

# BÀI TẬP THỰC HÀNH: THEO DÕI ĐỐI TƯỢNG TRONG VIDEO

## I. MỤC TIÊU

Bài thực hành này được thiết kế nhằm trang bị cho người học những kiến thức và kỹ năng nền tảng trong lĩnh vực theo dõi đối tượng (Object Tracking) bằng các phương pháp thị giác máy tính. Sau khi hoàn thành, người học có khả năng:

1. **Hiểu rõ** các khái niệm cốt lõi và thách thức trong bài toán theo dõi đối tượng.
2. **Phân biệt và trình bày** nguyên lý hoạt động của các thuật toán theo dõi phổ biến.
3. **Cài đặt và áp dụng** một thuật toán theo dõi hoàn chỉnh để xử lý một chuỗi video thực tế.
4. **Đánh giá** hiệu suất của thuật toán một cách định lượng bằng các độ đo tiêu chuẩn.

## II. YÊU CẦU KIẾN THỨC NỀN TẢNG

- Kiến thức cơ bản về xử lý ảnh số và thị giác máy tính.
- Kinh nghiệm lập trình với ngôn ngữ Python.
- Hiểu biết về các thư viện khoa học dữ liệu: OpenCV, NumPy.
- Kiến thức về học sâu (Deep Learning) và các kiến trúc mạng như CNN là một lợi thế.

## III. TÀI NGUYÊN

1. **Bộ dữ liệu (Dataset):**
  - Bài thực hành sẽ sử dụng bộ dữ liệu **MOTChallenge**. Đây là một benchmark tiêu chuẩn trong cộng đồng nghiên cứu, cung cấp các chuỗi video đa dạng về bối cảnh, mật độ đối tượng và điều kiện ánh sáng, kèm theo các chú thích (annotations) chi tiết.
  - **Truy cập tại:** <https://motchallenge.net/>
  - **Nhiệm vụ:** Lựa chọn và tải về một chuỗi video từ bộ dữ liệu MOT17 hoặc MOT20 để tiến hành thực nghiệm.
2. **Thuật toán đề xuất:**
  - Người học được khuyến khích tìm hiểu và cài đặt một trong các thuật toán theo dõi dựa trên mô hình "Tracking-by-Detection" (Theo dõi dựa trên phát hiện). Hai phương pháp tiêu biểu bao gồm:
    - **SORT (Simple Online and Realtime Tracking):** Một phương pháp hiệu quả, kết hợp bộ lọc Kalman để dự đoán trạng thái chuyển động của đối tượng và thuật toán Hungarian để giải quyết bài toán gán ghép (data association) giữa các đối tượng đã theo dõi và các phát hiện mới.
    - **DeepSORT:** Một phiên bản cải tiến của SORT, tích hợp thêm một mô hình học sâu để trích xuất đặc trưng ngoại hình (appearance features) của đối tượng. Điều này giúp cải thiện khả năng duy trì định danh (ID) cho đối tượng, đặc biệt trong các trường hợp bị che khuất trong thời gian dài.

## IV. NỘI DUNG VÀ QUY TRÌNH THỰC HIỆN

## Bước 1: Thiết lập môi trường

- Cài đặt Python 3.x.

Sử dụng `pip` để cài đặt các thư viện cần thiết:

```
pip install opencv-python numpy scikit-learn
```

# Đối với DeepSORT, cần cài đặt thêm TensorFlow hoặc PyTorch

```
pip install tensorflow # hoặc torch, torchvision
```

- 

## Bước 2: Tải và phân tích dữ liệu

- Tải chuỗi video và tệp chú thích tương ứng từ MOTChallenge.
- Nghiên cứu cấu trúc của tệp chú thích (`gt/gt.txt`). Mỗi dòng thường chứa thông tin: `<frame_id>`, `<object_id>`, `<bb_left>`, `<bb_top>`, `<bb_width>`, `<bb_height>`, `<confidence>`, `<class>`, `<visibility>`.

