

2.2. Tiền xử lý ảnh

Tiền xử lý ảnh là một bước can thiệp kỹ thuật mang tính nền tảng trong mọi hệ thống thị giác máy tính, đóng vai trò bản lề nhằm triệt tiêu các tác nhân gây nhiễu, chuẩn hóa không gian dữ liệu đầu vào và khoanh vùng chính xác khu vực chứa đối tượng mục tiêu. Đối với đặc thù của bài toán nhận diện thực vật, hình ảnh thu thập thường chịu ảnh hưởng mạnh bởi điều kiện môi trường chiếu sáng, bóng đổ và các cấu trúc hậu cảnh phức tạp. Nhằm giải quyết triệt để những thách thức này, hệ thống đã thiết kế và triển khai một luồng xử lý hoàn toàn tự động, tích hợp ba kỹ thuật cốt lõi được tối ưu hóa khắt khe cả về mặt cơ sở toán học lẫn logic lập trình.

2.2.1. Không gian màu và đọc ảnh

Hệ thống ưu tiên sử dụng không gian màu HSV thay vì không gian RGB truyền thống để tiến hành các thao tác tách nền. Căn nguyên của sự lựa chọn này nằm ở việc không gian HSV tách biệt hoàn toàn kênh sắc độ đại diện cho màu sắc thực sự khỏi cường độ ánh sáng, giúp thuật toán phân lập vùng lá cây xanh duy trì được độ chính xác và tính ổn định cao ngay cả dưới các môi trường chiếu sáng biến thiên phức tạp. Về mặt toán học, công thức chuyển đổi từ không gian màu RGB với các giá trị chuẩn hóa $R, G, B \in [0, 1]$ sang HSV được thực hiện thông qua việc xác định các giá trị cực đại và cực tiểu. Gọi $V_{max} = \max(R, G, B)$, $V_{min} = \min(R, G, B)$ và độ chênh lệch là $\Delta = V_{max} - V_{min}$. Kênh giá trị sáng được xác định trực tiếp bằng giá trị cực đại:

$$V = V_{max}$$

Kênh độ bão hòa được tính toán dựa trên tỷ lệ của độ chênh lệch biên độ so với mức cường độ sáng cực đại, nhằm đánh giá mức độ thuần khiết của màu sắc:

$$S = \begin{cases} 0 & \text{nếu } V_{max} = 0 \\ \frac{\Delta}{V_{max}} & \text{nếu } V_{max} \neq 0 \end{cases}$$

Cuối cùng, kênh sắc độ mô tả thuộc tính màu sắc cốt lõi được tính toán bằng hệ phương trình điều kiện phân nhánh, tùy thuộc hoàn toàn vào việc kênh màu cơ bản nào đang chiếm cường độ ưu thế nhất trên điểm ảnh đang xét:

$$H = \begin{cases} 0 & \text{nếu } \Delta = 0 \\ 60^\circ \times \left(\frac{G - B}{\Delta} \bmod 6 \right) & \text{nếu } V_{max} = R \\ 60^\circ \times \left(\frac{B - R}{\Delta} + 2 \right) & \text{nếu } V_{max} = G \\ 60^\circ \times \left(\frac{R - G}{\Delta} + 4 \right) & \text{nếu } V_{max} = B \end{cases}$$

2.2.2. Bộ lọc làm mờ Gaussian

Trước khi thực thi các thuật toán phân đoạn ảnh phức tạp như Otsu hay GrabCut, hệ thống bắt buộc tiến hành một bước làm mịn không gian ảnh thông qua bộ lọc Gaussian. Thao tác này nhằm mục đích loại bỏ các nhiễu hạt tần số cao và làm mượt các chi tiết vi mô không cần thiết trên bề mặt. Quy trình này giúp các thuật toán tách nền không bị đánh lừa bởi các biến động cường độ sáng cục bộ, từ đó tập trung toàn bộ năng lực tính toán vào cấu trúc hình học tổng thể của chiếc lá.

Về mặt lý thuyết, bộ lọc Gaussian thực hiện phép chập ảnh với một hạt nhân được xây dựng dựa trên hàm phân phối chuẩn hai chiều. Hàm Gaussian dùng để tạo kernel với kích thước $(2k + 1) \times (2k + 1)$ và độ lệch chuẩn σ được định nghĩa như sau:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$

