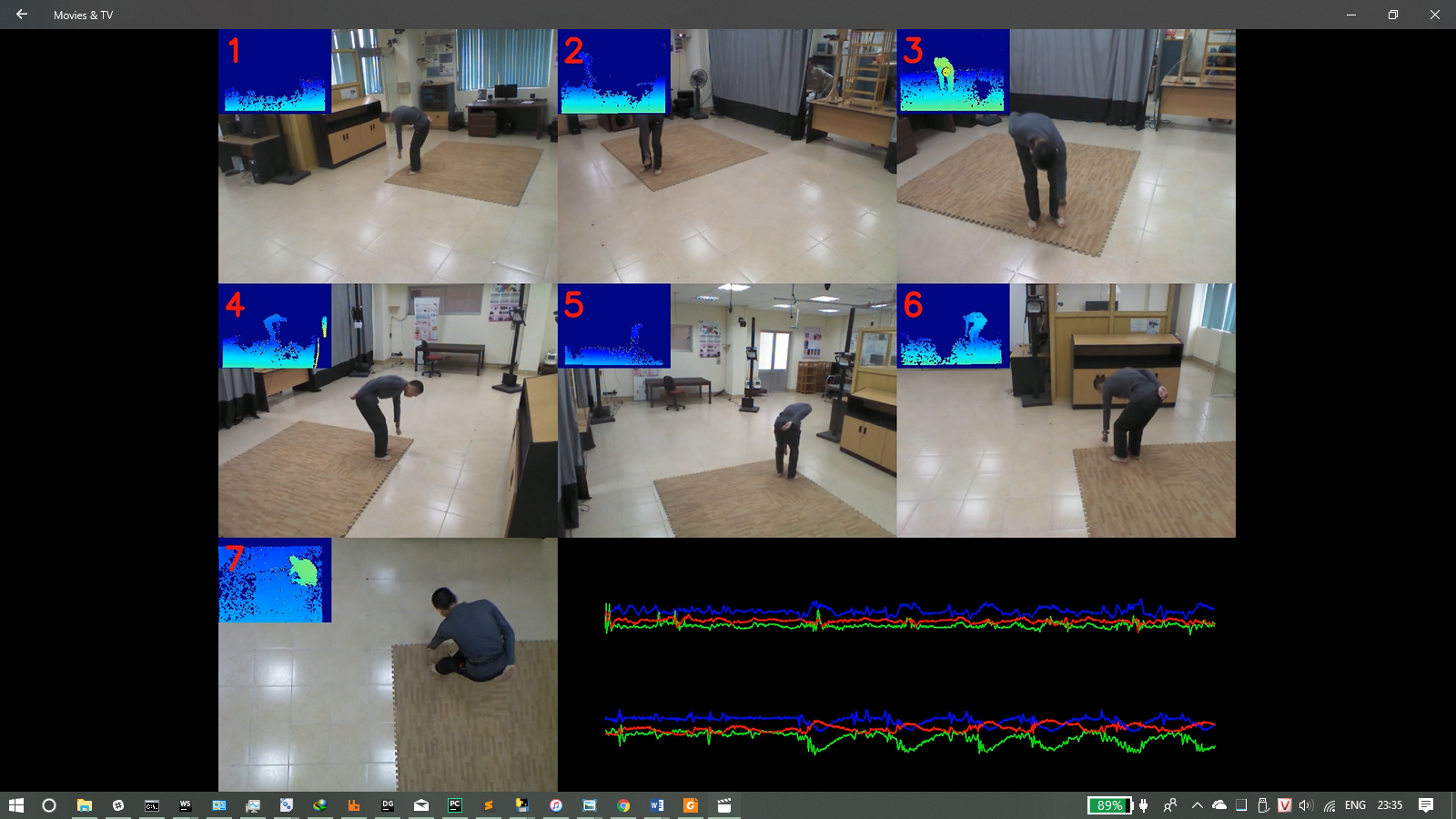
**KẾT QUẢ THỬ NGHIỆM PHÁT HIỆN NGƯỜI VỚI YOLO**

**Dữ liệu thử nghiệm:** Bộ cơ sở dữ liệu MICA được em sử dụng để đánh giá mô hình đề xuất. Bộ dữ liệu MICA bao gồm các video được ghi hình trong điều kiện trong phòng thí nghiệm, các camera được gắn cố định và ghi hình theo 7 góc khác nhau theo dõi hoạt động của người. Đây là bộ dữ liệu multiview lớn với số lượng người tham gia thu thập dữ liệu là 50 người, với 21 hành động khác nhau được sử dụng để đánh giá, thử nghiệm với các bài toàn phát hiện, nhận dạng hành động của người. Dữ liệu được thu thập bởi camera Kinect với độ phân giải 640x480, bao gồm dữ liệu ảnh màu RGB, dữ liệu ảnh độ sâu và dữ liệu cảm biến gia tốc gắn trên người.

