ToF Based Wearable Sensing for Passive Food Intake Monitoring

Harshavardhan Sasikumar[‡], Rashmi Wijesundara[§], Sarah Deemer[†], Xiaohui Yuan[‡], Megan Wesling[*], Mahdi Pedram[‡]
[‡]Department of Computer Science & Engineering, University of North Texas, Denton, TX, USA
[*] Department of Pharmacotherapy, University of North Texas Health Science Center, Fort Worth, TX, USA
[†]Department of Kinesiology Health Promotion and Recreation, University of North Texas, Denton, TX, USA
[§]Department of Biomedical Engineering, University of North Texas, Denton, TX, USA

Abstract—**Diet plays a crucial role in preventing chronic diseases such as type 2 diabetes and heart disease. Most existing diet monitoring systems require manual input or raise privacy concerns by continuously recording video data that often captures the user's face or the surrounding environment. In this paper, we present a chest-mounted wearable device that preserves user privacy while passively tracking dietary intake using a Time-of-Flight (ToF) sensor. Captured RGB images are masked using ToF depth data to isolate food items and eliminate background elements. A FOMO-based food detection model achieved an F1 score of 96% and a mean Average Precision (mAP) of 74% on masked images, outperforming its performance on unmasked RGB inputs. Also, ToF depth frames were used to build an eating gesture recognition model that achieved 88% accuracy, indicating reliable identification of eating gestures. All models and image processing steps were executed on-device, demonstrating the feasibility of the system. This work presents a novel approach for real-time dietary monitoring that addresses both user privacy and food detection accuracy in a wearable health system.**

I. Introduction

Diet plays a vital role in everyone's life, significantly influencing health outcomes and the risk of chronic diseases such as type 2 diabetes (T2D), cardiovascular disease, and obesity [1]. Numerous studies have established a strong connection between eating habits and overall health, highlighting diet as a key determinant in the development and management of these diseases [2], [3]. Type 2 diabetes, in particular, is a chronic disease that affects millions worldwide, with an estimated 462 million people affected worldwide [4]. Self-monitoring of dietary intake has been shown to be especially effective in facilitating change in diet behavior and improving glycemic control [5]. With diet, activity, and metabolism closely linked, tracking tools should support behavior change with minimal burden and align with real-life habits.

Over the years, numerous methods have been developed to monitor diet and eating habits in real time, offering live feedback to promote healthier behaviors and enhance self-awareness [6]. However, some wearable systems designed for dietary monitoring pose security risks, as they often collect sensitive health data [7], [8]. Additionally, many of these devices adopt a smartwatch form factor, which can be uncomfortable for extended use, reducing user convenience and long-term adherence [9]. A major concern with current food intake monitoring systems is the inherent privacy risk of unintentionally capturing individuals' faces during meal tracking. This issue is particularly critical when data is processed off-device, increasing the risk of unauthorized access, misuse, and identity exposure [10]. Furthermore, many existing systems focus on capturing visible food rather than eating gestures, resulting in unnecessary on-device storage consumption, increased battery usage, and a failure to capture the core behavior of intake. The reliance on constant video recordings in certain systems not only presents ethical issues but also results in data overload, which complicates post-processing, making it time-consuming and inefficient in resource-limited environments [11].

Even though the recent work by Doulah et al. [12] demonstrates that their system effectively captures food intake events and detects non-eating scenarios using machine learning techniques, it still has several limitations. During food intake sessions, the system frequently captures the user's face or nearby individuals within the frame, raising privacy concerns despite attempts to filter out irrelevant frames. Additionally, the system's usability is restricted by its spectacle-based form factor, which is unsuitable for many users. Similarly, sensors that detect chewing [13] and bio-impedance systems such as iEat [14] both rely on continuous skin contact, making them uncomfortable for long-term use. While they can identify intake activities with good accuracy in controlled settings, their performance is sensitive to noise, electrode placement, and utensil choice, which limits their practicality in real-world environments.

In this paper, we present a wearable device that passively captures and stores images of meals consumed in a manner that preserves privacy to track the diets of participants. Unlike continuous food monitoring systems that risk capturing people's faces through live video feeds, our system captures images only during detected eating gestures. This approach, enabled by a Time-of-Flight (ToF) sensor, ensures that image capture occurs only when a user is actively eating, thus conserving storage and battery resources. The captured images and their depth information are processed to detect regions containing food. Background elements such as people's faces are masked using ToF depth data, retaining only the food item for analysis. This approach provides strong privacy protection and addresses the limitations of existing wearable systems that risk exposing sensitive visual data.