Trong đó, x và y là khoảng cách từ tâm của hạt nhân theo các trục tương ứng. Hệ thống sẽ tự động xây dựng ma trận hạt nhân bằng cách tính toán giá trị phân bố tại từng tọa độ (x, y) , ngay sau đó thực hiện bước chuẩn hóa dữ liệu để đảm bảo tổng của tất cả các phần tử cấu thành kernel luôn bằng 1, ngăn chặn hiện tượng làm thay đổi độ sáng trung bình của bức ảnh sau khi lọc:

$$G_{normalized}(x, y) = \frac{G(x, y)}{\sum(i, j)}$$

Quá trình lọc thực tế được triển khai bằng phép chập cửa sổ trượt, trong đó hạt nhân Gaussian quét tuần tự qua từng điểm ảnh của bức ảnh gốc. Tại mỗi vị trí tọa độ, cường độ điểm ảnh mới được nội suy bằng tổng có trọng số của các điểm ảnh lân cận tương ứng với ma trận hạt nhân:

$$I_{blurred}(i, j) = \sum_{m, n} I(i + m, j + n) \cdot G_{normalized}(x, y)$$

Để xử lý triệt để hiệu ứng hiệu ứng lè đối với các điểm ảnh nằm sát biên, hệ thống áp dụng kỹ thuật đệm bằng cách bổ sung các dải điểm ảnh có giá trị bằng 0 bao quanh khung hình trước khi chụp. Kết quả đầu ra là một bức ảnh có độ chuyển màu mượt mà, các gai nhiễu biên được triệt tiêu hoàn toàn, tạo ra một vùng dữ liệu an toàn để định hướng cho các thuật toán phân đoạn phía sau hoạt động với độ tin cậy cao nhất.

2.2.3. Tách nền và tạo mặt nạ

Để đối phó với sự đa dạng của các loại hậu cảnh thực tế, hệ thống không phụ thuộc vào một phương pháp đơn lẻ mà xây dựng một chiến lược pipeline bao gồm ba bước dự phòng tuần tự, hướng tới mục tiêu tối thượng là sinh ra một mặt nạ nhị phân hoàn hảo nhất. Ở lớp xử lý đầu tiên, hệ thống thực thi việc tách nền trực tiếp bằng dải màu xanh trên không gian HSV kết hợp chặt chẽ với các phép toán hình thái học. Do bề mặt thực vật thường tồn tại các mạng lưới gân lá khác màu hoặc chịu tác động từ phản quang, mặt nạ ban đầu thu được thường xuất hiện các lỗ rỗng cục bộ. Để khắc phục triệt để khiếm khuyết này, hệ thống áp dụng phép toán hình thái học đóng với một phần tử cấu trúc B tác động lên không gian ảnh nguyên bản A . Phép toán này là sự giao thoa tuần tự giữa phép giãn $\oplus$ và phép co $\ominus$, chuyên đảm nhiệm vai trò lấp kín các khe hở và lỗ hổng bên trong khu vực nội nhân:

$$A \bullet B = (A \oplus B) \ominus B$$

Ngay sau đó, phép mở được áp dụng nhằm loại bỏ các nhiễu hạt nhỏ lẻ nằm bên ngoài phần nền:

$$A \circ B = (A \ominus B) \oplus B$$

Trong những trường hợp đặc thù khi vật thể lá mang màu sắc dị biệt khiến màng lọc HSV mất tác dụng, hệ thống lập tức chuyển hướng sang phương án dự phòng thứ nhất bằng thuật toán GrabCut. Bản chất của thuật toán phân đoạn tiên tiến này là quá trình phân tách tiền cảnh và hậu cảnh thông qua việc giải bài toán tối ưu hóa một hàm năng lượng đồ thị tổng quát:

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z)$$

Trong đó, hàm dữ liệu U đánh giá xác suất một điểm ảnh thuộc về vùng lá hay nền thông qua mô hình trộn Gaussian, và hàm làm trơn V có nhiệm vụ khuyến khích

các điểm ảnh lân cận mang màu sắc tương đồng sẽ được gán chung một nhãn phân loại.

Cuối cùng, nếu cấu trúc ảnh có độ nhiễu loạn vượt quá giới hạn xử lý của GrabCut, bức ảnh sẽ được đưa về dạng ma trận đa mức xám để kích hoạt lá chắn bảo vệ cuối cùng là thuật toán phân ngưỡng tự động Otsu. Thuật toán kinh điển này sẽ quét toàn bộ biểu đồ tần suất mức xám để tìm kiếm một ngưỡng phân cách T tối ưu phân tách ảnh thành hai lớp, sao cho phương sai liên lớp σ_b^2 đạt giá trị cực đại:

$$\sigma_b^2(T) = \omega_0(T)\omega_1(T)[u_0(T) - u_1(T)]^2$$

Tại công thức này, $\omega_0(T)$ và $\omega_1(T)$ là xác suất phân bố của hai lớp điểm ảnh được phân tách bởi ngưỡng T , trong khi $u_0(T)$ và $u_1(T)$ cung cấp giá trị trung bình cường độ mức xám của từng lớp. Thông qua chiến lược ba lớp phòng ngự vững chắc này, hệ thống đảm bảo duy trì khả năng cô lập vật thể cực kỳ mạnh mẽ trước mọi điều kiện nhiễu loạn của bộ dữ liệu thực tế.