

# IMUTube: Converting Videos of Human Activity into Virtual IMU Data Streams

---

## An approach to make use of large video repositories for behavior analysis with wearables

IMUTube is an automated processing pipeline that converts videos of human activity into virtual streams of inertial measurement unit (IMU) data. Georgia Tech's innovation is designed to improve the performance of a variety of models on known human activity recognition (HAR) datasets. It can be used to apply existing video repositories to on-body, wearable sensor-based HAR for a wide range of behavioral analysis applications.

IMUTube provides four essential functions:

1. It applies standard pose tracking and 3D scene understanding techniques to estimate full 3D human motion from a video segment that captures a target activity.
2. It translates the visual tracking information into virtual motion sensors that are “placed” on dedicated body positions.
3. It adapts the virtual sensors' IMU data towards the target domain through distribution matching.
4. It derives activity recognizers from the generated virtual sensor data, which can be augmented with small amounts of real sensor data in some cases.

Human curation is needed only for selecting the appropriate activity video content; much of the IMUTube process is automated and applicable to existing videos, thanks to the use of off-the-shelf computer vision and graphics techniques.

With promising initial test results in three preliminary benchmark studies, IMUTube's full potential can be realized through a collective approach involving computer vision, signal processing, and HAR communities. Further work in this area could lead to massive increases in the volume of available real-movement data, making it possible to develop substantially more complex and robust HAR systems with broader scope than is currently possible in the field.

## Summary Bullets

- **Automated:** Automates wearable sensor data collection, which can in turn be used to train HAR models by simply giving them access to large-scale video repositories
- **Accurate:** Demonstrated recognition accuracy in some cases comparable to models trained only with non-virtual (“real”) sensor data in benchmark studies

- **Superior:** Outperformed models using real sensor data alone by adding only small amounts of real sensor data to the virtual sensor dataset in benchmark studies

### Solution Advantages

- **Automated:** Automates wearable sensor data collection, which can in turn be used to train HAR models by simply giving them access to large-scale video repositories
- **Accurate:** Demonstrated recognition accuracy in some cases comparable to models trained only with non-virtual (“real”) sensor data in benchmark studies
- **Superior:** Outperformed models using real sensor data alone by adding only small amounts of real sensor data to the virtual sensor dataset in benchmark studies
- **Economical:** Has the potential to replace costly, time-consuming, error-prone efforts involved in collecting non-virtual sensor data from people in real life
- **Revolutionary:** Offers the possibility of substantially increasing the volume of available movement data, catching up to advances made in complementary fields such as speech recognition and language processing

### Potential Commercial Applications

Georgia Tech’s innovation is broadly applicable to large-scale sensor video datasets for training wearable sensor-based HAR models. These systems are particularly useful for behavioral analysis for fields such as:

- User authentication
- Health care
- Fitness and wellness
- Security
- Other fields in which tracking of everyday activities may be beneficial

### Background and More Information

On-body sensor-based human activity recognition systems have lagged behind other fields in terms of large breakthroughs in recognition accuracy. In fields such as speech recognition, natural language processing, and computer vision, it is possible to collect huge amounts of labeled data—the key for deriving robust recognition models that strongly generalize across application boundaries. By contrast, collecting large-scale, labeled data sets in sensor-based HAR has been limited. Labeled data in HAR is scarce and hard to obtain, sensor data collection is expensive, and the annotation is time-consuming and sometimes even impossible for privacy or other practical reasons. As such, the scale of typical datasets remains small, covering only limited sets of activities. With further research and collaboration among HAR, signal processing, and computer vision communities, IMUTube may directly address these shortcomings of the field, leading to exponential increases in the amount of movement data available.

### Inventors

- Hyeokhyen Kwon  
PhD student - Georgia Tech School of Interactive Computing, College of Computing
- Dr. Gregory Abowd  
Former Distinguished Professor - Georgia Tech School of Interactive Computing
- Harish Haresamudram  
PhD Student - Georgia Tech School of Electrical and Computer Engineering
- Dr. Thomas Ploetz  
Associate Professor in College of Computing, and Faculty Member of Ubiquitous Computing Research

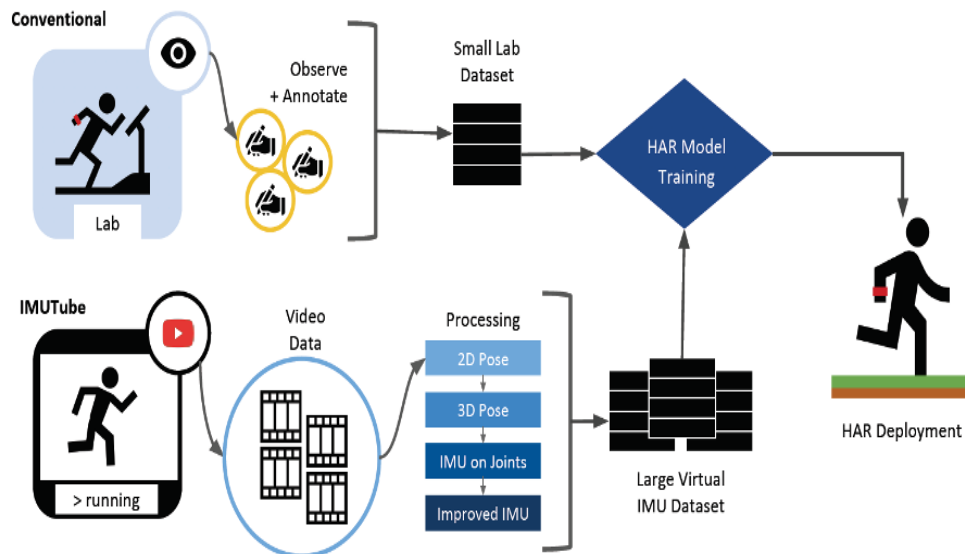
## IP Status

<p>Patent application has been filed</p>: US63/073009

## Publications

[IMUTube: Automatic Extraction of Virtual On-Body Accelerometry from Video for Human Activity Recognition](#), Proceedings of the Association for Computing Machinery (ACM) on Interactive, Mobile, Wearable and Ubiquitous Technologies - September 2020

## Images



IMUTube has the potential to replace the conventional data recording and annotation protocol (upper left) for developing sensor-based human activity recognition (HAR) systems. Georgia Tech’s system (bottom) uses existing, large-scale video repositories from which it generates virtual IMU data that are then used for training the HAR system.

Visit the Technology here:

[IMUTube: Converting Videos of Human Activity into Virtual IMU Data Streams](#)

---

<https://s3.sandbox.research.gatech.edu//print/pdf/node/3316>