** FGBG separated, cropped and centered image, **c** Depth estimated image of the same gesture

4 Results and discussion

This particular data can help train the machine to recognize gestures in the wild. As seen in Fig. 2. The gesture caught in low light is first resized, then background noise is totally removed using the FGBG separation technique, and finally depth in the picture is discovered where the gesture is automatically highlighted. This method provides researchers with new perspectives and ideas to work with.

A meaningful set of gestures is chosen on the basis of their use in HMI and one-on-one communication amongst persons with special needs. A total of 21 gesture folders are organised into three groups: Depth-Gestures, FGBGS-Gestures, and Resized-Gestures. Each of the first two sub-folders has 50 images. Depth gesture images are unique in that they are not only resistant to poor light and background noise, but they also allow users to comprehend disparity in the given frame. In terms of hand gestures, there is a considerable risk of occlusion attributed to finger proximity and intertwining, which can impair the precision of FGBG gesture separation. Having a depth map estimation of the same gesture allows for substantially better investigation. The Resized Gestures folder contains 100 samples of the same gesture that have been scaled and normalised but still provide a realistic setting for training a machine to recognise gestures in the wild. The FabDepth I dataset is summarised in Table 2 based on its description.

Aside from gestures indicating digits 0 through 9, other unconventional hand gestures are provided that might be readily employed in HMI or other computer vision solutions. Figure 3 outlines a few of them.