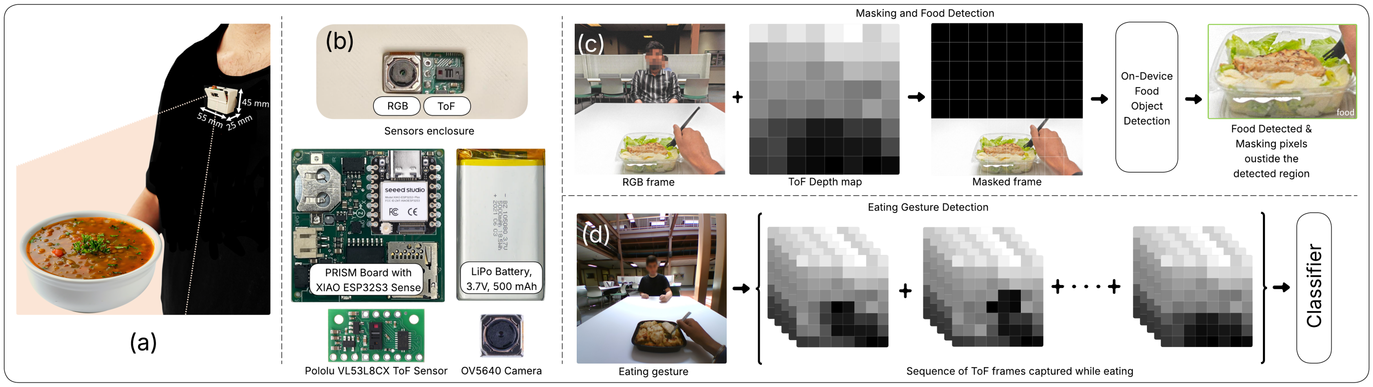
(a)
45 mm
55 mm
25 mm
(b)
RGB
ToF
Sensors enclosure
PRISM Board with XIAO ESP32S3 Sense
LiPo Battery, 3.7V, 500 mAh
Pololu VL53L8CX ToF Sensor
OV5640 Camera
Masking and Food Detection
(c)
RGB frame
ToF Depth map
Masked frame
On-Device Food Object Detection
food
Food Detected & Masking pixels oustide the detected region
Eating Gesture Detection
(d)
Eating gesture
Sequence of ToF frames captured while eating
Classifier

Fig. 1. Diagram of the methodological framework; (a) Chest-level wearable device that collects food data along with its depth information. (b) Components and sensors in the enclosure. (c) We mask the RGB data to protect the privacy of the background elements using the ToF depth map and feed it to an on-device food object detection. (d) We collect depth information from the ToF sensor during eating sessions and process it for eating gesture recognition.

II. Methods

Figure 1 illustrates our chest-level wearable, which captures RGB and ToF data with cameras enclosed in a compact case. ToF depth maps are used to mask far-distance objects for privacy before food detection using Faster Objects, More Objects (FOMO) [15], and also support eating gesture recognition.

A. System Design

The main objective of our design is to record RGB videos during the eating sessions. For good data collection, the feed must be as fast and continuous as possible. Adopting a compact form factor improves the user experience by making the device small and lightweight, also making it a more reliable and practical system.

The core of the system is built on an XIAO ESP32S3 Sense microcontroller, selected for its compact size, onboard Wi-Fi, and camera support. We used an OV5640 camera to capture the raw RGB images and the VL53L8CX ToF sensor to capture the depth array. The ToF sensor communicates over the I2C interface, delivering real-time depth data that is used to mask distant or irrelevant regions of the RGB frame. These masked frames are then logged onto an SD card or streamed via SoftAP Wi-Fi, depending on the operation phase.

A DS3231 RTC module synchronized timestamps between depth and RGB frames, enabling accurate post-processing for gesture recognition and behavioral monitoring. The system is powered by a 500 mAh LiPo battery, providing portability and on-the-go data capture. For system integration, a custom-made PRISM board is produced to house the ESP32S3 Sense, the ToF sensor, the RTC, and an SD card interface. All I2C peripherals and GPIO sections are integrated in this board, which simplifies the wiring and improves reliability for prolonged deployments.

To facilitate user inspection or debugging, the ESP32S3 intermittently switches to SoftAP mode, establishing a local Wi-Fi hotspot. While in this mode, live RGB frames are transmitted over HTTP to connected devices. After a set period, Wi-Fi is turned off to save power, and the system reverts to SD-based logging.

B. Usability Study

1) Participants: The study comprised 31 meal sessions with 15 participants instructed to eat naturally while wearing the device. The system was tested in various environmental conditions, including 21 indoor meals, 5 in low-light conditions, and 5 outdoor sessions in direct sunlight, collecting 21,052 images. The data was split into 70% training, 15% validation, and 15% testing with a near-balanced distribution of 10,500 positive and 10,552 negative samples.

