

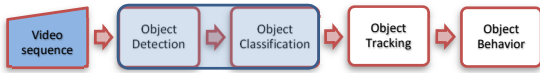
1 GIỚI THIỆU

Trong lý thuyết quản trị, việc nhận biết và đánh giá được hành vi khách hàng để đưa các chiến lược hiệu quả đóng vai trò quan trọng cho sự thành công trong kinh doanh. Trong đó, có một phần là đánh giá được hành vi, biểu hiện của khách hàng khi họ đến tham quan mua sắm tại các khu vực trưng bày, các gian hàng kinh doanh.

Để quyết định được một chiến thuật trưng bày, một thông tin không thể thiếu hỗ trợ cho người quản lý là phải đánh giá được hành vi của khách hàng qua các con số thống kê, tự động hóa việc nhận biết và đánh giá hành vi là một biện pháp hiện đại mang tính chính xác cao rất cần thiết trong trường hợp này.

Việc xây dựng hệ thống nhận dạng hành vi đối tượng thông qua việc áp dụng kỹ thuật theo dõi đối tượng qua camera xuất phát từ nhu cầu đó.

Hệ thống theo dõi đối tượng nhận vào các khung hình video thu nhận từ các camera, qua một số bước xử lý, phân tích và cuối cùng là đưa ra quỹ đạo đường đi của đối tượng theo thời gian làm cơ sở cho việc tiếp theo là nhận biết hành vi. Bắt đầu là quá trình phát hiện đối tượng chuyển động trong các khung hình. Sau đó các đối tượng này sẽ qua quá trình phân lớp để phân biệt các đối tượng thuộc lớp nào, sự vật nào như: con người, xe, mây bay, cây lác lư Tiếp theo là quá trình xử lý để theo vết nhằm tìm ra đường chuyển động của đối tượng, từ đó có thể phân tích, nhận biết hành vi của đối tượng. Hình 1 giới thiệu mô hình theo dõi đối tượng chuyển động, trong đó mỗi bước là một lĩnh vực nghiên cứu rộng lớn.



Hình 1: Mô hình theo dõi đối tượng chuyển động (Ragland & Tharcis, 2014)

2 THEO DÕI ĐỐI TƯỢNG THEO GIẢI THUẬT CMT (CONSENSUS-BASED TRACKING AND MATCHING)

Giải thuật CMT gốc đã được G.Nebhay khởi tạo từ năm 2014 (Nebhay & Pflugfelder, 2014), sau đó vào năm 2015 CMT được cải tiến thành giải thuật CMT thích nghi (Nebhay & Pflugfelder, 2015). Cũng giống với CMT, CMT thích nghi sử dụng vùng khởi tạo b_0 ở khung hình đầu tiên của video. Trong vùng này rút ra được tập các keypoint $P_0 = \{x_1^0, \dots, x_m^0\}$. Một cặp so khớp m_i được định nghĩa là $m_i = (x_i^t, x_i^0)$, với x_i^t là vị trí mới của x_i^0 trong khung ảnh thứ t . Ở khung ảnh thứ t , ta cần phải xác định được một tập các cặp so khớp $\mathcal{L}_t =$

$\{m_1, \dots, m_n\}$ biểu diễn cho đối tượng đang theo dõi càng chính xác càng tốt.

2.1 Tương ứng thích nghi tĩnh (static-adaptive correspondences)

Mô hình hình dáng tĩnh dựa trên mô hình hình dáng khởi tạo của đối tượng được tạo nên từ bộ mô tả của tập các keypoint $P_0 = \{x_1^0, \dots, x_m^0\}$. Chúng ta gọi các cặp so khớp thu được từ mô hình này là sự tương ứng tĩnh. Trước hết, chúng ta sử dụng tìm kiếm toàn cục để thiết lập các cặp so khớp giữa các keypoint x_i^0 trong khung ảnh khởi tạo và x_j^t trong khung ảnh hiện tại bằng cách sử dụng một ngưỡng (threshold) và điều kiện dựa trên khoảng cách láng giềng gần nhất (Lowe, 2004) sử dụng khoảng cách $d(\dots)$ giữa các mô tả của keypoint.

$$d(x_i^0, x_j^t) < \theta \wedge \frac{d(x_i^0, x_j^t)}{d(x_i^0, x_k^t)} < \gamma, j \neq k \quad (1)$$

Ngoài ra, ta cần phải loại bỏ các keypoint ứng viên mà nó được so khớp với các keypoint nền trong khung ảnh đầu tiên. Mô hình tĩnh này rất hiệu quả và có thể xử lý được các trường hợp phát hiện lại các keypoint sau khi bị che khuất. Tuy nhiên, mô hình này không thể phát hiện thêm các keypoint mới cho đối tượng đang được theo dõi.

Ngược lại, mô hình thích nghi được cập nhật lại ở mỗi khung ảnh, bao gồm các vùng ảnh nhỏ xung quanh các keypoint $x_i^{t-1} \in \mathcal{L}_{t-1}$. Trong khi ở mô hình tĩnh, ta cần phải tìm kiếm sự tương ứng trên toàn bộ khung ảnh hiện tại, thì với mô hình thích nghi ta có thể giả sử thời gian giữa hai khung ảnh liên tiếp tương đối nhỏ. Vì thế, ta có thể thiết lập sự tương ứng một cách hiệu quả bằng cách sử dụng luồng quang học từ khung ảnh $t-1$ đến khung ảnh t . Hợp của các keypoint được tìm kiếm bằng cách so khớp toàn cục và các keypoint có được do truy vết từ các keypoint trong khung ảnh $t-1$ tạo thành tập các keypoint \mathcal{L}_t^*

2.2 Gom cụm các tương ứng

Ý tưởng chính là sử dụng độ đo D phản ánh sự khác nhau giữa các tương ứng m_i và m_j dựa trên khả năng tương thích hình học của chúng, điều này phản ánh trực tiếp sự biến dạng của đối tượng. Độ đo D được định nghĩa như sau:

$$D(m_i, m_j) = \|(x_i^t - Hx_i^0) - (x_j^t - Hx_j^0)\| \quad (2)$$

Trong đó $\|\cdot\|$ là khoảng cách Eclide và H là phép biến đổi tương đồng được ước lượng từ \mathcal{L}_t^* .

D được sử dụng để phân hoạch \mathcal{L}_t^* vào trong các tập con sử dụng giải thuật phân cụm phân cấp từ trên