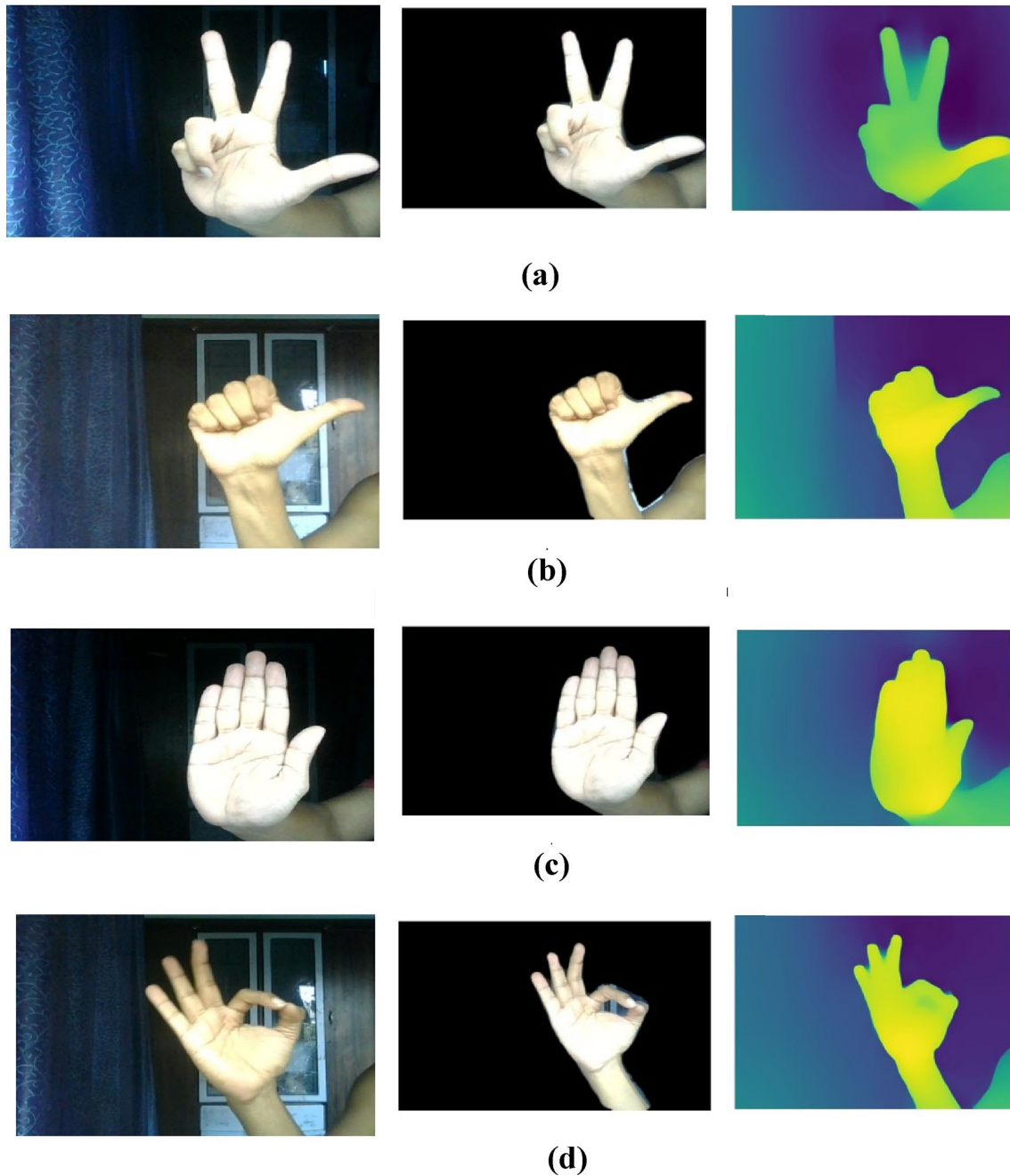
5 Conclusion and future scope

FabDepth I dataset is designed for hand-gesture identification and includes 21 distinct gestures that may be utilised in HMI. In contrast to the other public datasets, this is the first

Table 2 Key points of FabDepth I dataset

Type of images	Number of gestures	Number of images per gesture	Total number of images
Depth-gestures	21	50	1050
FGBGS-gestures	21	50	1050
Resized-gestures	21	100	2100

to combine hand-gestures in the wild with low light background and noise along with FGBG separated and Depth Estimated ones. There is no special camera used to measure depth, and no additional software or processing is done to discover FGBG split photos for various gestures. This dataset may be used to investigate hand-gesture detection challenges from numerous perspectives and improve the system's depth accuracy. Potential applications include sign language comprehension, contactless intelligent systems, augmented reality, virtual reality and interactive gaming.

**Fig. 3** Few of the gestures which are Captured-resized, Center cropped, FGBG Separated and Depth estimated. **a** On, **b** Right, **c** Stop, **d** Ok

Researchers may approach this dataset through three different selections and combine any two of the three modalities in accordance with their requirements. As a result of this deconstruction, FabDepth I stands apart as a benchmark dataset for Hand gesture detection.

For future work, a variety of diverse samples, such as women, children, and senior citizens, can be employed [31]. A dataset that focuses entirely on HMI and can incorporate gestures done with two hands can also be created.

Funding No funding has been received for this research.

Declarations

Conflict of interest The authors declare no conflict of interest. The authors have no competing interests to declare that are relevant to the content of this article.

