The Visual Accelerometer: A High-fidelity Optic-to-Inertial Transformation Framework for Wearable Health Computing

Chenhan Xu*, Huining Li*, Zhengxiong Li†, Xingyu Chen†, Aditya Singh Rathore*,
Hanbin Zhang*, Kun Wang‡, Wenyao Xu*
*Dept. of Computer Science and Engineering, University at Buffalo, USA
†Dept. of Computer Science and Engineering, University of Colorado Denver, USA
‡Dept. of Electrical and Computer Engineering, University of California, Los Angeles, USA
*{chenhanx, huiningl, asrathor, hanbinzh, wenyaoxu}@buffalo.edu,
†{zhengxiong.li, xingyu.chen}@ucdenver.edu, ‡kun.wang@ieee.org

2022 IEEE 10th International Conference on Healthcare Informatics (ICHI) | 978-1-6654-6845-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICHI54592.2022.00053

***Abstract*—Human activities of daily life (ADL) monitoring has been applied in life-critical applications such as occupational safety and stroke rehabilitation tracking. However, wearable computing, as the main technical paradigm of ADL monitoring, requires tremendous efforts to obtain satisfactory inertial data with labels for training. In this paper, we develop *VisualAcc*, a high-fidelity optic-to-inertia framework of human locomotion for wearable computing, which leverages harvested light-intensity data from public videos to reconstruct authentic wearable motion data. Specifically, a two-step optical motion estimator is first designed to infer the high-quality optical motion field (OMF) from the time-varying light intensity. Then, the obtained OMF is fed to an optic-to-inertia transformer, which leverages human kinematics constraints in light ray projection to recover time-sequential inertial data in a convolution-based process. Experimental results show over 0.86 Pearson Correlation Coefficient between reconstructed data via *VisualAcc* and ground truth from authentic off-the-shelf MEMS sensors. Furthermore, we conduct a case study on IMU inverse human dynamics analysis to show *VisualAcc*'s potential in empowering and transforming fine-grained wearable computing.**

***Index Terms*—wearable computing; activities of daily life monitoring; health data system**

I. Introduction

Human activities of daily life (ADL) are a key indicator of individual health status [1], [2] and have been applied to different areas, including smart buildings [3], smart health [4], and human-computer interaction [5]. Particularly, monitoring human activity and behaviors in daily life is highly related to life-critical applications, such as stroke rehabilitation tracking [6] and heart disease prediction [7]. The marketplace of ADL monitoring technologies is forecast to reach 2.6 billion by 2023 [8]. However, ADL data, the cornerstone of developing various applications, is still lacking due to the humongous efforts (e.g., building sensor systems, recruiting participants, labeling, etc.) required in the data collection stage.

Numerous efforts have been made to address the data starvation problem. Large datasets of human motion [9], [10] were established to include as many activities and subjects as possible. Multiple Inertial Measurement Unit (IMU) sensors

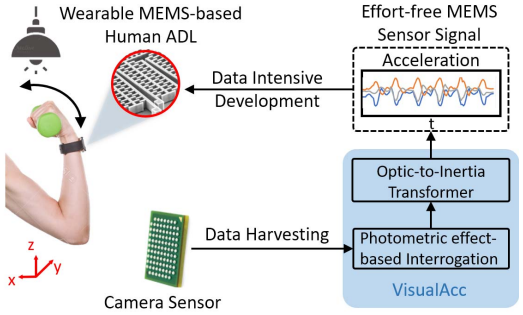


Fig. 1: *VisualAcc* facilitates the optic-to-inertia conversion from harvested light-intensity data to authentic wearable accelerometer data for effort-free wearable computing.

are used in datasets to cover various body areas. However, the amount and diversity of data are still not satisfactory for ever-emerging wearable computing applications, which is in stark contrast to the massive datasets in other domains (e.g., image, audio) that are far more diverse in terms of scenarios and activities. The development of machine learning in the past decade reveals an opportunity to use the Generative Adversarial Network (GAN) for wearable data generation by approximating the distribution of existing wearable data [11]. Nevertheless, GAN-created data is usually biased and it cannot generate new labels in terms of demographics and scenes to fulfill the demand of training data.

The recent success of vision-based human activity recognition reveals that videos also carry rich motion data. More importantly, as a device-free solution, camera has less obligation for humans and is used everywhere. Thereby, it is convenient to obtain diverse labeled data for training from public video sources, such as Youtube [12] and TikTok [13]. Recent studies [14]–[16] envision a reconciliation between vision and wearable computing methods, i.e., harvesting IMU data from videos to train the model for the always-on wearable

2575-2634/22/$31.00 ©2022 IEEE
DOI 10.1109/ICHI54592.2022.00053

computing applications. Although these existing works have achieved good performance in activity recognition, they compromise the authenticity of the harvested IMU data. Thereby, the generated IMU data can not be used for life-critical applications, such as post-traumatic rehabilitation [17] and neural disease progress control [6]. Our work shares the vision of these prior systems and aims to extract authentic inertial motion data for the first time to facilitate the development of fine-grained wearable computing.

In this study, we explore and unveil the opportunity of extracting high-fidelity inertial motion from video optic information for wearable computing. Specifically, different light rays are reflected as the subject changes position and orientation, resulting in sensible time-varying light intensity distribution, namely photometric effect. Therefore, with the accurate modeling of the photometric effect and human locomotion, authentic inertial motion information can be inferred from the video optical variations. To achieve this goal, there are three key technical challenges. 1) How to accurately interrogate human motion with the modeling of photometric effect and intensity variation, given the inevitable hinders such as illumination alteration and ambient occlusion? 2) How to extract and reconstruct three-dimensional (3-D) inertial data of specific human body areas using interrogated information? 3) How to evaluate the system usability in wearable human activity monitoring applications?

To this end, we present *VisualAcc*, an effort-free wearable computing framework, to facilitate the optic-to-inertia conversion from harvested light-intensity data to authentic wearable motion data, as illustrated in Fig. 1. *VisualAcc* takes ordinal optic data from camera optics sensors as the input, the processing chain comprises of three steps. First, *VisualAcc* relates intensity variation in optic data to optical motion via Horn and Schunck's theory [18], and optical motion reconstruction can be formulated as an energy function minimization problem. To address the challenges in optic data integrity and ambiguity, we augment intensity patterns and estimate optical motions in a two-stage fashion, i.e., formulating the condition-selective energy function and initial Optical Motion Field (OMF) [19] refinement. Second, an optic-to-inertia transformer is developed to track different components in OMF. We utilize the kinematics constraints in light ray projection to reconstruct time-series inertial data in a convolution-based process. Finally, the *VisualAcc* evaluation includes both publicly available datasets and in-situ wearable computing. We comprehensively examine the optic-to-inertia conversion in terms of fidelity, integrity, and authenticity in three presentation levels, i.e., motion data, locomotion features, and real-world applications (e.g., human activity recognition). Results show that the Pearson Correlation Coefficient (PCC) between reconstructed motion data and ground truth can achieve over 0.86. Also, an activity recognition model using *VisualAcc* can reach above 92% accuracy on average. Furthermore, to demonstrate the capability of *VisualAcc* in empowering fine-grained wearable computing applications, a case study on multiple-IMU human kinematics and dynamics analysis is conducted. These results indicate that *VisualAcc* is a promising framework to empower and transform wearable computing.

