

$$\delta[\emptyset(x)] = \begin{cases} 1, & \emptyset(x) = 0 \\ 0, & |\emptyset(x)| < \varepsilon \\ D[\emptyset(x)] & |\emptyset(x)| \geq \varepsilon \end{cases} \quad (3)$$

$$\text{với } D[\phi(x)] = \frac{1}{2\pi} \left(1 + \cos\left(\frac{\pi\phi(x)}{\varepsilon}\right) \right).$$

Gọi y là biến không gian thứ hai, độc lập với biến x . Cả hai biến này đều dùng để biểu diễn một điểm nào đó trong miền không gian của ảnh. Hàm đặc trưng $B(x, y)$ dùng để đánh dấu các vùng cục bộ có dạng hình tròn với tâm là x và bán kính bằng r được định nghĩa như sau:

$$B(x, y) = \begin{cases} 1, & \|x - y\| < r \\ 0, & \text{Khác} \end{cases} \quad (4)$$

Hàm năng lượng E trong vùng cục bộ định nghĩa bởi $B(x, y)$ được định nghĩa như sau:

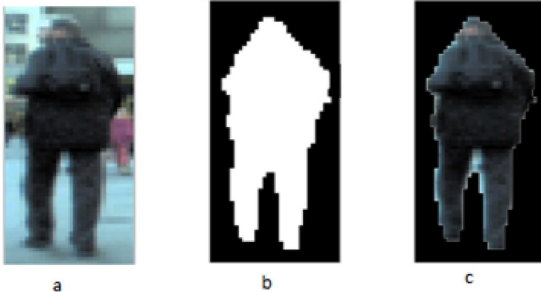
$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} B(x, y) F(I(y), \phi(y)) dy dx \quad (5)$$

với F là hàm năng lượng mô tả mức độ giống nhau và khác nhau giữa các pixel trong các vùng cục bộ dọc đường biên. Các dạng khác nhau của hàm F được định nghĩa chi tiết như trong nghiên cứu (Shawn Lankton et al., 2008).

Để áp dụng giải thuật phân đoạn bằng đường biên động vào việc tách miền nền trong khung ảnh người, một mặt nạ ảnh thô tương ứng với ba vùng đầu, thân và chân của người được khởi tạo, làm cơ sở cho các vòng lặp tìm kiếm. Giải thuật phân đoạn bằng đường biên động sau đó được áp dụng để có được mặt nạ M ứng với vùng đối tượng. Ảnh loại bỏ nền I_M thu được thông qua phép nhân chập ma trận:

$$I_M = I \otimes M \quad (6)$$

Hình 6 trình bày kết quả của quá trình tách đối tượng ra khỏi nền.



Hình 6: Tách đối tượng ra khỏi ảnh nền

3.3 Trích xuất các đặc trưng cục bộ

Để mô tả hình dáng bên ngoài của các đối tượng cần truy vấn, khung ảnh của đối tượng sau khi đã loại bỏ phần nền trước hết sẽ được chia thành các mảnh ghép cục bộ (local patch), với mỗi mảnh ghép có kích thước là $k \times l$ pixel. Mỗi mảnh ghép tiếp theo sẽ được trích hai đặc trưng là histogram trong không gian màu LAB và đặc trưng SIFT. Vector tính được từ hai loại đặc trưng này sẽ được ghép lại thành một vector tổng quát, đặc tả thông tin của từng mảnh ghép, và được gọi là vector đặc trưng dCOLORSIFT (R. Zhao et al., 2013).

3.4 Tìm kiếm mảnh ghép tương ứng

Do góc nhìn của các camera giám sát là khác nhau và người trong quá trình di chuyển sẽ có sự thay đổi về hình dáng bên ngoài, vì vậy các miền tương ứng giữa hai khung ảnh của cùng một đối tượng sẽ không chính xác ở cùng một vị trí. Bước xử lý này sẽ thực hiện việc tìm kiếm các mảnh ghép “tương ứng” trong một miền linh động xung quanh vị trí của mảnh ghép truy vấn. Mảnh ghép tương ứng được xem là mảnh ghép có độ giống nhất với mảnh ghép đem đi truy vấn.

Gọi $\mathbf{x}_{m,n}^{A,p}$ là vector đặc trưng trích từ mảnh ghép ở hàng m và cột n trong ảnh thứ p thu được từ camera A . Gọi $\mathcal{T}^{A,p}(m)$ là tập hợp các mảnh ghép ở hàng thứ m :

$$\mathcal{T}^{A,p}(m) = \{\mathbf{x}_{m,n}^{A,p} | n = 1, 2, \dots, N\} \quad (7)$$

Tất cả các mảnh ghép trong tập hợp $\mathcal{T}^{A,p}(m)$ có cùng một tập tìm kiếm mảnh ghép tương ứng trong ảnh thứ q thu được từ camera B :

$$S(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q}) = \mathcal{T}^{B,q}(m), \quad \forall \mathbf{x}_{m,n}^{A,p} \in \mathcal{T}^{A,p}(m) \quad (8)$$

với $\mathbf{x}^{B,q}$ là tập hợp các mảnh ghép trong ảnh q thu được từ camera B . Trong trường hợp lý tưởng, tập S là hàng thứ m của ảnh q ở camera B . Tuy nhiên, việc tìm kiếm theo hàng tương ứng thường không mang lại hiệu quả cao vì sự thay đổi góc nhìn của camera giám sát, cũng như sự thay đổi về hình dáng của người khi di chuyển. Để đối phó với các biến thể không gian, việc tìm kiếm được mở rộng ra trên các hàng liền kề với hàng cần tìm kiếm, việc tìm kiếm này được mô tả theo công thức sau:

$$\hat{S}(\mathbf{x}_{m,n}^{A,p}, \mathbf{x}^{B,q}) = \{\mathcal{T}^{B,q}(b) | b \in N(m)\}, \quad (9)$$

$$\forall \mathbf{x}_{m,n}^{A,p} \in \mathcal{T}^{A,p}(m)$$