

với $N(m) = \{m-l, \dots, m, \dots, m+l\}$, $m-l \geq 0$ và $m+l \leq M$, M là số hàng của khung ảnh p , q . Giá trị l xác định số hàng lân cận cho việc tìm kiếm.

Đối với mỗi mảnh ghép có vector đặc trưng $\mathbf{x}_{m,n}^{A,p}$ trong khung ảnh p của camera A , chúng tôi thực hiện tìm kiếm mảnh ghép tương ứng trong khung ảnh q của camera B , nghĩa là tìm kiếm vector gần nhất trong tập hợp $\hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q})$. Hàm ước lượng độ giống nhau của hai vector đặc trưng được định nghĩa thông qua hàm Gaussian:

$$s(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{(\|\mathbf{x} - \mathbf{y}\|_2)^2}{2\sigma^2}\right) \quad (10)$$

với σ là tham số cho trước.

3.5 Ước lượng chỉ số đánh giá độ nổi bật

Ở phần này chúng tôi trình bày hai phương pháp được áp dụng là k -Nearest Neighbors (k -NN) (S. Byers và A. Raftery, 1998) và One Class Support Vector Machines (OCSVM) (K. Heller *et al.*, 2003) để xây dựng ma trận chỉ số đánh giá độ nổi bật của các đặc trưng hình dáng bên ngoài đối tượng.

3.5.1 k -NN

Gọi N_r là số lượng các ảnh trong tập hợp ảnh tham chiếu cho quá trình tái nhận dạng. Sau khi xây dựng vector đặc trưng (dCOLORSIFT) và thực hiện việc tìm kiếm các mảnh ghép tương ứng giữa ảnh truy vấn và các ảnh trong tập tham chiếu, ứng với mỗi mảnh ghép truy vấn $\mathbf{x}_{m,n}^{A,p}$, ta có được N_r mảnh ghép tương ứng tìm được từ N_r ảnh tham chiếu. Tập hợp các mảnh ghép này, tạm gọi là $X_{nn}(\mathbf{x}_{m,n}^{A,p})$, được định nghĩa theo công thức sau:

$$X_{nn}(\mathbf{x}_{m,n}^{A,p}) = \left\{ \mathbf{x} \mid \arg \max_{\hat{\mathbf{x}} \in \hat{S}_{p,q}} s(\mathbf{x}_{m,n}^{A,p}, \hat{\mathbf{x}}), q = 1, \dots, N_r \right\} \quad (11)$$

với $\hat{S}_{p,q} = \hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q})$ là tập tìm kiếm lân cận trong phương trình (9) và s là hàm ước lượng độ giống nhau định nghĩa trong phương trình (10).

Khoảng cách KNN (S. Byers và A. Raftery, 1998) được sử dụng để ước lượng chỉ số đánh giá độ nổi bật (salience score):

$$\text{score}_{knn}(\mathbf{x}_{m,n}^{A,p}) = D_k(X_{nn}(\mathbf{x}_{m,n}^{A,p})) \quad (12)$$

với D_k là khoảng cách giữa mảnh ghép truy vấn và mảnh ghép tương ứng gần nhất thứ k .

3.5.2 OCSVM

Giải thuật thứ hai được áp dụng trong phần này là One Class Support Vector Machines (OCSVM) (K. Heller *et al.*, 2003). Ý tưởng cơ bản của OCSVM là tìm kiếm một mặt siêu cầu biểu diễn được hầu hết các điểm dữ liệu trong không gian đặc trưng và có kích thước nhỏ nhất có thể. Mặt siêu cầu này tìm được thông qua việc giải bài toán tối ưu với hàm quyết định $f(\mathbf{X})$ được định nghĩa trong (R. Zhao *et al.*, 2013).

Chỉ số đánh giá độ nổi bật trong trường hợp này được ước lượng như sau:

$$\text{score}_{ocsvm}(\mathbf{x}_{m,n}^{A,p}) = d(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^*) \quad (13)$$

với $\mathbf{x}^* = \arg \max_{\mathbf{x} \in X_{nn}(\mathbf{x}_{m,n}^{A,p})} f(\mathbf{x})$, d là khoảng cách

Euclidean giữa các vector đặc trưng.

3.6 Tái nhận dạng đối tượng

Phần này trình bày bước cuối cùng trong chuỗi xử lý bài toán tái nhận dạng đối tượng. Xét một cặp ảnh gồm ảnh của đối tượng cần truy vấn (ảnh p) và ảnh trong tập tìm kiếm (ảnh q). Ứng với mỗi mảnh ghép $\mathbf{x}_{m,n}^{A,p}$, ta tìm được mảnh ghép tương ứng trong miền tìm kiếm $\hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q})$ là:

$$\mathbf{x}_{i,j}^{B,q} = \arg \max_{\hat{\mathbf{x}} \in \hat{S}_{p,q}} s(\mathbf{x}_{m,n}^{A,p}, \hat{\mathbf{x}}) \quad (14)$$

Ảnh của đối tượng trong tập ảnh tìm kiếm có độ tương đồng cao nhất với ảnh truy vấn được xác định như sau:

$$q^* = \arg \max_q \text{Sim}(\mathbf{x}^{A,p}, \mathbf{x}^{B,q}) \quad (15)$$

với hàm Sim là hàm đo độ giống nhau giữa hai đối tượng trong khung ảnh p và q . Hàm Sim được ước lượng như sau:

$$\text{Sim}(\mathbf{x}^{A,p}, \mathbf{x}^{B,q}) = \sum_{m,n} \frac{\text{score}(\mathbf{x}_{m,n}^{A,p}) \cdot s(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}_{i,j}^{B,q}) \cdot \text{score}(\mathbf{x}_{i,j}^{B,q})}{\alpha + |\text{score}(\mathbf{x}_{m,n}^{A,p}) - \text{score}(\mathbf{x}_{i,j}^{B,q})|} \quad (16)$$