Our contribution in the work has three-fold:

- We are the first to explore the data conversion from the optic domain to the inertia domain for wearable health applications. We discover that the reflected light rays by moving subjects can result in time-varying intensity in the optic data, which can be leveraged to reconstruct inertial motion data.
- We develop *VisualAcc*, an effort-free framework, to facilitate the optic-to-inertia conversion from harvested light-intensity data to authentic wearable motion data for wearable computing. We design a two-step optical motion estimator to extract high-quality Optical Motion Field (OMF) from light intensity variation. Then, a novel optic-to-inertia transformer with kinematics prior knowledge is proposed to reconstruct inertial data from the OMF.
- We conduct extensive experiments to evaluate *VisualAcc* in both simulated and real-world scenarios. We demonstrate that reconstructed motion data is close to the data from off-the-shelf wearable MEMS sensors and can facilitate real-world applications. A case study of using *VisualAcc* for fine-grained multi-IMU human inverse kinematics and dynamics is investigated.

II. *VisualAcc* Overview

In this paper, we present *VisualAcc*, an optic-to-inertia conversion paradigm. The overview is illustrated in Fig. 2, consisting of two parts:

Photometric Effect based Interrogation: We design a photometric effect-based interrogation module to passively sense the human motion under the influence of inevitable hinders. Specifically, gradient augmentation is first applied to the incoming optic data for strengthening the intensity pattern. Then, *VisualAcc* establishes and solves a selective energy function minimization problem to obtain the initial Optical Motion Field (OMF). After that, an occlusion-resilient method is developed to disambiguate the OMF further.

Optic-to-Inertia Transformer: The interrogated information (i.e., OMF) is then fed to an Optic-to-Inertia transformer for inertial data reconstruction. *VisualAcc* adopts a specifically designed transformer with prior knowledge of human kinematics. This transformer first tracks the movement of different human body areas-of-interest in the OMFs and then cognizes the human kinematics constraints among them. The following inertial data reconstructing module leverages the tracked movement and the cognized constraints to recover the 3-D authentic inertial data. As the reconstruction is formulated as an optimization problem, we specifically illustrate the data flow for optimization in Fig. 2.

III. Photometrics Effect based Motion Interrogation

In this section, we illustrate how *VisualAcc* passively interrogates human motion with the model of photometric effect

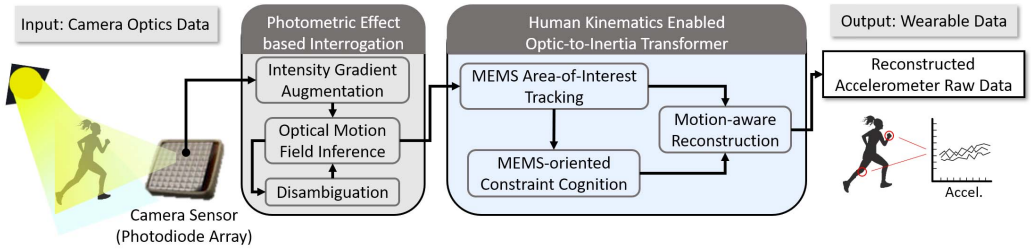


Fig. 2: The paradigm of *VisualAcc*. *VisualAcc* leverages device-free light intensity data to reconstruct wearable accelerometer raw data via Photometric effect-based interrogation and human kinematrics enabled optic-to-inertia transformer.

and intensity variation, and get an Optical Motion Field (OMF) as the output of interrogation.

A. Intensity Gradient Augmentation

The light intensity faces the attenuation problem in the propagation process, which will compromise optic data integrity and the quality of the intensity pattern. To resolve this problem, we augment intensity patterns. We adopt the following enhanced isotropic Laplacian filtering to capture the intensity gradient:

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \sum f(\alpha, \beta) - 8f(x, y), \tag{1}$$

where Δ is the Laplacian operator, $f(x, y)$ is the gray-scale function of the optic data, and (α, β) are the surrounding pixels that satisfy $||(x - \alpha, y - \beta)||_2 = 1$. Note that the enhanced filtering includes the second derivatives on $\pm 45^\circ$ besides x and y axis, which retain patterns in extra directions. The filtering result is masked to the original pattern by:

$$\mathbf{I}(x, y) = f(x, y) + c(\Delta f(x, y)), \tag{2}$$

where $\mathbf{I}(x, y)$ is the augmented intensity value for each pixel and c controls the strength of augmentation. The augmented pattern $\mathbf{I}(x, y)$ is then fed to the next stage.

B. Optical Motion Field Inference

We use Horn and Schunck's theory [18] to relate intensity variation of optic signals to optical motion and infer OMF via solving an energy minimization problem. This energy minimization problem is usually formulated based on two widely used constraints [18]: (1) *intensity constancy constraint*: the intensity remains unchanged when a pixel flows from one frame to another; (2) *smooth constraint*: OMF varies (piecewise) smoothly in the space. Considering illumination changing cases in real-world scenarios, we further introduce *gradient constraint* [20] as an optional replacement of *intensity constancy constraint*, which enables switching between stable illumination case and illumination alteration case.

The two-dimensional OMF is defined as $\mathbf{w}(\mathbf{m}) = (u(\mathbf{m}), v(\mathbf{m}))^T$, where $\mathbf{m} = (x, y)$ represents the pixel in the intensity pattern. Based on *intensity constancy constraint*, we set a data penalty function as: $\Gamma_{\mathbf{I}}(\mathbf{w}, \mathbf{m}) = \| \mathbf{I_2}(\mathbf{m} + \mathbf{w}) - \mathbf{I_1}(\mathbf{m}) \|$, where $\| \cdot \|$ represents $L1$ norm. As for *gradient constraint*, the penalty function is given by: $\Gamma_{\nabla \mathbf{I}}(\mathbf{w}, \mathbf{m}) = \kappa \| \nabla \mathbf{I_2}(\mathbf{m} + \mathbf{w}) - \nabla \mathbf{I_1}(\mathbf{m}) \|$, where ∇ is the discrete approximation of gradient operator, and κ is a weight coefficient. To choose more fitting constraint for different cases, we develop a selective data function as:

$$E_D(\mathbf{w}) = \sum_{\mathbf{m}} [\eta(\mathbf{m})\Gamma_{\mathbf{I}}(\mathbf{w}, \mathbf{m}) + (1 - \eta(\mathbf{m}))\Gamma_{\nabla \mathbf{I}}(\mathbf{w}, \mathbf{m})], \tag{3}$$

where $\eta(\mathbf{m})$ is a binary weight map for switching between two terms, denoted as $\eta(\mathbf{m}) : \mathbb{Z}^2 \to \{0, 1\}$.

For *smooth constraint*, we define smoothness penalty function as:

$$E_S(\mathbf{w}) = \sum_{\mathbf{m}} \| \nabla \mathbf{w}(\mathbf{m}) \| . \tag{4}$$

The whole energy function can be given by:

$$E(\mathbf{w}) = E_D(\mathbf{w}) + \lambda E_S(\mathbf{w}), \tag{5}$$

where λ is a weight balancing the two terms.

