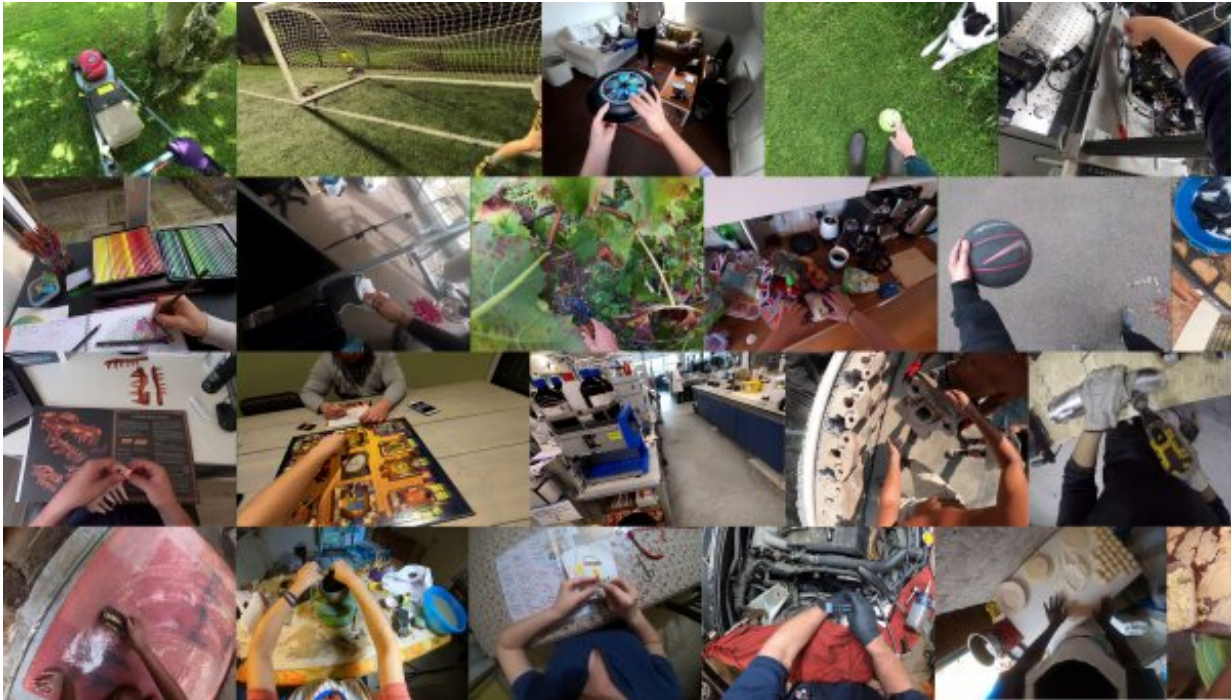# World's first massive egocentric dataset

October 14 2021

Egocentric 4D Live Perception (Ego4D) is a massive-scale dataset that compiles 3,025 hours of footage from the wearable cameras of 855 participants in nine countries. Credit: Ego4D

The University of Bristol is part of an international consortium of 13 universities in partnership with Facebook AI, that collaborated to advance egocentric perception. As a result of this initiative, we have built the world's largest egocentric dataset using off-the-shelf, head-mounted cameras.

Progress in the fields of artificial intelligence (AI) and augmented reality (AR) requires learning from the same data humans process to perceive the world. Our eyes allow us to explore places, understand people, manipulate objects and enjoy activities—from the mundane act of opening a door to the exciting interaction of a game of football with friends.

Egocentric 4D Live Perception (Ego4D) is a massive-scale dataset that compiles 3,025 hours of footage from the wearable cameras of 855 participants in nine countries: UK, India, Japan, Singapore, KSA, Colombia, Rwanda, Italy and the US. The data captures a wide range of activities from the 'egocentric' perspective—that is from the viewpoint of the person carrying out the activity. The University of Bristol is the only UK representative in this diverse and international effort, collecting 270 hours from 82 participants who captured footage of their chosen activities of daily living—such as practicing a musical instrument, gardening, grooming their pet or assembling furniture.

"In the not-too-distant future you could be wearing smart AR glasses that guide you through a recipe or how to fix your bike—they could even remind you where you left your keys," said Principal Investigator at the University of Bristol and Professor of Computer Vision, Dima Damen.

"However, for AI to move forward, it needs to understand the world, and the experiences within it. AI attempts to learn about all aspects of human intelligence through digesting data we perceive. To allow such automated learning, we have to capture and record our daily experiences 'through our eyes'. This is what Ego4D provides."

In addition to the captured footage, a suite of benchmarks is available for researchers. A benchmark is a problem definition along with manually collected labels to compare models. EGO4D benchmarks are related to understanding places, spaces, ongoing actions, upcoming

actions as well as social interactions.

"Our five new, challenging benchmarks provide a common objective for researchers to build fundamental research for real-world perception of visual and social contexts," says Professor Kristen Grauman from Facebook AI—technical lead.

The ambitious project was inspired by the University of Bristol's successful EPIC-KITCHENS dataset, which recorded the daily kitchen activities of participants in their homes and has been, until now, the largest dataset in egocentric computer vision. EPIC-KITCHENS has pioneered the approach of "pause and narrate" to give a near-accurate time of where each action takes place in the long and varied videos. Using this approach, the EGO4D consortium collected 2.5 million timestamped statements of ongoing actions in the video, which is crucial for benchmarking the collected data.

Ego4D is a huge and diverse dataset, with benchmarks, that will prove invaluable to researchers working in the fields of augmented reality, assistive technology and robotics. The datasets will be publicly available in November of this year for researchers who sign Ego4D's data use agreement.

More information: Website: ego4d-data.org/

Provided by University of Bristol