2) Procedures: Participants were provided with three identical devices for data collection. Each device was chest-mounted with magnets, providing an unobstructed view of the plate and food. This placement of the device replicates the natural perspective of diners. Before each session, participants were instructed to eat normally without enforcing their posture relative to the camera. After each session, the data were evaluated to assess sensor accuracy, food visibility, and system performance under varying lighting conditions.

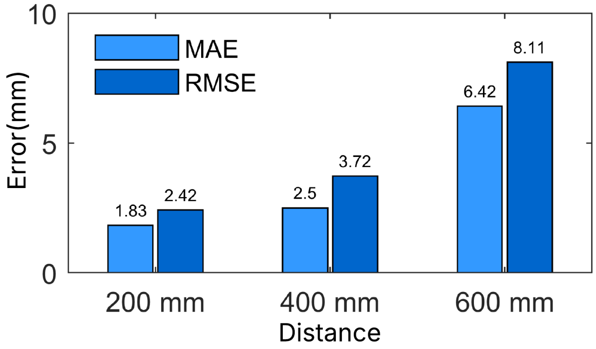

Fig. 2. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) of the ToF sensor at different distances. The plot illustrates increasing error trends with distance, reflecting the sensor's reduced accuracy at longer ranges.

III. Results

As illustrated in Figure 2, the ToF sensor demonstrates high accuracy at shorter distances, with error increasing progressively as distance grows. Both MAE and RMSE metrics captured this trend, corresponding with the expected reduction of precision in depth measurements with an increase in distance. To compute these metrics, an object was positioned at the center, and the Time of Flight (ToF) measurements were taken in comparison with the ground truth distances. Table I shows the ToF sensor's SNR metrics, indicating a high signal-to-noise ratio and reliable depth measurements under the tested conditions.

TABLE I
SNR METRICS OF THE TOF SENSOR

Metric	Value
SNR (Linear)	74.36
SNR (dB)	37.42 dB

A. Food Detection Performance

To evaluate the impact of depth-based cropping on food detection performance, we trained two FOMO object detection models [15] under identical configurations. One model used full-frame RGB images, while the other employed ToF sensor depth-masked images. In the masked version, the outer regions of the image, determined by the ToF sensor's depth measurement, were blacked out. This enabled the model to concentrate on the region that is closest to the user and where food is more likely to be present.

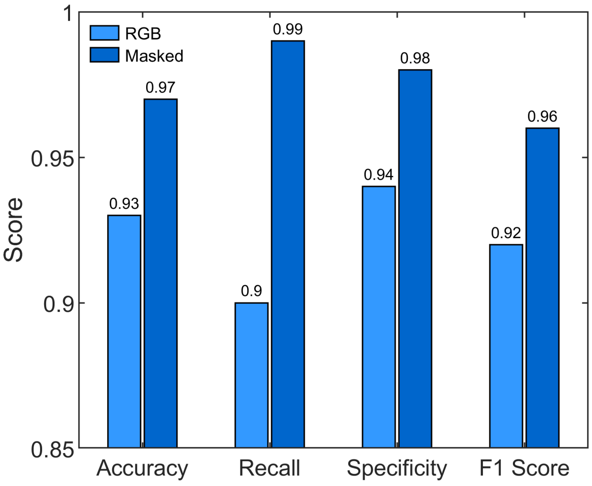

Fig. 3. Comparison of metrics between models trained on unmasked RGB images and ToF-masked images. The masked model highlights the effectiveness of ToF-based masking in enhancing food detection accuracy.

As shown in Figure 3, the results demonstrate a significant performance improvement with the ToF-masked input. The masked model achieved a mean average precision (mAP) of 73.4%, compared to 63.1% with unmasked RGB images. Specifically, the F1 score increased from 92% (RGB) to 96% (Masked), indicating improved balance between precision and recall. Although the full-frame RGB model performed well overall, the ToF-masked model showed improvement in suppressing noisy detections and maintaining consistency in cluttered environments. This is because irrelevant background areas masked in the images tend to have lower detection rates, resulting in fewer falsely positive identifications and greater inter-class confidence. This validates the assumption that using spatial attention by means of depth-based masking increases model robustness, especially in complex visuals.

Both models use a MobileNetV2-based architecture, which includes a frozen ImageNet-trained backbone for streamlined feature extraction. A GlobalAveragePooling2D layer is followed by a dense head with 128 ReLU units and a final Dense layer yielding six sigmoid values: four for bounding box coordinates, one for object confidence, and one for binary class (food/background). This architecture is particularly effective for precise edge inference on 96 × 96 RGB and masked images.