To solve this energy function minimization problem, we use the Mean Field (MF) approximation to avoid binary process [21], then the whole energy function is transformed to:

$$\hat{E}(\mathbf{w}) = \sum_{\mathbf{m}} -\frac{1}{\xi} \ln(e^{-\xi \Gamma_{\mathbf{I}}(\mathbf{w}, \mathbf{m})} + e^{-\xi \Gamma_{\nabla \mathbf{I}}(\mathbf{w}, \mathbf{m})}) + \lambda E_S(\mathbf{w}), \tag{6}$$

where we set $\xi \geq 1$ to guarantee that minimizing Eq. (6) has the same effect as minimizing Eq. (5) for OMF estimation. Since Eq. (6) is non-convex, we first estimate potential OMF candidates via scale-invariant feature transform (SIFT) [22] detection and matching, and then apply an optimal combination method [23] to select the optimal $\hat{\mathbf{w}}(\mathbf{m})$ from candidates.

C. Optical Motion Field Disambiguation

Ambient occlusion is common in real-world scenarios that may result in ambiguous OMF estimation. Thereby, we develop an occlusion detection method to disambiguate the initial OMF $\hat{\mathbf{w}}(\mathbf{m})$. We first identify the occlusion pixel by examining if there exist multiple pixels mapping to the same position between two frames. The number of pixels in frame S_1 mapping to the same target position in frame S_2 is recorded

as $h(\mathbf{m})$. Then, we develop an occlusion confidence $\zeta(\mathbf{m})$ to weaken the influence of occluded pixels, given by:

$$\zeta(\mathbf{m}) = \begin{cases} 1 & h(\mathbf{m}) \leq 1, \\ 2 - h(\mathbf{m}) & 1 < h(\mathbf{m}) < 2 - \theta, \\ \theta & h(\mathbf{m}) \geq 2 - \theta, \end{cases} \tag{7}$$

where θ is set as 0.05. Thereby, the whole energy function will be updated as:

$$E(\mathbf{w}) = \zeta(\mathbf{m}) E_D(\mathbf{w}) + \lambda E_S(\mathbf{w}). \tag{8}$$

The minimization problem of Eq. (8) can be solved using continuous optimization via iteratively updating $\hat{\eta}(\mathbf{m})$ and $\mathbf{w}(\mathbf{m})$, similar to [24]. $\hat{\eta}(\mathbf{m})$ is the MF-approximation of binary weight $\eta(\mathbf{m})$, given by [24]: $\hat{\eta}(\mathbf{m}) = \frac{1}{1+e^{\xi(\Gamma_{\mathbf{I}}(\mathbf{w},\mathbf{m}) - \Gamma_{\nabla\mathbf{I}}(\mathbf{w},\mathbf{m}))}}$. The accurate $\mathbf{w}(\mathbf{m})$ can be finally obtained when convergence.

IV. Optic-to-Inertia Transformer

In this section, we need to solve the inertial data reconstruction problem. The problem is formulated as optimizing a transformer function $\psi(\mathbf{w_t}) = a_t$, where $\mathbf{w_t}$ is OMF, a_t is inertial data from MEMS accelerometer, and subscript t indicates the time. However, finding a fixed closed-form ψ is an ill-posed problem [25] due to the diversity of wearable computing application scenarios. Considering wearable MEMS is attached to the human body to measure the motion acceleration, we manually designed ψ with prior human kinematics knowledge by neural network components, whose advantages have been proven in learning a complex mapping between data [26]. The prior knowledge, as shown in Fig. 3, is that human motion is constrained by the body posture and the status of different body areas [27]. For example, the angle, as well as the muscle strength of the body joint constrain the bending and twisting. Also, the body skeleton keeps various anatomical features at fixed distances from one another. These skeleton constraints collapsed in a 2-D optical motion field limiting the inertia of connected body areas. With this knowledge, we then elaborate on the mechanism of ψ in the following subsections.

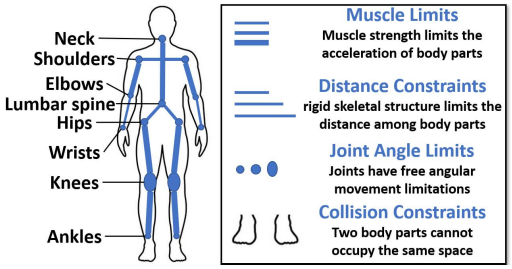


Fig. 3: Kinematics constraints applied to wearable MEMS.

A. MEMS Area-of-Interest Tracking

To understand and leverage the above constraints, we are motivated to track different MEMS area-of-interest and extract their motion-related intimations in OMF. Moreover, as the body moves and the posture changes, it is crucial and challenging to continuously track these area-of-interests in dynamic OMF, in which way we can narrow the potential inertial data because the previous status of the human body will influence the next status [28].

In *VisualAcc*, we pose temporal convolutional module. Specifically, the first convolutional module in Fig. 4 takes $\mathcal{T}$ adjacent motion fields $\mathbf{W} = \{\mathbf{w}_t | t = 0, 1, \cdots, \mathcal{T}\}$ as input and calculates the n-th feature map $\mathcal{F}$ as:

$$\mathcal{F}(x, y, t, n) = \sum_{j_1=0}^{J_1} \sum_{j_2=0}^{J_2} \sum_{j_3=0}^{J_3} K_n(j_1, j_2, j_3) \mathbf{W}(x - j_1, y - j_2, t - j_3), \tag{9}$$

where J_1, J_2, and J_3 are the sizes of convolution kernel in height, width, and time, respectively, N is the number of kernels and K_n is the n-th convolution kernel whose parameters can be determined by optimization. Feature maps can be regarded as the motion intimations of area-of-interest. Note that the following convolutional modules take the output of previous one instead of the original OMF and generate more feature maps. This multiple-convolution design increases the number of feature maps, which facilitates extracting fine-grained time-varying motion intimations [29]. Moreover, it enables the multi-scale capture of different area-of-interest because the later module executes convolution on the spatial down-sampled feature maps, i.e., larger-scale convolution. Knowing these human motion intimations, kinematics constraint cognition for MEMS is then performed.

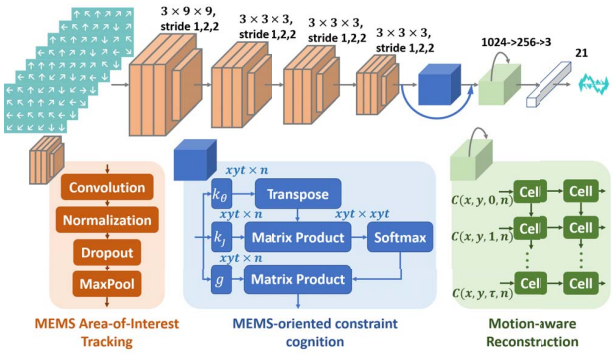


Fig. 4: Optic-to-Inertia Transformer Design.

B. MEMS-oriented Constraint Cognition

Although motion intimations of different area-of-interest are tracked by convolution, leveraging those intimations to cognize the kinematics constraints applied to wearable MEMS are still challenging because the convolution operations are local. Particularly, Eq. (9) shows that the unit in feature maps is the weighted sum of its neighborhoods indicated by convolution kernel size. Those neighbor units usually belong to the same body area.

