

Tra cứu ảnh dựa trên nội dung với thuật toán SURT

Bùi Thị Hương*

*ThS. Trường Đại học Hải Phòng

Received: 6/2/2023; Accepted: 10/2/2023; Published: 14/2/2023

Abstract: Content-Based Image Retrieval - CBIR (Content-Based Image Retrieval) includes concepts, purposes, models, components, functions and several content-based image retrieval systems. The image matching problem is a subproblem of the image query problem. Image matching is the matching of the disproportionate features of two images. In this paper, I mainly study the image matching method based on invariant features using SURF algorithm.

SURF is an invariant detector and descriptor of points of interest with scale and rotation. This method is equivalent or even faster than the previous proposed methods with regard to repetition, specificity and robustness, it also makes computation and comparison faster. SURF achieves this by building on integrated images with multiple image folds through building on the strengths of leading detectors and descriptors (here using the phantom method). Hessian's matrix to measure the detector and based on the distribution method for the descriptors).

Keywords: Content-Based Image Retrieval; CBIR; image matching; SURF; QBIC;

1. Đặt vấn đề

Trong những thập kỷ gần đây, cùng với sự phát triển ngày càng mạnh mẽ của khoa học kỹ thuật, lĩnh vực xử lý ảnh vẫn còn là một ngành khoa học rất mới mẻ so với các ngành khoa học khác nhưng nó đã là một lĩnh vực thu hút rất đông đảo các nhà khoa học quan tâm và phát triển, thúc đẩy các lĩnh vực nghiên cứu chuyên sâu về lĩnh vực này.

Xử lý ảnh (XLA) đóng vai trò quan trọng trong nhiều ứng dụng thực tế và khoa học kỹ thuật như: giám sát an ninh, nhận dạng đối tượng, nhận dạng khuôn mặt, phát hiện chuyển động, theo dõi chuyển động, nhận dạng các khối u trong y học, hiệu chỉnh các ảnh và video...

Một trong những bài toán quan trọng trong XLA là bài toán truy vấn ảnh. Tuy nhiên do sự bùng nổ thông tin ngày nay, khối lượng dữ liệu khổng lồ đã làm cho bài toán truy vấn ảnh vô cùng phức tạp, vì thế mà bài toán truy vấn ảnh theo nội dung ra đời khắc phục nhược điểm của các phương pháp tìm kiếm truyền thống. Có nhiều thuật toán đối sánh ảnh dựa trên các đặc điểm bất biến, trong khuôn khổ của bài báo này tác giả sử dụng thuật toán SURF cũng như phân tích các ưu nhược điểm của thuật toán này.

2. Nội dung nghiên cứu

2.1 Cơ sở lý thuyết của thuật toán

Khi làm việc với các đặc tính cục bộ, vấn đề đầu tiên cần phải được giải quyết là mức độ yêu cầu của bất biến. Điều này dựa trên sự biến dạng hình học và trắc quang mà bản thân nó được xác định bằng khả

năng bất biến trong điều kiện hướng nhìn thay đổi. Bài báo tập trung vào bộ phát hiện và bộ mô tả bất biến về tỷ lệ và góc xoay. Điều này dường như mang đến một sự kết hợp tốt giữa các đặc tính phức tạp, rộng lớn và thường xảy ra sự biến dạng.

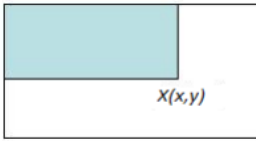
Lowe đã viết sự phức tạp của các thuộc tính bất biến thường có một tác động tiêu cực đến tính bền vững của ảnh trừ khi thay đổi hướng nhìn thực sự lớn. Trong một số trường hợp, thậm chí bất biến quay có thể được bỏ qua, kết quả là chỉ một phiên bản mô tả bất biến mà ta gọi là U-SURF.

SURF là máy dò được dựa vào ma trận Hessian nhưng sử dụng một xấp xỉ cơ bản giống như DoG là một máy dò dựa trên cơ sở Laplacian. Nó dựa trên ảnh tích hợp để làm giảm thời gian tính toán và do đó chúng ta gọi nó là phát hiện 'Fast-Hessian'. Mặt khác, PP này còn mô tả sự phân bố của các đặc trưng Haar-wavelet trong khu vực điểm quan tâm.