B. RGB Cropping Based on ToF Proximity

To evaluate the consistency and accuracy of our ToF-based cropping method, we computed the Intersection-over-Union (IoU) between the cropped region and manually annotated food bounding boxes across the entire dataset. Our method achieved a mean IoU of 0.712 with a standard deviation of 0.08, indicating reliable and precise localization of food regions across diverse meal scenarios. As shown in Figure 4 and Figure 5, the ToF-based masking utilizes depth proximity to retain only the near-field regions, usually where food is located, thus isolating the relevant area in the RGB frame. This level of cropping enhances the effectiveness of subsequent food detection. By analyzing only the relevant parts of the RGB frame, rather than dealing with extensive background clutter, computational resources are optimized, enhancing detection accuracy.

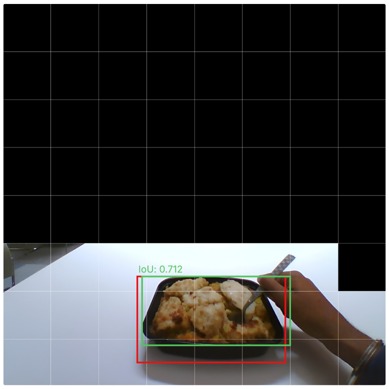

Fig. 4. Visualization of predicted and ground truth bounding boxes over a food item to illustrate Intersection-over-Union (IoU) calculation. Higher IoU indicates better alignment with the ground truth.

C. Eating Gesture Detection Using ToF Sensing

The objective was to recognize the distinctive hand-to-mouth movement executed when using eating utensils. The participants had the device mounted on their chest while they had their meals. The ToF sensor captured depth frames of size 8 x 8 at 5 Hz. The frames were subjected to median filtering to remove noise and magnify the distracting movement patterns.

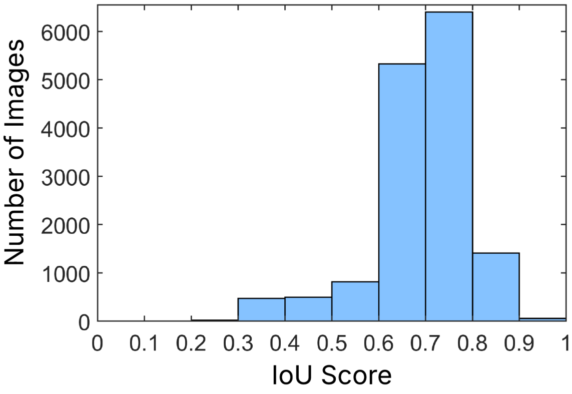

Fig. 5. Distribution of IoU scores between the cropped regions and ground truth food bounding boxes. Most samples fall within the 0.7–0.8 range, indicating consistent and accurate cropping of food-relevant areas.

Eating gestures were identified as changes in depth zones, with the hand being taken from the plate to the mouth. Each gesture sample included 15 sequential ToF frames; therefore, each gesture sample had a feature vector of 960 dimensions (15 x 64). We applied a sliding window technique to segment gestures in each session, allowing for overlap. We created ground truth labels using RGB frames synchronized with RTC by marking eating or non-eating gestures based on the hand's position relative to the food or mouth. ToF depth frames taken during an eating gesture sequence are shown in Figure 6, which captures the patterns correlating with hand movement. The gesture classification model was trained using the Adam optimizer with a binary cross-entropy loss function. This lightweight fully connected model achieved a test accuracy of 88%, proving that ToF data can effectively capture temporal hand dynamics during food intake. Importantly, this method allows for the energy-efficient capture of RGB images only while eating, thus conserving battery life, saving on memory, enhancing privacy by minimizing irrelevant data collection, and reducing data storage.

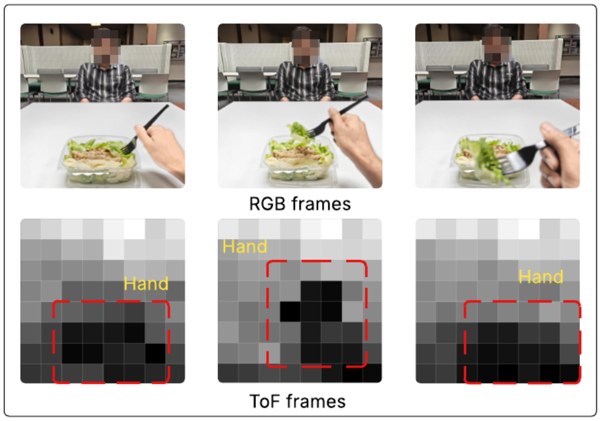

Fig. 6. RGB frames and corresponding ToF depth maps showing an eating gesture. The visible hand in depth maps highlights the ToF sensor's ability to capture user-proximal motion for gesture recognition.