To cognize the collapsed constraints from those local feature maps, we use the following non-local calculation [30] to take

the intimation of each body area into consideration:

$$\Psi(x,y,t,n) = \sum_{\hat{\mathcal{F}} \in \forall \mathcal{F}} exp(\mathcal{F}(x,y,t,n)^T K_\theta^T K_\jmath \hat{\mathcal{F}}) g(\hat{\mathcal{F}}), \tag{10}$$

where g is a linear transformation and $K_\jmath, K_\theta$ are two embedding kernels. The above equation captures the similarity between every two feature map units, i.e., motion intimations, in embedding spaces. By optimizing these two kernels, we are able to find embedding spaces where the similarity between two feature map units represents the constraints. Moreover, the constraints in embedding spaces are not only spatial but also temporary because the $\hat{\mathcal{F}}$ in Eq. (10) can be $\mathcal{F}(x,y,t{+}1,n)$ or $\mathcal{F}(x,y,t{-}1,n)$, which are two examples providing intimations from future and past.

C. Motion-Aware Inertia Reconstruction

In this part, we reconstruct the MEMS inertial data based on the combination of intimations from area-of-interest and the constraints among them. The combination can be formulated as the following residual connection [31]:

$$C(x,y,t,n) = K_z \Psi(x,y,t,n) + \mathcal{F}(x,y,t,n), \tag{11}$$

where K_z is a linear embedding kernel. Typically, a fully connected network is enough to decode the extracted intimations to 1-D or 2-D data. However, C is a temporo-spatial combination, where temporary motion intimations are extracted by convolution and non-local operations along the time dimension. Therefore, Long-short Term Memory [32] design is adopted in *VisualAcc* to perform motion-aware reconstruction as

$$\text{FoldR}(CELL, [0, C(x,y,0,n), \cdots, C(x,y,\mathcal{T},n)]), \tag{12}$$

where $CELL$ is a function closure that takes (hidden_state, cell_state) and a combination C, and returns (hidden_state', cell_state'). FoldR is a high-order function that can recursively apply $CELL$ on the combination list.

V. Implementation and Benchmark

In this section, we introduce the *VisualAcc* implementation and the benchmark preparation for evaluation and performance metrics in wearable computing.

A. System Implementation

Software: We augment intensity pattern (see Section III-A) via a 3×3 filter, which can be formulated as $[[-2,-2,-2]^T, [-2,32,-2]^T, [-2,-2,-2]^T]$. We build a six-layer pyramid in an intensity pattern and estimate OMF in each layer of the pyramid. Once the OMF at level l is obtained, it is propagated to the next level $l+1$, and becomes an OMF candidate together with other candidates estimated using SIFT feature detection and matching. After the OMF disambiguation process, the finally obtained OMF is represented as an RGB frame through the Hue-Saturation-Value color space using a similar approach introduced in [33]. The Optic-to-Inertia Transformer is developed following the parameter setting illustrated in Fig. 4 and the dropout rate is set as 50%. We set a cost function as Mean-Square Error (MSE) and adopt Stochastic Gradient Descent (SGD) as the optimizer. The batch size we select is 64. The optimization is performed on a workstation with Intel Xeon CPU E5-1620 and NVidia TITAN Xp GPUs.

Data Preparation: To evaluate the system performance, we leverage the data from two distinct resources: 1) publicly available database, i.e., Berkeley Multimodal Human Action Database (MHAD) [9] in a controlled environment for training and testing the *VisualAcc*'s overall performance; 2) self-collected data from a real-world environment for studying the *VisualAcc*'s reliability (see Section VII-B in detail). Both motion-caption sensors and wearable sensors are deployed in validating the *VisualAcc* performance. In the MHAD dataset, human motion is captured with two cameras. In the same while, participants wear two wearable IMU units on the left wrist and left hip, respectively. Three-axis accelerometer sensors are integrated into the IMU units, working with the sample rate at 30Hz. The clock in the camera and Shimmer wearable sensors are synchronized for ease of comparison. Nine subjects are enrolled for training. Each subject performs four motion actions, and each action is repeated five times, including jumping in place, jumping jacks, bending - hands up all the way down, and sitting down then standing up (hereafter A1, A2, A3, and A9, respectively).

Data Partition: We centrally crop the raw image and then resize it to 256×256 pixels. We then leverage the nearest interpolation to pair every frame image with the nearest acceleration data record in time stamps, which forms 43097 image-acceleration pairs for the left wrist and 42950 image-acceleration pairs for the left hip. We group every seven adjacent pairs into a sample, and the overlapping between samples is six pairs. Since every subject performs each activity five times, we leverage the first four times for training, and the last time for the test, resulting in 33558 training samples and 8459 test samples for the left wrist, 33558 training samples and 8318 test samples for the left hip.

B. Performance Metrics

We are interested in measuring the *fidelity*, *integrity*, and *authenticity* of the reconstructed inertial data by *VisualAcc*, and investigating if these data can truly boost wearable computing applications. Therefore, our evaluation is from three perspectives:

1) Human Motion Data Fidelity. We evaluate the *fidelity* of the reconstructed acceleration data compared to the accelerometer readings (hereafter, ground truth) using two metrics. *Cross-correlation (XCorr)* measures the similarity between two signals with the following formulation:

$$(a \star a')[\tau] = \sum_{j=-\infty}^{\infty} a[j] a'[j+\tau], \tag{13}$$

where τ represents the lag, a and a' are ground truth and reconstructed acceleration, respectively. In evaluation, we compare XCorr $(a \star a')$ with the Auto Correlation of ground

truth ($a \star a$, hereafter, ACorr). The high-fidelity reconstructed acceleration would give a close XCorr to ACorr. Moreover, we adopt *Pearson product-moment correlation coefficient (PCC)*, which is a measure of linear correlation between two signals, to quantitatively evaluate the reconstructed acceleration:

$$r = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(a'_i - \bar{a'})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2}\sqrt{\sum_{i=1}^{n}(a'_i - \bar{a'})^2}}, \qquad (14)$$

where $\bar{\cdot}$ is the mean function. A high-fidelity reconstructed acceleration would give r close to $+1$.

2) Locomotion Feature Integrity. Feature engineering is widely adopted for analyzing inertial data since features describe the statistical characteristics of motions and decrease the demand for computational resources for many classification-based tasks, e.g., activity recognition. Therefore, we expect the reconstructed acceleration to have integral locomotion characteristics as the ground truth. Particularly, we adopt *Skewness*, *Kurtosis*, and *Interquartile Range (IQR)*, which are practical locomotion features widely used in wearable motion sensing [34], as the locomotion feature integrity metrics from the kinematics perspective.

3) Activity Recognition Authenticity. The activity recognition accuracy (ARA) is to further evaluate the *authenticity* of reconstructed acceleration from the application perspective. Specifically, we train Random Forest models RF_{gnd} and RF_{rec} on the ground truth and reconstructed data using the aforementioned locomotion features, respectively. Two models share the same setting: the number of contained decision trees is 200; GINI function is used as a criterion; the random seed is fixed as 2. We test two models on another independent ground truth data set.