PP này khai thác những hình ảnh thích hợp để cải tiến tốc độ. Hơn nữa, việc chỉ sử dụng kích thước 64 chiều nên giảm thời gian cho tính toán và đối sánh và tăng tính ổn định. Bài báo cũng trình bày một bước đánh chỉ số mới dựa trên các ký hiệu của Laplacian, làm tăng tốc độ đối sánh và sự vững mạnh của các mô tả.

Ảnh tích hợp được đưa ra bởi Viola và Jones là một mảng 2 chiều với kích thước bằng kích thước của ảnh cần tính các đặc trưng, mỗi phần tử của mảng này được tính bằng cách tính tổng của điểm ảnh phía trên (dòng-1) và bên trái (cột-1) của nó. Bắt đầu từ vị trí trên, bên

trái đến vị trí dưới, phải của ảnh, việc tính toán này đơn thuần chỉ dựa trên phép cộng số nguyên đơn giản, do đó tốc độ thực hiện rất nhanh.



$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

Với I_{Σ} được tính toán thì chỉ mất bốn phép cộng để tính toán tổng các cường độ bất kỳ của một hình chữ nhật độc lập với kích thước của nó.

Ảnh tích hợp cho phép thực hiện nhanh chóng các loại bộ lọc nếp cuộn. Sự tham gia của một hình ảnh tích hợp $I_{\Sigma}(x)$ tại một vị trí $x = (x, y)$ đại diện cho tổng của tất cả các điểm ảnh trong hình ảnh đầu vào I của một khu vực hình chữ nhật được hình thành bởi các điểm x với $I_{\Sigma}(x) = \sum_{i=0}^{i < x} \sum_{j=0}^{j < y} I(i, j)$. Khi tính toán I_{Σ} chỉ phải mất bốn phép tính để tính tổng các cường độ phía trên bên phải khu vực hình chữ nhật độc lập với kích thước của nó.

2.2. Bộ mô tả SURF

Hiệu quả của SIFT tốt hơn nhiều so với các mô tả khác. Sự pha trộn của các thông tin sơ sài cục bộ hóa và sự phân bố của gradient liên quan đến các thuộc tính đường như mang lại sức mạnh đặc biệt tốt tránh ảnh hưởng của lỗi cục bộ hóa về tỉ lệ hoặc không gian. Việc sử dụng thể mạnh và hướng của gradient làm giảm ảnh hưởng của thay đổi trắc quang.

Bộ mô tả SURF được đề xuất dựa trên tính chất tương tự với độ phức tạp giảm dần. Bước đầu tiên là gán hướng dựa trên các thông tin từ một khu vực hình tròn xung quanh điểm quan tâm. Sau đó, chúng ta xây dựng một khu vực hình vuông phù hợp với hướng lựa chọn và trích xuất các mô tả SURF từ nó. Ngoài ra một đề xuất phiên bản của mô tả (U-SURF) không bất biến khi xoay hình ảnh nhưng tính toán nhanh hơn và phù hợp hơn cho các ứng dụng máy ảnh.

Đối với việc khai thác các mô tả, bước đầu tiên bao gồm xây dựng một cấu trúc hình vuông tập trung vào các điểm quan tâm và định hướng theo hướng lựa chọn trong phần trước. Các chuyển đổi này là không cần thiết. Kích thước của cửa sổ này là 20s. Khu vực này thường được chia ra nhỏ hơn 4×4 tiểu vùng vuông. Đối với từng tiểu vùng ta tính toán một vài thuộc tính đơn giản với không gian 5×5 điểm mẫu đồng đều. Vì lý do đơn giản, ta gọi dx là phản ứng Wavelet Haar theo hướng ngang và dy là phản ứng Wavelet

Haar theo hướng thẳng đứng (2s kích thước lọc). “Ngang” và “thẳng đứng” ở đây là khái niệm liên quan đến hướng điểm quan tâm được lựa chọn.

Để tăng độ vững mạnh theo hướng biến dạng hình học và các lỗi cục bộ hóa, các phản ứng dx và dy trước hết được gán trọng số Gaussian ($\sigma=3.3s$) tập trung tại các điểm quan tâm.