Dưới đây là hình minh họa bộ cơ sở dữ liệu MICA thể hiện các dữ liệu ảnh màu, ảnh độ sâu và dữ liệu gia tốc.



Hình 1: Các dữ liệu trong bộ dữ liệu MICA

Để phục vụ bài toán phát hiện người, em đã trích xuất một phần trong bộ cơ sở dữ liệu MICA để phục vụ huấn luyện và thử nghiệm. Dữ liệu này bao gồm các ảnh màu RGB có độ phân giải 640x480 trích xuất từ các clip chứa ảnh màu.

Bộ dữ liệu này được trích xuất từ clip của 15 người/50 người trong bộ cơ sở dữ liệu, bao quát 21 hành động của người bao gồm:

1: Đi bộ theo các hướng

2: Chạy chậm

3: Nhảy tại chỗ

4: Cử động tay và đầu gối

5: Cúi xuống nhặt đồ bằng tay trái

6: Cúi xuống nhặt đồ bằng tay phải

7: Đi loạng choạng

8: Ngã về phía trước

9: Ngã về phía sau

10: Ngã về bên trái

11: Ngã về bên phải

12: Bò

13: Ngồi lên ghế rồi đứng lên

14: Di chuyển ghế sang vị trí khác

15: Ngã về bên trái khi đang ngồi trên ghế

16: Ngã về bên phải khi đang ngồi trên ghế

17: Ngồi lên giường rồi đứng lên

18: Nằm lên giường rồi ngồi dậy

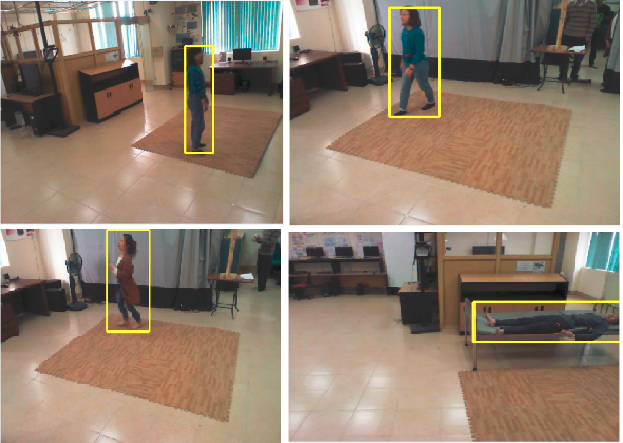
19: Ngã xuống về bên trái khi đang nằm trên giường

20: Nằm ngược lại, ngã xuống về bên phải khi đang nằm trên giường

Bộ dữ liệu sau đó được ground truth bằng tay và được chia làm 2 phần cho giai đoạn huấn luyện và kiểm thử phương pháp:

|  |  |  |
| --- | --- | --- |
|  | Huấn luyện | Thử nghiệm |
| Số lượng ảnh | 15000 | 4430 |

Dưới đây là một số hình ảnh minh họa dữ liệu trong bộ dữ liệu sử dụng trong đồ án này:



Hình 2. Minh họa cơ sở dữ liệu cho huấn luyện và thử nghiệm (bounding box màu vàng thể hiện ground truth của dữ liệu)

**Phương pháp đánh giá**

Việc đánh giá mô hình nhận dạng dựa trên các thước đo: Precision, Recall, Accuracy.

* Precision hay positive predictive value: là tỷ lệ của số lượng đầu vào được dự đoán chính xác (true positive) trong số được dự đoán là đúng,
* Recall hay sensitivity: là tỷ lệ của số lượng đầu được dự đoán chính xác (true positive) trong số được dự đoán là đúng.



Hình 3. Minh họa cách tính Recall và Precision

Bên cạnh đó em sử dụng Jaccard Index để đánh giá một phát hiện là đúng khi:

**Kết quả đạt được**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Kinect** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| **Số phát hiện đúng** | 681 | 587 | 665 | 688 | 683 | 581 | 424 |
| **Tổng số phát hiện** | 718 | 1276 | 975 | 779 | 733 | 638 | 455 |
| **Recall** | 0.9970 | 0.8519 | 0.9525 | 0.9666 | 1 | 1 | 1 |
| **Precision** | 0.9484 | 0.46003 | 0.68205 | 0.88318 | 0.93179 | 0.91066 | 0.93187 |

Bảng 3: Kết quả đạt được khi thử nghiệm

Tuy kết quả đạt được với các kinect khá cao, tuy nhiên với kinect 2, 3, 4 thì kết quả đạt được chưa tốt. Lý do là bởi dữ liệu huấn luyện chưa được tốt, gây ra sự sai sót ground truth đối với dữ liệu huấn luyện.



Hình 4. Minh hoạt một trường hợp phát hiện thiếu, bounding box vàng là dữ liệu ground truth của dữ liệu, bounding box màu tím là đối tượng được phát hiện.