VI. Performance Evaluation

In this section, we evaluate the overall performance of *VisualAcc* for left wrist and left hip, respectively.

A. Accelerometer Raw Data Fidelity

The reconstructed acceleration and the corresponding accelerometer readings of bending action are shown in Fig. 5. The reconstructed data exhibits the same periodicity and period length compared with the ground truth. In each period, the local maximum and minimum of reconstructed acceleration are close to that of the ground truth.

We further evaluate the fidelity of raw-acceleration data via XCorr and PCC. As shown in Fig. 6, The XCorr curve and ACorr curve are quite close. This indicates that the reconstructed acceleration contains the same periodicity as the ground truth. Also, we observe there is always a peak at lag=0 in these curves, and this peak in the XCorr curve is nearly close to that in the ACorr curve. These results demonstrate that the reconstructed acceleration confirms the same temporary distribution as the ground truth. In addition, the PCC values are all over 0.8, which shows a significant correlation between the reconstructed data and the ground truth. In conclusion, *VisualAcc* can reconstruct high-fidelity acceleration data.

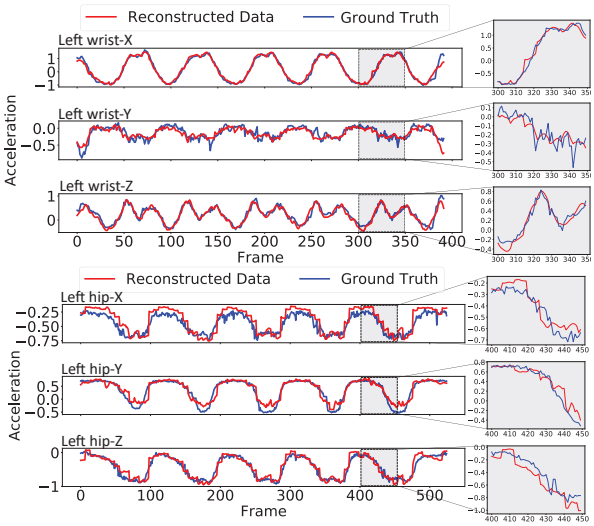


Fig. 5: The comparison between the acceleration data reconstructed by *VisualAcc* and that collected from accelerometer (ground truth) of bending action.

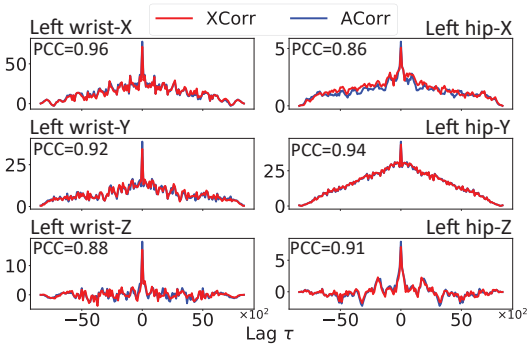


Fig. 6: Evaluation of raw data fidelity.

B. Locomotion Feature Integrity

To study the locomotion feature-level integrity of the reconstructed data, we compare the skewness, kurtosis, and IQR of the reconstructed acceleration with those of the ground truth. For each action, we first segment the ground truth and reconstructed acceleration data with a sliding window (length=24), and then calculate these three features in each window. The average values of these features over different actions are illustrated in Fig. 7. We observe that the values of these features vary from action to action on reconstructed acceleration, which indicates the reconstructed acceleration carries distinct locomotion characteristics. Moreover, the feature values of the reconstructed data are similar to that of the ground truth on each action, which shows that the reconstructed acceleration reserves most of the locomotion characteristics existing in the ground truth. Thus, we conclude *VisualAcc* can reconstruct integral data from the kinematics view.

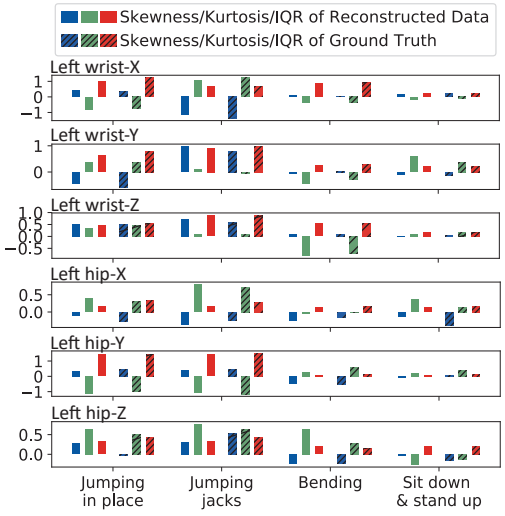
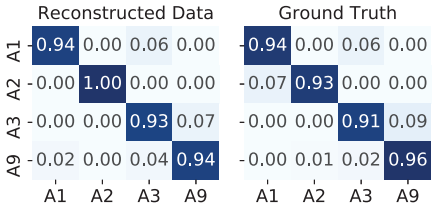
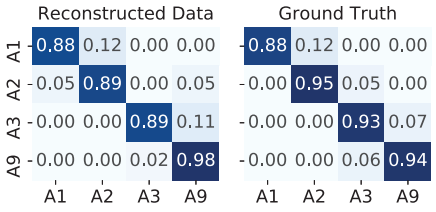


Fig. 7: The comparison of locomotion features among different activities.



(a) Left wrist, mean accuracy are 94.1% and 93.6% for RF_{rec} (left) and RF_{gnd} (right)



(b) Left hip, mean accuracy are 92.6% and 93.1% for RF_{rec} (left) and RF_{gnd} (right)

Fig. 8: The comparison of activity recognition performance.

C. Authenticity in Activity Recognition

We leverage the aforementioned locomotion features to evaluate ARA via RF_{gnd} and RF_{rec} models with the same parameter settings (see Section V-B). RF_{gnd} is trained on 439 samples (70%) from ground truth, and RF_{rec} is trained on the corresponding 439 samples from the reconstructed acceleration. We use the remaining 188 samples from the ground truth as an independent test set. The recognition performance of both models are shown in the Fig. 8. RF_{rec} achieves the recognition accuracy of over 93% on the left wrist and 88% on the left hip for all actions. Also, the recognition accuracy of RF_{rec} on each activity is close to those of RF_{gnd}. These results indicate that *VisualAcc* can reconstruct authentic inertial data at the application level.

D. Model Understanding

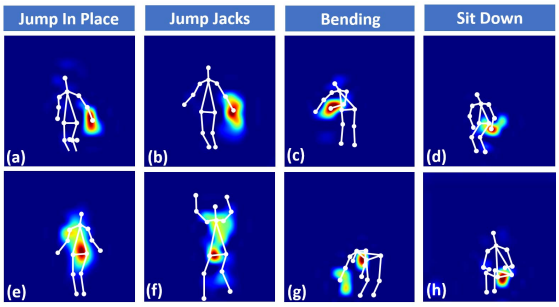


Fig. 9: Visualization of the body areas' contribution to the reconstruction. (a-d): left wrist; (e-h): left hip.

