

$$P = \{(a_i, f_i)\}_{i=1}^{N^P} \quad (3)$$

Mỗi keypoint trong tập P, cần phải tính khoảng cách Hamming theo công thức (4) giữa bộ mô tả của nó với bộ mô tả của từng keypoint được tìm thấy trong I₁, bao gồm cả các keypoint nền.

$$d(f^1, f^2) = \sum_{i=1}^d XOR(f_i^1, f_i^2) \quad (4)$$

Để so khớp các keypoint trong P với các keypoint trong I₁, khoảng cách đến láng giềng gần nhất phải gần hơn các láng giềng khác theo một tỷ lệ cố định ρ nào đó (Lowe, 2014). Tập các keypoint M là một tập con chứa các keypoint đã so khớp giữa keypoint trong P và các keypoint trong O, sau đó bổ sung thêm các chỉ số (index) của keypoint tương ứng (là các định danh m trong công thức (2))

Để theo vết, chúng ta tính toán sự dịch chuyển của K_{t-1} từ I_{t-1} đến I_t bằng cách dùng phương pháp ước lượng luồng quang học (Lucas & Kanate, 1981). Tại thời điểm t=2, K₁ bằng với tập các keypoint của O. Tập các keypoint T được tìm ra bằng cách cập nhật vị trí của keypoint trong K_{t-1} nhưng vẫn giữ lại các định danh của keypoint.

Sau đó thực hiện kết hợp hai tập T và M và đưa vào tập K' với kích thước NK', khi kết hợp hai tập T, M thì các keypoint giống nhau ở hai tập này sẽ được loại bỏ. Đến bước này trong tập K' vẫn còn chứa các keypoint ngoại lai hoặc các điểm mơ hồ không chính xác.

b. Bình chọn

Để xác định vị trí của đối tượng, mỗi keypoint (a, m) trong K' sẽ bình chọn một giá trị $h(a, m) \rightarrow \mathbb{R}^2$ vào tâm của đối tượng, tập các bình chọn đó được thể hiện như sau:

$$V = \{h(a_i, m_i)\}_{i=1}^{N^{K'}} \quad (5)$$

Ta có thể xem mỗi bình chọn (vote) là một phép biến đổi hình học affine gồm: tịnh tiến, tỉ lệ và quay lên các keypoint trong không gian 2 chiều.

Trước hết, với phép tịnh tiến, $h^T(a, m)$ được tính như sau:

$$h^T(a, m) = a - r_m \quad (6)$$

Trong đó, r_m là vị trí tương đối của keypoint tương ứng với chỉ số (index) của m trong O. Khi đối tượng bị thay đổi tỉ lệ (ví dụ thu nhỏ), các vector bình chọn này tiến gần đến trung tâm đối tượng (các vector bị thu ngắn lại xem Hình 5). Khi đó tỷ lệ thu nhỏ của các vector bình chọn được tính lại như sau:

$$h^S(a, m) = a - s \cdot r_m \quad (7)$$

Để tính s, chúng ta tính khoảng các Euclide giữa các cặp a_i và a_j trong K', và so sánh khoảng cách vừa tính được với khoảng cách của các keypoint tương ứng r_{m_i} và r_{m_j} trong O.

Ta có: $a_{i,j} = a_i - a_j$ và $r_{i,j} = r_{m_i} - r_{m_j}$

$$D_s = \left\{ \frac{\|a^{i,j}\|}{\|r^{i,j}\|}, i \neq j \right\} \quad (8)$$

Từ đó rút ra được $s = \text{median}(D_s)$

Khi đối tượng bị quay, các vector bình chọn bị quay theo, và tâm quay là tâm của đối tượng:

$$h^R(a, m) = a - s \cdot R r_m \quad (9)$$

Với R là ma trận quay trong không gian 2D

$$R = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad (10)$$

Để tính được góc quay α , cần phải tính được các góc $\alpha_{i,j}$ (theo công thức 11) là góc tạo bởi sự chênh lệch góc quay của hai đường thẳng so với trục ngang:

– Đường thẳng thứ nhất nối hai điểm thứ i và thứ j trong khung hình đầu tiên

– Đường thẳng thứ hai nối hai điểm thứ i và thứ j trong khung hình thứ t.

$$\alpha_{i,j} = \text{atan2}(a_y^{i,j}, a_x^{i,j}) - \text{atan2}(r_y^{i,j}, r_x^{i,j}) \quad (11)$$

Duyệt qua tất cả các cặp keypoint ta được tập D_a như sau:

$$D_a = \{\alpha_{i,j}, i \neq j\} \quad (12)$$

Từ đó rút ra được $\alpha = \text{median}(D_a)$

c. Đồng thuận (Consensus)

Khi có một keypoint trong K' không thuộc về đối tượng (key point ngoại lai) thì vector bình chọn của nó sẽ không hướng đến tâm của đối tượng mà nó chỉ đến một vùng hình ảnh bất kỳ nào đó. Trước khi tìm tâm của đối tượng μ , chúng ta cần loại bỏ các keypoint ngoại lai này và các vector bình chọn của nó bằng cách tìm kiếm sự đồng thuận của các vector bình chọn.

Để thực hiện việc này, chúng ta phân V thành các cụm nhỏ hơn V₁, V₂, ..., V_m bằng cách áp dụng giải thuật phân cụm phân cấp theo hướng từ dưới lên (hierarchical agglomerative clustering) trên các tập điểm trong V sử dụng khoảng cách Euclide. Trong cách phân cụm này, dữ liệu được tổ chức thành các cấu trúc phân cấp theo ma trận khoảng cách, kết quả là một cây phân cấp (dendrogram), sau đó cây phân cấp này sẽ bị cắt theo một ngưỡng δ xác định nào đó (dùng phân cấp khi khoảng cách Euclide giữa hai cụm lớn hơn hoặc bằng ngưỡng δ).