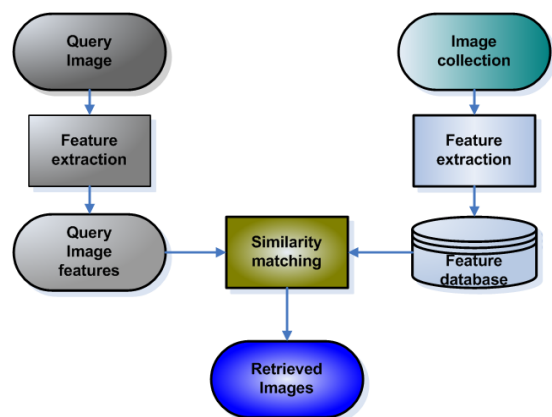
Sau đó, các phản ứng dx và dy Wavelet được tổng hợp qua từng tiểu vùng và tạo thành một thực thể đầu tiên của tập hợp các mục cho vector đặc trưng. Để có được thông tin về sự phân cực của việc thay đổi cường độ ta cũng trích tổng các giá trị tuyệt đối của các phản ứng $|dx|$ và $|dy|$. Do đó, mỗi tiểu vùng có bốn chiều vector mô tả v, cơ cấu cường độ của nó $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$.

Điều này dẫn đến một vector mô tả cho tất cả 4×4 tiểu vùng có độ dài 64. Các phản ứng Wavelet là bất biến đối với một sự ưu tiên trong vùng chiếu sáng (offset). Sự bất biến trong tương phản (một yếu tố tỉ lệ) đạt được bằng cách chuyển các mô tả vào một vector đơn vị.

Việc đối sánh chỉ mục thực hiện rất nhanh bao gồm các dấu hiệu của Laplacian (tức là các dấu vết của ma trận Hessian) cho các điểm quan tâm cơ bản. Thông thường những điểm quan tâm được tìm thấy ở các cấu trúc blob-type. Các dấu hiệu của Laplacian phân biệt các đốm màu sáng trên nền tối từ hình ngược lại. Thuộc tính này có sẵn mà không mất thêm chi phí tính toán vì nó đã được tính toán trong giai đoạn phát hiện. Trong giai đoạn tiếp ta chỉ so sánh thuộc tính nếu chúng có cùng một độ tương phản. Do đó, thông tin tối thiểu này cho phép đối sánh nhanh hơn và tăng hiệu suất.

2.3. Xây dựng hệ truy vấn ảnh

Mô hình hệ truy vấn ảnh (CBIR)



Hình 2.1: Mô hình truy vấn ảnh

Ảnh đầu vào (còn gọi là ảnh truy vấn) được trích chọn đặc trưng và việc truy vấn 1 ảnh trở thành truy vấn các đặc trưng của ảnh đó. Các ảnh trong cơ sở dữ liệu cũng lần lượt được trích chọn đặc trưng và lưu các đặc trưng này trong cơ sở dữ liệu.

Việc truy vấn ảnh trở thành việc đối sánh độ tương đồng của các đặc trưng cần truy vấn với từng đặc trưng lưu trong cơ sở dữ liệu. ảnh nào có độ tương đồng nhỏ nhất với ảnh cần tìm sẽ là ảnh kết quả. Thông thường nếu ảnh là tương đồng thì độ tương đồng thường <1 , còn lại là không tương đồng (đây là kết quả được lấy ra trong phần cài đặt chương trình).

Chương trình truy vấn ảnh thực hiện trên cơ sở dữ liệu gồm 70 ảnh. Để giảm độ phức tạp trong tính toán và tăng tốc độ chương trình ta nên chọn các ảnh nhỏ, dung lượng nhỏ. Khi thực hiện các thao tác với ảnh như trích chọn keypoint, xây dựng bộ mô tả...đều thực hiện trên ảnh xám, do đó ta cần chuyển đổi ảnh màu sang ảnh xám.

Cơ sở dữ liệu ảnh là ảnh màu, vì vậy sau khi đọc ảnh xong ta phải convert ảnh sang ảnh xám bằng lệnh `cvtColor(src2, img_object, CV_BGR2GRAY)`;

Trước tiên ta chọn ảnh cần truy vấn, chương trình sẽ trích chọn các đặc trưng của ảnh này và xây dựng bộ mô tả cho các đặc trưng này. Sau đó các keypoint sẽ được trích rút ra từ bộ mô tả và lưu lại trong một mảng.