The models of photometric effect and human locomotion dominate the data reconstruction in *VisualAcc*. In this part, we validate this domination from the physical perspective. The key idea is to evaluate the contribution of different intensity pattern regions (i.e., *VisualAcc*'s focus) to the inertial data reconstruction. We leverage Gradient-weighted Class Activation Mapping (Grad-CAM) method [35] to evaluate the weight of each body area's intensity pattern and mask the weights to the original frame as a heatmap, shown in Fig. 9. For the reconstructed data of left wrist, Fig. 9 (a-d) show that the *VisualAcc* correctly extracts motion-related intimations mainly from the left forearm region, where the left wrist (red region) contributes most. Besides, Fig. 9 (d) illustrates that the left tibia region contributes a little to the reconstruction when the subject sits down. This is consistent with the observation that the subject's left wrist applies a force to the left thigh when sitting down, and the force propagates to the tibia. As for the reconstructed data of left hip, we observe that *VisualAcc* focuses on multiple body regions in Fig. 9 (e-h). The reason is that the hip is a point connecting pelvis and thigh bones, which propagates the force applied on the hip to both the upper and lower parts of the human body. Evidently as exhibited, the left hip area and lower back area (red region) make the greatest contribution whatever actions the subjects perform. In Fig. 9 (g), the left arm region (yellow region) is the second focus of *VisualAcc*. It is because the arm's movement can constrain the forward-and-backward motion of the left hip by changing the body center of gravity, thereby affecting the acceleration estimation. Based on these observations and analysis, we conclude that *VisualAcc* has learned the OMF to an inertia mapping function. The above observation and analysis validate the domination of photometric effect and human locomotion models to *VisualAcc*.

VII. Reliability Study

In this section, we investigate the *VisualAcc*'s reliability against alien activities and different real-world environments.

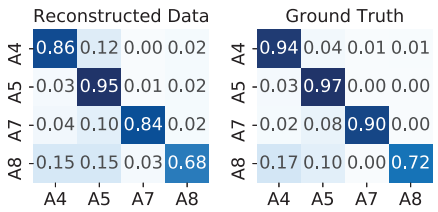


Fig. 10: The comparison of alien activity recognition performance.

A. Impact of Alien Actions

VisualAcc should be applicable with various human activities due to the wearable computing requirements. Therefore, it is necessary to investigate whether the *VisualAcc* can reconstruct reliable acceleration data of the target body area from the optic data of actions unseen during the optimization. Specifically, we select the optic data (from front view) and corresponding accelerometer readings of the left wrist when subjects 1-9 are performing the activities of punching (A4), waving hands (A5), clapping hands (A7), and throwing a ball (A8). The alien action recognition model is trained with the same locomotion features and parameter setting as mentioned in Section V. Specifically, we use 1044 samples from reconstructed acceleration for training RF_{rec}, and 1044 samples from corresponding ground truth for training RF_{gnd}. The remaining 448 samples from the ground truth are used to test these two recognition models. As shown in Fig. 10, we observe that the recognition accuracy of RF_{rec} on each alien action is close to that of RF_{gnd}. Also, RF_{rec} can achieve over 84% recognition accuracy over punching, waving hands, and clapping hands. For throwing a ball, the recognition accuracy of both RF_{rec} and RF_{gnd} are lower than 73%. The reason is that the left wrist keeps almost static or moves slightly, indicating that the left wrist is not ideal for ball throwing recognition. In conclusion, the *VisualAcc* exhibits reliable performance against alien actions.

B. Impact of Real-World Environment

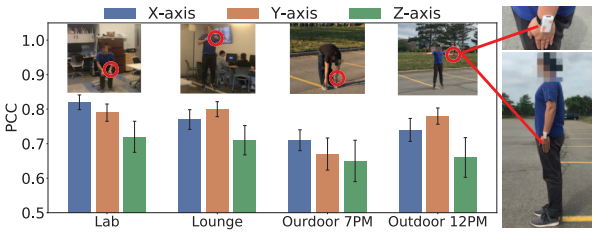


Fig. 11: The PCC between reconstructed data and the ground truth in four real-world scenarios. A shimmer is strapped to subject's left wrist for labeling the acceleration ground truth.

The real-world environment can introduce some daily-life interference or various illumination conditions. Thereby, we are curious about whether *VisualAcc* can be applied in real-world scenarios.

To collect data from the real-world environment, we leverage the camera on iPhone 6s (4.15 mm focal length) to record the video of activities from the front view of the subject, with a frame rate of 30 Hz and an image resolution of 1280×720 pixels. A Shimmer is strapped to the subject's left wrist and captures the acceleration data (ground truth) with the frequency of 30 Hz, as shown in Fig. 11. We recruit three volunteers to perform the activities jump in place, jump jacks, bending, and sit down and repeat five times in (1) a lab with some desks and serves, with 354 lux; (2) a student lounge with several people chatting, with 385 lux; (3) an empty parking lot at 12 pm, with 1826 lux; (4) an empty parking lot at 7 pm, with 903 lux. The data partition method is the same as we mentioned in Section V. Finally, we obtain approximately 14000 samples for each scenario.

For each action sequence, we calculate the PCC value between the reconstructed acceleration and the ground truth. Fig. 11 illustrates the average PCC values in every scenario. We observe that the average PCC values are over 0.7 when the experiments are conducted indoors. In the outdoor parking lot, the PCC values are lower than that of the indoor environment but still over 0.65. This is because the sharp shadow of subjects caused by sunlight moves as the subjects move, which has a similar shape to the subject to interfere with the transformer's tracking.

VIII. Case Study: Real-World IMU Inverse Human Kinematics and Dynamics

This case study explores the potential of applying the *VisualAcc* in real-world IMU-based inverse human kinematics and dynamics [36], a fine-grained framework that estimates human posture and force via multiple IMU sensors. Inverse kinematics is an established tool for human body analysis. For example, muscle fiber force can be estimated leveraging multiple IMU sensors, thereby assessing the rehabilitation from a movement disorder [17]. The main idea is to reconstruct acceleration data for multiple IMU sensors via *VisualAcc* from the recorded video of human motion. We solve the inverse kinematics and dynamics problems with the reconstructed IMU data and compare the results with those solved with the IMU sensor data. As the problems use multiple IMU data inputs, this case study can provide more insights of *VisualAcc* on the data authenticity in the multi-IMU application context.

A. Method

Problem Statement: The IMU inverse human kinematics and dynamics are problems that take inputs from multiple IMU sensors and finally output the force of body parts of interest. In detail, IMU-based inverse human kinematics is to find a posture (*i.e.*, the angles of joints) of the human model so that the orientation error between the model IMU sensors and real-world IMU sensors can be minimized. It can be formulated as [37]:

$$\min_{q} \sum_{i \in IMUs} w_i \theta_i^2, \tag{15}$$

where q is the vector of generalized coordinates representing the model posture, w_i is the weight of i-th IMU sensor, and θ_i is the error of orientation of the i-th IMU sensor. The following inverse dynamics is to calculate the generalized force χ of body parts via the posture q as [37]:

$$\chi = M(q)q'' + C(q, q') + G(q), \tag{16}$$

where $M(\cdot)$ is the mass matrix, $C(\cdot)$ is the vector of Coriolis and centrifugal forces, $G(\cdot)$ is the vector of G-forces, q'' and q' are the generalized velocities and accelerations.