## Bước 3: Cài đặt thuật toán

- **Phát hiện đối tượng (Object Detection):** Vì đây là phương pháp "Tracking-by-Detection", bước đầu tiên trong mỗi khung hình là phát hiện đối tượng. Người học có thể sử dụng một mô hình phát hiện đối tượng đã được huấn luyện trước (pre-trained) như YOLO, SSD, hoặc Faster R-CNN. Đối với bài thực hành này, có thể sử dụng các phát hiện đã được cung cấp sẵn trong bộ dữ liệu MOTChallenge để tập trung vào logic của thuật toán theo dõi.
- **Logic theo dõi (Tracking Logic):**
  1. **Khởi tạo:** Tạo một danh sách rỗng để quản lý các đối tượng đang được theo dõi (trackers).
  2. **Dự đoán:** Với mỗi khung hình mới, sử dụng bộ lọc Kalman để dự đoán vị trí tiếp theo của từng đối tượng trong danh sách trackers.
  3. **Gán ghép (Association):**
    - Tính toán ma trận chi phí (ví dụ: khoảng cách IoU - Intersection over Union) giữa các hộp giới hạn được dự đoán và các hộp giới hạn được phát hiện trong khung hình hiện tại.
    - Sử dụng thuật toán Hungarian để tìm ra phương án gán ghép tối ưu, giảm thiểu tổng chi phí.
  4. **Cập nhật:**
    - Đối với các cặp được gán ghép thành công, cập nhật trạng thái của bộ lọc Kalman bằng thông tin từ phát hiện mới.
    - Đối với các phát hiện không được gán ghép, khởi tạo một tracker mới.
    - Đối với các tracker không được gán ghép, tăng bộ đếm "tuổi". Nếu một tracker không được cập nhật trong một khoảng thời gian nhất định, hãy xóa nó khỏi danh sách.
- **(Tùy chọn cho DeepSORT):** Tích hợp mạng học sâu để trích xuất vector đặc trưng cho mỗi phát hiện. Sử dụng khoảng cách cosine giữa các vector đặc trưng này làm một thành phần bổ sung trong ma trận chi phí để cải thiện độ chính xác của việc gán ghép.

#### Bước 4: Trực quan hóa kết quả

- Đọc từng khung hình của video.
- Chạy thuật toán theo dõi.
- Vẽ hộp giới hạn và ID của mỗi đối tượng được theo dõi lên khung hình.
- Hiện thị hoặc lưu video kết quả.

#### V. ĐÁNH GIÁ VÀ BÁO CÁO

##### 1. Độ đo hiệu suất:

- Tính toán các độ đo tiêu chuẩn để đánh giá định lượng hiệu suất của thuật toán. Các độ đo quan trọng bao gồm:
  - **MOTA (Multiple Object Tracking Accuracy):** Phản ánh độ chính xác tổng thể, có tính đến các lỗi như phát hiện sai (false positives), bỏ sót (false negatives) và chuyển đổi ID (ID switches).
  - **MOTP (Multiple Object Tracking Precision):** Đo lường độ chính xác về vị trí của các đối tượng được theo dõi.
  - **IDS (ID Switches):** Số lần một đối tượng bị gán sai ID.

##### 2. Nội dung báo cáo:

- **Giới thiệu:** Trình bày tổng quan về bài toán và phương pháp tiếp cận.
- **Cài đặt thuật toán:** Mô tả chi tiết về cách cài đặt và các tham số đã sử dụng.
- **Kết quả thực nghiệm:**
  - Trình bày các kết quả định lượng (bảng các độ đo MOTA, MOTP, IDS).
  - Cung cấp các kết quả định tính (hình ảnh hoặc video minh họa các trường hợp theo dõi thành công và thất bại).
- **Thảo luận:** Phân tích ưu và nhược điểm của thuật toán dựa trên kết quả. Đề xuất các hướng cải tiến khả thi.
- **Kết luận:** Tóm tắt các kết quả chính đã đạt được.