



NHẬN DẠNG MẶT NGƯỜI VỚI GIẢI THUẬT HAAR LIKE FEATURE – CASCADE OF BOOSTED CLASSIFIERS VÀ ĐẶC TRƯNG SIFT

Châu Ngân Khánh¹ và Đoàn Thanh Nghi²

¹ThS. Khoa Kỹ thuật-Công nghệ và Môi trường, Trường Đại học An Giang

²TS. Khoa Kỹ thuật-Công nghệ và Môi trường, Trường Đại học An Giang

Thông tin chung:

Ngày nhận bài: 03/03/14

Ngày nhận kết quả bình duyệt:
06/05/14

Ngày chấp nhận đăng:
30/07/14

Title:

Human face recognition by
Haar Like Feature – Cascade
of Boosted Classifiers
Algorithm and SIFT Features

Từ khóa:

Biểu diễn đặc trưng không đổi
SIFT, nhận dạng khuôn mặt,
Bayes thơ ngây với láng giềng,
đặc trưng Haar-like, k láng
giềng (kNN), mô hình phân
tầng cascade (CBC)

Keywords:

Scale-invariant feature
transform - SIFT, face
recognition, naive Bayes
nearest neighbor, Haar-like
features, k nearest neighbor
(kNN), cascade of boosted
classifiers (CBC)

ABSTRACT

Human face recognition is one of the very challenging problems in computer vision and machine learning. Recent studies have significantly increased the accuracy of recognition systems. In this paper, however, we present a new methodology that combines Haar Like Feature - Cascade of Boosted Classifiers and Scale-Invariant Feature Transform (SIFT) for the recognition of human face. Haar Like Features combine with the AdaBoost algorithm and the Cascade stratified model for fast and accurate detection and extraction of facial images. Reflection of the images is expressed by Scale-Invariant Feature Transform method, which is invariant to image scale, translational movement, rotation, partly to affine distortion, and strongly to illumination, addition of noise and obscuration. For object recognition, we propose the k nearest neighbors (kNN), the reversibility of kNN and Naive Bayes Nearest Neighbor (NBNN). Numerical testing on several benchmark datasets such as AT&T, Face94, Face95, Face96, Grimace, Jaffe using our proposed method for face recognition was more improved than the other methods and the accuracy levels obtained for kNN, reversed kNN, and NBNN are 94.35%, 86.05%, and 98.83%, respectively.

TÓM TẮT

Nhận dạng mặt người là một trong những vấn đề rất thách thức trong lĩnh vực thị giác máy tính và máy học. Các công trình nghiên cứu gần đây đã nâng cao độ chính xác của hệ thống nhận dạng. Tuy nhiên, trong bài này chúng tôi trình bày một phương pháp mới, kết hợp Haar Like Feature - Cascade of Boosted Classifiers (CBC) và các đặc trưng cục bộ không đổi (Scale-Invariant Feature Transform - SIFT) cho nhận dạng mặt người. Các đặc trưng Haar Like kết hợp thuật toán AdaBoost và mô hình phân tầng Cascade cho phép phát hiện và trích ảnh khuôn mặt nhanh và chính xác. Ảnh khuôn mặt được biểu diễn bằng các đặc trưng cục bộ không đổi (SIFT), được xem như bất biến đối với những biến đổi tỷ lệ ảnh, tịnh tiến, phép quay, không bị thay đổi một phần với phép biến đổi hình học affine (thay đổi góc nhìn) và mạnh với những thay đổi về độ sáng, sự nhiễu và che khuất. Để nhận dạng đối tượng, chúng tôi đề xuất sử dụng các thuật toán k láng giềng gần nhất (k Nearest Neighbor – kNN), kNN đảo ngược và Bayes thơ ngây với láng giềng gần nhất (Naive Bayes Nearest Neighbor – NBNN). Kết quả thử nghiệm trên nhiều tập dữ liệu kiểm chuẩn AT&T, Face94, Face95, Face96, Grimace, Jaffe cho thấy phương pháp nhận dạng mặt người được đề nghị bởi chúng tôi đạt được các kết quả tốt hơn so với các phương pháp khác, độ chính xác đạt được bởi kNN, kNN đảo và NBNN lần lượt là 94.35%, 86.05% và 98.83%.

1. GIỚI THIỆU

Cùng với sự bùng nổ thông tin, sự phát triển công nghệ cao, sự giao tiếp giữa con người và máy tính đang thay đổi rất nhanh, giờ đây giao tiếp này không còn đơn thuần dùng những thiết bị như chuột, bàn phím, v.v, mà có thể thông qua các biểu hiện của khuôn mặt. Các hệ thống giao tiếp người máy đang được phát triển rất nhiều. Trong số đó, có thể nói đến hệ thống nhận dạng mặt người bằng hình ảnh. Nhận dạng mặt người là xác định danh tính tự động cho từng ảnh đối tượng người dựa vào nội dung của ảnh. Nhận dạng mặt người được ứng dụng nhiều trong thực tế như xác minh tội phạm, camera chống trộm, hệ thống chấm công, lưu trữ thông tin khuôn mặt ở các máy ATM, các bãi giữ xe siêu thị, v.v.