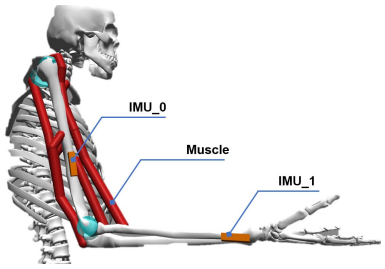


Fig. 12: The human kinematic and dynamics model and the placement of the two IMU sensors used in case study.

Data Preparation: The data we prepare is of right arm motion. Three participants are recruited in the case study. Two IMU sensors are attached to the right wrist and right upper arm. We collect data from both accelerometers and gyroscopes with a sampling rate of 30 Hz as the ground truth. A camera (resolution is 1024×576) is placed in the right front of the subject to provide the input for *VisualAcc*. Data collection is done by a laptop so that the IMU data and camera data are synchronized. During the data preparation, the three participants perform random arm movement in the sagittal plane so that more arm postures are covered. We collect five minutes of data for each participant.

IMU Data Reconstruction: We start with the model we used in Sec. VI. The data of the first four minutes are used for training and transferring this pre-trained model to the right arm IMU data generation. Then, we feed the camera data of the last minute to the transferred *VisualAcc* to reconstruct the IMU data.

Metrics: We solve the inverse kinematic and dynamics problems using the reconstructed and ground truth IMU data, following the three setups detailed in TABLE I. We adopt OpenSim [38] as our solver. Three dynamics metrics are used, which are the Mean square error of normalized muscle fiber length MSE_{NMFL} and fiber force MSE_{FF}, respectively.

	Setup A	Setup B	Ground Truth
Accelerometer_0 Data	Reconstructed	Reconstructed	Ground Truth
Accelerometer_1 Data	Ground Truth	Reconstructed	Ground Truth
Gyroscope Data	Ground Truth	Ground Truth	Ground Truth

TABLE I: The setups to evaluate *VisualAcc* reconstruction's authenticity.

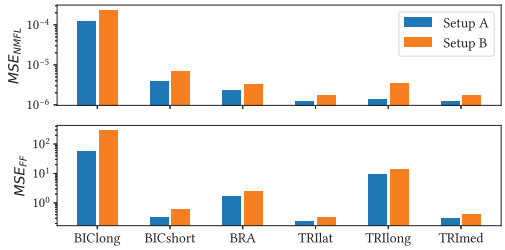


Fig. 13: The performance of inverse human kinematic and dynamics facilitated by *VisualAcc* reconstructed IMU data.

B. Performance Results

We report the performance results on six different muscles, which are Biceps brachii long head (BIClong), biceps brachii short head (BICshort), brachioradialis (BRA), triceps brachii lateralis (TRIlat), triceps brachii long head (TRIlong), triceps brachii medialis (TRImed), respectively. Fig. 13 shows that the muscle fiber length (typical value is less than 2) is less impacted by the reconstructed IMU data compared with fiber force (typical value is around 500). In addition, the MSE_{NMFL} and MSE_{FF} of muscle TRIlat and TRImed are relatively less than that of other muscles. These observations reveal that the data reconstructed by *VisualAcc* could be better on specific applications related to triceps filer length. We also observe that as the use of reconstructed data increases, both error metrics also increase. However, the human kinematics and dynamics problems are usually over-determined [39], which implies that if more IMU sensors are reconstructed by *VisualAcc*, the errors might be well controlled.

IX. Related Work

Light-based Human Sensing: Light-based human sensing has a long history. Zhou *et al.* proposed to apply light sensing (photodiode) in hand gesture reconstruction [40] and human skeleton posture sensing [41], [42]. Ma *et al.* [43] proposed to use solar panel as a light sensor for gesture recognition. In the biomedical area, light-based respiration [44] and heart rate [45] monitoring (Photoplethysmographic, PPG) is widely adopted. Some researchers also explored human motion sensing with infrared light [46], [47]. Different from those works, *VisualAcc* is the first photometric effect-based framework to sense complex human motion and reconstruct fine-grained inertial data, with the models of photometric effect and human locomotion.

Data Boosting and Generation: Data boosting and generation is an emerging topic in wearable computing. Many works generate new data by transferring harvested information in prior studies. For example, motion studies [48], [49] tried to generate full-body motion data via computer models of the muscle and joint. However, due to the complexity and individuality of the human body, several practices [50], [51] have approved that generated data usually far differ from real-world data. Other works leverage the fast development of neural networks for data boosting and generation. For example,

generative adversarial models (GAN) [11] as one type of neural network that can create new data via approximating the real data distribution are the primary tools used in this field. However, GAN-created data is usually biased because the discriminator that helps improve data quality does not carry any kinematics knowledge [52]. Different from these works, *VisualAcc* bridges data-rich and device-free camera repositories to wearable computing, boosting authentic data for wearable computing.

X. Conclusion and Future work

In this paper, we propose *VisualAcc*, an optic-to-inertia framework to facilitate effort-free wearable computing for human activities of daily life. We explore the principled connection between harvested light-intensity data and authentic wearable motion data, with the models of photometric effect and human locomotion. Based on our exploration, we develop a photometric effect-based human motion interrogation module to extract high-quality optical motion field from light intensity variation. Then, we design an optic-to-inertia transformer with prior human kinematics knowledge to reconstruct the inertial data. This transformer consists of three modules: MEMS area-of-interest tracking, MEMS-oriented constraint cognition, and motion-aware reconstruction. The extensive evaluation and real-world case study show the potential of *VisualAcc* to empower effort-free wearable computing.

VisualAcc extends the scope of wearable computing application to more scenarios. For example, *VisualAcc* can be applied to generate more inertial data for researchers to analyze workers' awkward postures, improving occupational safety. Besides, *VisualAcc* is able to provide in-situ analytics for athletes during professional sports games, where the wearable devices are not easy to deploy. In future work, we plan to evaluate *VisualAcc* using gyroscope data and improve its performance in the cases of deep shadow.

ACKNOWLEDGMENTS

We thank all anonymous reviewers for their insightful comments on this paper. This work was supported by the National Science Foundation under grant No. ECCS-2028872, CNS-1718375.

References

[1] A. Fleury, M. Vacher, and N. Noury, "Svm-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 2, pp. 274–283, 2009.

[2] P. Kodeswaran, R. Kokku, M. Mallick, and S. Sen, "Demultiplexing activities of daily living in iot enabled smarthomes," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[3] C. Debes, A. Merentitis, S. Sukhanov, M. Niessen, N. Frangiadakis, and A. Bauer, "Monitoring activities of daily living in smart homes: Understanding human behavior," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 81–94, 2016.

[4] S. A. Shah and F. Fioranelli, "Rf sensing technologies for assisted daily living in healthcare: A comprehensive review," *IEEE Aerospace and Electronic Systems Magazine*, vol. 34, no. 11, pp. 26–44, 2019.

[5] S. M. Gerber, R. M. Müri, U. P. Mosimann, T. Nef, and P. Urwyler, "Virtual reality for activities of daily living training in neurorehabilitation: a usability and feasibility study in healthy participants," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 1–4.

