**1. Tổng quan**

Đây là một khảo sát tổng quan về **các phương pháp nhận diện hành động người (HAR)** dựa trên thị giác máy tính.

Bài báo tập trung vào ba hướng chính:

**Biểu diễn đặc trưng hành động**: sử dụng cả đặc trưng thủ công và học sâu.

**Nhận diện tương tác người - vật**: phát hiện hành động khi có sự tham gia của vật thể.

**Phát hiện hành động (Action Detection)**: xác định thời điểm và vị trí xảy ra hành động trong video.

**2. Công nghệ sử dụng**

**Phương pháp truyền thống (Handcrafted Features)**

**Dữ liệu RGB**: STIP, HOG3D, iDT, BoVW (From Youtube, camera, video)

**Dữ liệu chiều sâu và bộ xương:** Depth Data and Skeletal Data

**Deep Learning**

* **CNN hai luồng (Two-Stream CNN)**: sử dụng cả ảnh RGB và quang trắc dòng chảy
* **Mạng tích chập 3D (3D CNNs)**: học đặc trưng không gian-thời gian từ video.
* **Mạng nơ-ron hồi tiếp (LSTM/RNNs)**: phân tích chuỗi thời gian và phát hiện chuyển động.
* **Mô hình lai (Hybrid Models)**: kết hợp giữa CNN + RNN để tối ưu hóa phát hiện hành động.

**3. Cách thu thập dataset**

**Dữ liệu video RGB**

UCF101 (13,320 video - 101 lớp hành động).

HMDB51 (6,766 video - 51 lớp hành động).

Kinetics (hơn 306,245 video - 400 lớp hành động).

**Dữ liệu RGB-D: kết hợp hình ảnh màu và cảm biến chiều sâu**

**Bộ dữ liệu phổ biến**:

* NTU RGB+D (56,880 video - 60 lớp hành động).
* MSR-Action3D (567 video - 20 lớp hành động).
* Daily Activity 3D (320 video - 16 lớp hành động).

**Dữ liệu thu thập từ cảm biến**

* Hệ thống Kinect để ghi lại dữ liệu chiều sâu và bộ xương.
* Camera hồng ngoại để phân tích chuyển động trong môi trường thực tế.

**4. Kết quả nghiên cứu**

**UCF101** Tốt nhất đạt **97.9%** (Deep Learning).

**HMDB51**: Độ chính xác **80.2%** khi dùng CNN 3D.

**Real time (Trang 12-13)**

**Hệ thống phát hiện tương tác người - vật**

Phương pháp mới nhất sử dụng mạng **R-CNN đa vùng** giúp nhận diện tốt hơn

**5.Research GAPs**

1.Khó giải thích vì sao ra kết quả vì sử dụng DL

2.Khó tổng quát hoá mô hình

3.Gặp khó khăn trong hoàn cảnh phức tạp: đông người, bị che khuất,…

4.Tấn công đối kháng (Adversarial Attacks in HAR) không hiểu

**Khả năng tổng quát hóa mô hình (Generalization Issues)**

* Các mô hình HAR hiện nay thường **không tổng quát tốt khi chuyển từ dataset này sang dataset khác**.

🔹 **Phát hiện hành động trong cảnh phức tạp (Complex Scene Understanding)**

* HAR vẫn gặp khó khăn khi có **đông người, thay đổi góc quay, hoặc bị che khuất**.