References

- Putro MD, Nguyen D, Jo K (2022) A fast CPU real-time facial expression detector using sequential attention network for human-robot interaction. *IEEE Trans Ind Inf*. <https://doi.org/10.1109/TII.2022.3145862>
- Chen Y, Wu H, Wang T, Wang Y, Liang Y (2021) Cross-modal representation learning for lightweight and accurate facial action unit detection. *IEEE Robot Autom Lett* 6:7619–7626
- Gangonda SS, Patavardhan P, Karande KJ (2021) VGHN: variations aware geometric moments and histogram features normalization for robust uncontrolled face recognition. *Int J Inf Technol* 1:1–12
- Liang J, Wang J, Quan Y, Chen T, Liu J, Ling H, Xu Y (2022) Recurrent exposure generation for low-light face detection. *IEEE Trans Multimedia* 24:1609–1621
- Lang Y, Hou C, Ji H, Yang Y (2021) A dual generation adversarial network for human motion detection using micro-doppler signatures. *IEEE Sens J* 21:17995–18003
- Dong Z, Li F, Li Z, Pahlavan K (2022) A study of on-body RF characteristics based human body motion detection. *IEEE Sens J* 22:3442–3454
- Koli SM, Shamalik RM (2019) Transformation of video signal processing techniques from 2D to 3D: a survey. *Lect Notes Electr Eng* 570:63–70
- Zhang X, Sun Y, Zhang Y (2022) Evolutionary game and collaboration mechanism of human-computer interaction for future intelligent aircraft cockpit based on system dynamics. *IEEE Trans Hum Mach Syst* 52:87–98
- Yadav S, Chakraborty P, Kochar G, Ansari D (2020) Interaction of children with an augmented reality smartphone app. *Int J Inf Technol* 12:711–716
- Mohamed N, Mustafa MB, Jomhari N (2021) A review of the hand gesture recognition system: current progress and future directions. *IEEE Access* 9:157422–157436
- Wei C, Zhao J, Zhou W, Li H (2021) Semantic boundary detection with reinforcement learning for continuous sign language recognition. *IEEE Trans Circuits Syst Video Technol* 31:1138–1149
- Zhao T, Liu J, Wang Y, Liu H, Chen Y (2021) Towards low-cost sign language gesture recognition leveraging wearables. *IEEE Trans Mob Comput* 20:1685–1701
- Hisham B, Hamouda AE (2020) Arabic sign language recognition using Ada-Boosting based on a leap motion controller. *Int J Inf Technol* 13:1221–1234
- Ma J, Li T, He J (2020) A Comprehensive Hand Gesture Dataset. *Int Confer Data Sci (CDS) 2020*:328–333
- Liu C, Yang Y, Liu X, Fang L, Kang W (2021) Dynamic-Hand-Gesture Authentication Dataset and Benchmark. *IEEE Trans Inf Forensics Secur* 16:1550–1562
- Shamalik R, Koli S (2022) DeepHands: Dynamic hand gesture detection with depth estimation and 3D reconstruction from monocular RGB data. Sādhanā. <https://doi.org/10.1007/s12046-022-02026-7>
- Shamalik R, Koli SM (2021) Real time human gesture recognition: methods, datasets and strategies. *Recent trends in intensive computing*
- Zhang Y, Cao C, Cheng J, Lu H (2018) EgoGesture: a new dataset and benchmark for egocentric hand gesture recognition. *IEEE Trans Multimedia* 20:1038–1050
- Joze HR, Koller O (2018) MS-ASL: A large-scale data set and benchmark for understanding american sign language. *ArXiv, abs/1812.01053*
- Shukla D, Erkent Ö, Piater JH (2016) A multi-view hand gesture RGB-D dataset for human-robot interaction scenarios. 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 1084–1091
- Shamalik R, Koli SM (2020) Emergence and Functionality of 3D Videos. *Int J Eng Adv Technol* 9(3):4319–4322
- Ahmad F, Ahmad W (2022) An efficient astronomical image processing technique using advance dynamic workflow scheduler in cloud environment. *Int J Inf Technol* 14:2779–2791
- Shukla D, Erkent O, Piater JH (2015) The IMHG dataset : A multiview hand gesture RGB-D dataset for human-robot interaction
- Hoang VT (2020) HGM-4: A new multi-cameras dataset for hand gesture recognition. *Data Brief*. <https://doi.org/10.1016/j.dib.2020.105676>
- Khanh Nguyen-Trong, Cuong Pham, Nam Vu Hoai (2020) "GesHome Dataset". *IEEE Dataport*. <https://doi.org/10.21227/3qt8-yh80>
- Liu Y, Meng H, Swash MR, Gaus YF, Qin R (2018) Holoscopic 3D Micro-gesture database for wearable device interaction. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 802–807
- Mantecón T, del-Blanco CR, Jaureguizar F, García N (2016) Hand Gesture Recognition Using Infrared Imagery Provided by Leap Motion Controller. *ACIVS*
- Chen L, Papandreou G, Kokkinos I, Murphy KP, Yuille AL (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell* 40:834–848
- Ranftl R, Lasinger K, Hafner D, Schindler K, Koltun V (2022) Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer. *IEEE Trans Pattern Anal Mach Intell* 44:1623–1637
- Shamalik RM (2022) FabDepth I, Mendeley Data, V1. <https://doi.org/10.17632/vvdy2x5vpr.1>
- Shamalik R (2023) FabDepth HMI. *IEEE Dataport*. <https://doi.org/10.21227/ek5d-r453>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.