[6] J. Kim, S. Yang, and M. Gerla, "Stroketrack: wireless inertial motion tracking of human arms for stroke telerehabilitation," in *Proceedings of the First ACM Workshop on Mobile Systems, Applications, and Services for Healthcare*, 2011, pp. 1–6.

[7] Z. Jin, J. Oresko, S. Huang, and A. C. Cheng, "Hearttogo: a personalized medicine technology for cardiovascular disease prevention and detection," in *2009 IEEE/NIH Life Science Systems and Applications Workshop*. IEEE, 2009, pp. 80–83.

[8] "Motion Sensor Market by Motion Technology, Application, and Geography - Global Forecast to 2025," Tech. Rep. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/Motion-Sensor-Market-614.html

[9] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley mhad: A comprehensive multimodal human action database," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 2013, pp. 53–60.

[10] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, D. J. Cook, G. Blohm, and N. F. Troje, "Movi: A large multipurpose motion and video dataset," *arXiv preprint arXiv:2003.01888*, 2020.

[11] J. Wang, Y. Chen, Y. Gu, Y. Xiao, and H. Pan, "Sensorygans: An effective generative adversarial framework for sensor-based human activity recognition," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[12] K. C. Madathil, A. J. Rivera-Rodriguez, J. S. Greenstein, and A. K. Gramopadhye, "Healthcare information on youtube: a systematic review," *Health informatics journal*, vol. 21, no. 3, pp. 173–194, 2015.

[13] Y. Wang, "Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of tiktok (douyin)," *Computers in Human Behavior*, vol. 110, p. 106373, 2020.

[14] H. Kwon, C. Tong, H. Haresamudram, Y. Gao, G. D. Abowd, N. D. Lane, and T. Ploetz, "Imutube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–29, 2020.

[15] S. Zhang and N. Alshurafa, "Deep generative cross-modal on-body accelerometer data synthesis from videos," in *Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp-ISWC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 223–227.

[16] V. F. Rey, P. Hevesi, O. Kovalenko, and P. Lukowicz, "Let there be imu data: Generating training data for wearable, motion sensor based activity recognition from monocular rgb videos," in *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2019, pp. 699–708.

[17] V. B. Semwal, N. Gaud, P. Lalwani, V. Bijalwan, and A. K. Alok, "Pattern identification of different human joints for different human walking styles using inertial measurement unit (imu) sensor," *Artificial Intelligence Review*, pp. 1–21, 2021.

[18] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[19] A. Verri and T. Poggio, "Motion field and optical flow: qualitative properties," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 5, pp. 490–498, May 1989.

[20] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *European conference on computer vision*. Springer, 2004, pp. 25–36.

[21] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from mrfs: Surface reconstruction," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 5, pp. 401–412, 1991.

[22] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using sift features and mean shift," *Computer vision and image understanding*, vol. 113, no. 3, pp. 345–352, 2009.

[23] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[24] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 1293–1300.

[25] A. M. Tekalp and A. M. Tekalp, *Digital video processing*. Prentice Hall PTR Upper Saddle river, NJ, 1995, vol. 1.
[26] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3, p. 100004, 2019.
[27] J. O'rourke and N. I. Badler, "Model-based image analysis of human motion using constraint propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 6, pp. 522–536, 1980.
[28] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," in *Proceedings 1992 IEEE Computer Society conference on computer vision and pattern recognition*. IEEE, 1992, pp. 379–385.
[29] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.
[30] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[33] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.
[34] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *Proceedings of the 6th International Conference on Body Area Networks*. ICST, 2011, pp. 92–98.
[35] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
[36] D. Tolani and N. I. Badler, "Real-time inverse kinematics of the human arm," *Presence: Teleoperators & Virtual Environments*, vol. 5, no. 4, pp. 393–401, 1996.
[37] S. Chiaverini, O. Egeland, and R. Kanestrom, "Achieving user-defined accuracy with damped least-squares inverse kinematics," in *Fifth International Conference on Advanced Robotics' Robots in Unstructured Environments*. IEEE, 1991, pp. 672–677.
[38] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "Opensim: open-source software to create and analyze dynamic simulations of movement," *IEEE transactions on biomedical engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.
[39] V. Cahouët, M. Luc, and A. David, "Static optimal estimation of joint accelerations for inverse dynamics problem solution," *Journal of biomechanics*, vol. 35, no. 11, pp. 1507–1513, 2002.
[40] T. Li, X. Xiong, Y. Xie, G. Hito, X.-D. Yang, and X. Zhou, "Reconstructing hand poses using visible light," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 71:1–71:20, Sep. 2017. [Online]. Available: http://doi.acm.org/10.1145/3130937
[41] T. Li, Q. Liu, and X. Zhou, "Practical human sensing in the light," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '16. New York, NY, USA: ACM, 2016, pp. 71–84.
[42] C. An, T. Li, Z. Tian, A. T. Campbell, and X. Zhou, "Visible light knows who you are," in *Proceedings of the 2Nd International Workshop on Visible Light Communications Systems*, ser. VLCS '15. New York, NY, USA: ACM, 2015, pp. 39–44.
[43] D. Ma, G. Lan, M. Hassan, W. Hu, M. B. Upama, A. Uddin, and M. Youssef, "Solargest: Ubiquitous and battery-free gesture recognition using solar cells," in *The 25th Annual International Conference on Mobile Computing and Networking*. New York, NY, USA: ACM, 2019, pp. 1–15.
[44] A. Johansson, "Neural network for photoplethysmographic respiratory rate monitoring," *Medical and Biological Engineering and Computing*, vol. 41, no. 3, pp. 242–248, 2003.
[45] K. H. Shelley, "Photoplethysmography: beyond the calculation of arterial oxygen saturation and heart rate," *Anesthesia & Analgesia*, vol. 105, no. 6, pp. S31–S36, 2007.
[46] D. Ryu, D. Um, P. Tanofsky, D. H. Koh, Y. S. Ryu, and S. Kang, "T-less: A novel touchless human-machine interface based on infrared proximity sensing," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 5220–5225.
[47] R. Wechsler, F. Weiß, and P. Dowling, "Eyecon–a motion sensing tool for creating interactive dance, music, and video projections," in *Proc. of the SSAISB Convention*. Citeseer, 2004.
[48] M. G. Pandy, "Computer modeling and simulation of human movement," *Annual review of biomedical engineering*, vol. 3, no. 1, pp. 245–273, 2001.
[49] Y.-C. Lin and M. G. Pandy, "Three-dimensional data-tracking dynamic optimization simulations of human locomotion generated by direct collocation," *Journal of biomechanics*, vol. 59, pp. 1–8, 2017.
[50] A. Sucerquia, J. López, and J. Vargas-Bonilla, "Sisfall: A fall and movement dataset," *Sensors*, vol. 17, no. 1, p. 198, 2017.
[51] M. Jiang, H. Shang, Z. Wang, H. Li, and Y. Wang, "A method to deal with installation errors of wearable accelerometers for human activity recognition," *Physiological measurement*, vol. 32, no. 3, p. 347, 2011.
[52] I. Durugkar, I. Gemp, and S. Mahadevan, "Generative multi-adversarial networks," *arXiv preprint arXiv:1611.01673*, 2016.