Đối với cơ sở dữ liệu ảnh, thuật toán cũng quét từng ảnh, lưu lại đặc trưng cho từng bức ảnh và sau đó xây dựng bộ mô tả. Các keypoint của từng ảnh cũng được trích trọn và lưu lại trong từng mảng.

Khi đối sánh, ta duyệt từng keypoint, trước tiên ta tìm điểm láng giềng gần nhất với keypoint đang xét theo thuật toán láng giềng gần. Điểm láng giềng gần là điểm mà hàm tính khoảng cách `euclidDistance(Mat& vec1, Mat& vec2)` có giá trị nhỏ nhất và phải nhỏ hơn một ngưỡng `THRESHOLD` thì mới được chọn, trong chương trình này thì ngưỡng được gán giá trị 400.

Ảnh có độ tương đồng lớn nhất là ảnh có tỉ lệ phần trăm keypoint so khớp cao nhất. Thông thường, một ảnh là tương đồng nếu tỉ lệ phần trăm keypoint so khớp $\geq 50\%$, ngược lại là không tương đồng.

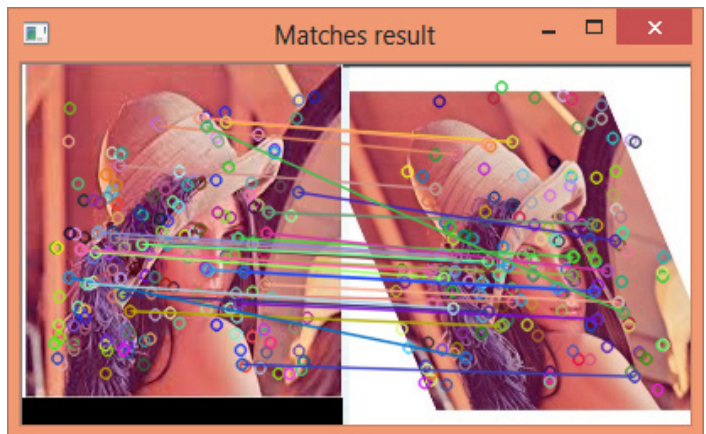
3. Kết luận

Chương trình thực hiện trên cơ sở dữ liệu có 70

ảnh và lọc ra 5 ảnh có độ tương đồng cao nhất với ảnh truy vấn. Kết quả lọc 5 ảnh có độ tương đồng cao nhất như sau:

```
Administrator: C:\Windows\system32\cmd.exe - image-search.exe 2 anh1.jpg l...
Extracting image Images\65.jpg
Extracting image Images\66.jpg
Extracting image Images\67.jpg
Extracting image Images\68.jpg
Extracting image Images\69.jpg
Extracting image Images\7.jpg
Extracting image Images\70.jpg
Extracting image Images\71.jpg
Extracting image Images\8.jpg
Extracting image Images\9.jpg
Extracting image Images\anh2.jpg
Extracting image Images\anh3.jpg
Extracting image Images\lena.jpg
1: <anh1.jpg, Images\anh3.jpg, 176, 168, 52, 0.178082>
2: <anh1.jpg, Images\28.jpg, 176, 151, 47, 0.167857>
3: <anh1.jpg, Images\57.jpg, 176, 327, 69, 0.158986>
4: <anh1.jpg, Images\37.jpg, 176, 148, 44, 0.157143>
5: <anh1.jpg, Images\56.jpg, 176, 327, 68, 0.156322>
result:5
```

Với độ tương đồng là 0.18.



Kết quả cho thấy thuật toán SURF có khả năng tốt khi truy vấn các ảnh có bất biến Affine tuy nhiên tốc độ của thuật toán còn chưa cao và thuật toán chưa thực sự hiệu quả với với các ảnh nhị phân.

Tài liệu tham khảo

[1] Arya, S., and Mount, D.M (1993). *Approximate nearest neighbor queries in fixed dimensions*. In *Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'93)*, pp. 271-280.

[2] Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., and Wu, A.Y. (1998). *An optimal algorithm for approximate nearest neighbor searching*. *Journal of the ACM*, 45:891-923.

[3] Ballard, D.H. (1981). *Generalizing the Hough transform to detect arbitrary patterns*. *Pattern Recognition*, 13(2):111-122.

[4] Lindeberg, T (1998): *Feature detection with automatic scale selection*. *IJCV* 30(2) 79 – 116