

สรุปการทำงาน week 5

1. โครงสร้างหลักของระบบ

1. Detection & Tracking

- ใช้ YOLO สองโมเดล (คน กับ กล้อง) ร่วมกับ ByteTrack
- เก็บ track ID และ bounding box ของคนกับกล้องต่อเฟรม

2. Pose Extraction

- ใช้ MediaPipe Pose เพื่อดึง landmarks 33 จุด
- เก็บเฉพาะ joint สำคัญ (ไหล่, ข้อศอก, ข้อมือ, สะโพก)

3. Feature Buffer

- เก็บ sequence ของ landmarks ย้อนหลัง SEQUENCE_LENGTH (15) เฟรม
- คำนวณความแปรปรวน (std) เพื่อแยก “idle” vs “moving” (rule-based)

4. Object–Person Matching

- ใช้ฟังก์ชัน point_in_bbox ตรวจสอบว่า key point ของกล้องตกในกรอบคนหรือไม่
- แมตช์ได้หลายคนต่อกล้องเดียว → เก็บ history ใน object_id_to_person_ids[obj_id] (deque maxlen=50)
- เลือกคนที่ถือกล้องจริง โดยดู freq ≥ 15 ใน deque → list carriers

2. กระบวนการจำแนกประเภท Carrying

เมื่อเจอคนที่ถือกล้อง (carrying) จะเรียกฟังก์ชัน `extract_features_from_skeleton(landmarks, track_id)` เพื่อตรวจรายละเอียด 6 กรณีหลัก:

Action	เงื่อนไขหลัก (Static + Smoothing)
push_forward	– มื่อยื่นไปข้างหน้า
pull_backward	– มื่อยู่หลังลำตัว $\text{avg_hand_x} < \text{avg_shoulder_x} - 0.03$ – ระดับมือระดับอก-เอว– $100^\circ \leq \text{avg_elbow} \leq 160^\circ$
carry_normal	– มื่อยู่ใกล้ตัว ($ \text{hand_x} - \text{shoulder_x} \leq 0.05$)– ระดับมือระหว่าง ($\text{shoulder_y} + 0.05, \text{hip_y} + 0.05$)– $100^\circ < \text{avg_elbow} \leq 165^\circ$

carry_heavy	– avg_elbow > 165° (เกือบเหยียดแขน)– avg_hand_y > hip_y – 0.03 (ถือแนบสะโพก)
carry_on_shoulder	– avg_hand_y < shoulder_y – 0.05 (มือสูงพาดบ่า)– avg_elbow < 70° (งอข้อศอก)
carry_together	– กล้องเดียวมี ≥ 2 carry IDs (freq ≥ 15)–

Default fallback: ถ้าไม่เข้าข้อใด จะยัง label เป็น "carrying" ทั่วไป

3. การจัดการ Flicker & Stability

- **Smoothing**
 - สำหรับ push_forward ใช้ deque เก็บ flag 5 เฟรมหลังสุด
 - ต้องมี ≥ 2 True ก่อนจึงตัดสินใจเป็น push_forward
- **Sequential priority (elif)**
 - เรียงลำดับตรวจเงื่อนไข: Push \rightarrow Pull \rightarrow Normal \rightarrow Heavy \rightarrow Shoulder
 - ป้องกัน overlap ของเงื่อนไข static

4. สรุปผลลัพธ์

- ระบบสามารถจับทั้ง คนเดียวถือ และ หลายคนช่วยกันถือ (carry_together)
- แยก 5 รูปแบบหลักได้โดยใช้ **Skeleton-based rules + history smoothing**