IV. Conclusion and Future Work

This work presents a chest-mounted wearable with RGB and ToF sensors that captures eating sessions and preserves privacy by masking backgrounds while detecting food items. The study shows the device works across lighting conditions while preserving privacy, and that ToF-based masking improves food detection by focusing on food regions. Future work includes using ToF depth data to estimate food volume and calorie intake. We also plan to add an Inertial Measurement Unit to assess activity levels through (Metabolic Equivalent of Task) MET values, which can further enhance calorie estimation accuracy.

References

[1] Y.-L. Xiao, Y. Gong, Y.-J. Qi, Z.-M. Shao, and Y.-Z. Jiang, "Effects of dietary intervention on human diseases: molecular mechanisms and therapeutic potential," *Signal Transduction and Targeted Therapy*, vol. 9, no. 1, p. 59, 2024.

[2] J. F. López-Gil and P. J. Tárraga-López, "Research on diet and human health," p. 6526, 2022.

[3] H. Cena and P. C. Calder, "Defining a healthy diet: evidence for the role of contemporary dietary patterns in health and disease," *Nutrients*, vol. 12, no. 2, p. 334, 2020.

[4] M. Abdul Basith Khan, M. J. Hashim, J. K. King, R. D. Govender, H. Mustafa, and J. Al Kaabi, "Epidemiology of type 2 diabetes—global burden of disease and forecasted trends," *Journal of epidemiology and global health*, vol. 10, no. 1, pp. 107–111, 2020.

[5] R. Misra and D. James, "The role of dietary tracking on changes in dietary behavior in a community-based diabetes prevention and management intervention," *Public Health Nutrition*, pp. 1–29, 2025.

[6] G. J. Fernandes, J. Zheng, M. Pedram, C. Romano, F. Shahabi, B. Rothrock, T. Cohen, H. Zhu, T. S. Butani, J. Hester *et al.*, "Habitsense: A privacy-aware, ai-enhanced multimodal wearable platform for mhealth applications," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 8, no. 3, pp. 1–48, 2024.

[7] A. Sifaoui and M. S. Eastin, "““whispers from the wrist": wearable health monitoring devices and privacy regulations in the us: the loopholes, the challenges, and the opportunities," *Cryptography*, vol. 8, no. 2, p. 26, 2024.

[8] X. Yuan and M. Gomathisankaran, "Secure medical image processing for mobile devices using cloud services," in *Electronic Imaging Applications in Mobile Healthcare*. SPIE Press, 2016.

[9] L. Wang, M. Allman-Farinelli, J.-A. Yang, J. C. Taylor, L. Gemming, E. Hekler, and A. Rangan, "Enhancing nutrition care through real-time, sensor-based capture of eating occasions: a scoping review," *Frontiers in Nutrition*, vol. 9, p. 852984, 2022.

[10] B. Zhang, C. Chen, I. Lee, K. Lee, and K.-L. Ong, "A survey on security and privacy issues in wearable health monitoring devices," *Computers & Security*, p. 104453, 2025.

[11] A. Yoon, K. M. Jones, and L. Spotts, "Ethical use of lifelogging data for research: Perceived value and privacy concerns of wearable camera users," *Nordic Journal of Information Science and Culture Communication*, vol. 6, no. 1, 2017.

[12] A. Doulah, T. Ghosh, D. Hossain, M. H. Imtiaz, and E. Sazonov, "““automatic ingestion monitor version 2"–a novel wearable device for automatic food intake detection and passive capture of food images," *IEEE journal of biomedical and health informatics*, vol. 25, no. 2, pp. 568–576, 2020.

[13] T. Vu, F. Lin, N. Alshurafa, and W. Xu, "Wearable food intake monitoring technologies: A comprehensive review," *Computers*, vol. 6, no. 1, p. 4, 2017.

[14] M. Liu, B. Zhou, V. F. Rey, S. Bian, and P. Lukowicz, "ieat: automatic wearable dietary monitoring with bio-impedance sensing," *Scientific Reports*, vol. 14, no. 1, p. 17873, 2024.

[15] Edge Impulse, "Announcing fomo (faster objects, more objects)," 2022, accessed: 2025-06-05. [Online]. Available: https://www.edgeimpulse.com/blog/announcing-fomo-faster-objects-more-objects