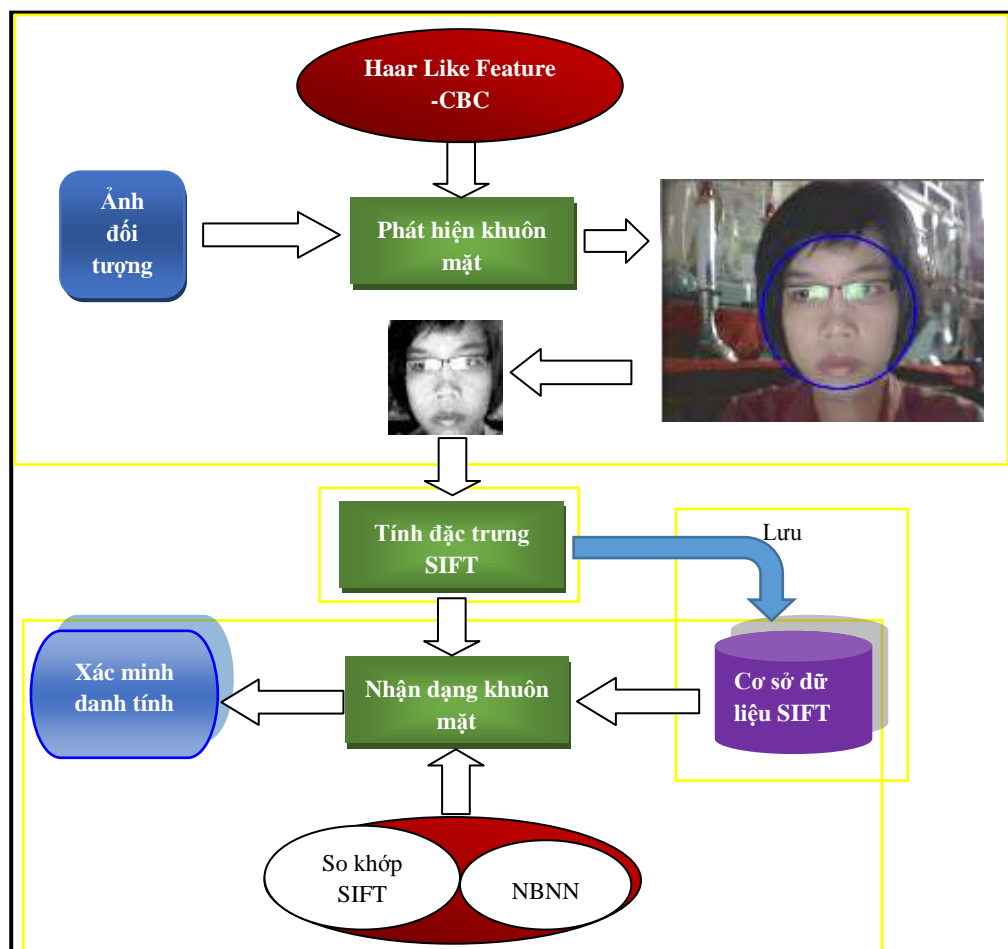
Hệ thống nhận dạng mặt người bao gồm hai bước: phát hiện khuôn mặt và định danh tự động đối tượng. Công việc chính dựa vào các kỹ thuật rút trích đặc trưng từ ảnh đối tượng và thực hiện đối sánh để định danh tự động. Hiệu quả của hệ thống nhận dạng phụ thuộc vào các phương pháp sử dụng.

Các nghiên cứu trước đây (Bledsoe, & cs., 1960; Goldstein, & cs., 1970), sử dụng tiếp cận dựa trên các đặc trưng như mắt, tai, màu tóc, độ dày môi để tự động nhận dạng (Ion Marqués, 2010). Kirby và Sirovich (1988); Turk và Pentland (1991) áp dụng phương pháp phân tích thành phần chính (PCA) và thuật toán eigenfaces để nhận dạng khuôn mặt. Trần Phước Long và Nguyễn Văn Lượng (2003), dùng mạng nơron để dò tìm khuôn mặt trong ảnh, kết hợp với phương pháp PCA và biến đổi cosine rời rạc (DCT) để rút ra các đặc trưng là đầu vào cho bộ nhận dạng máy học SVM và mô hình Markov ẩn HMM. Lu Boun Vinh và Hoàng Phương Anh (2004) sử dụng thuật toán AdaBoost để dò tìm khuôn mặt kết hợp với thuật toán FSVM để tiến hành nhận dạng mặt người. Zuo và cs. (2006) đã kết hợp phương pháp PCA hai chiều với phân tích độc lập tuyến tính (LDA); Chen và Yaou Zhao (2006) kết hợp DCT và

Hierarchical Radial Basis Function Network (HRBF); Sharif, Sajjad Mohsin, Muhammad Younas Javed, và Muhammad Atif Ali (2012) kết hợp thuật toán Laplacian of Gaussian (LoG) và DCT; Suhas, Kurhe Ajay B., và Dr. Prakash Khanale B (2012) áp dụng PCA kết hợp với DCT theo hướng tiếp cận toàn diện; Sunil (2013) kết hợp PCA với DCT; Bouzalmat, Jamal Kharroubi, và Arsalane Zarghili (2013) áp dụng thuật toán máy học SVM dựa trên LDA, để nhận dạng khuôn mặt người trên ảnh.

Gần đây, hướng tiếp cận dựa trên các đặc trưng cục bộ không đổi SIFT của Lowe (2004) được quan tâm nhiều. Đặc trưng cục bộ SIFT không bị thay đổi trước những biến đổi tỷ lệ ảnh, tịnh tiến, phép quay, không bị thay đổi một phần đối với phép biến đổi hình học affine (thay đổi góc nhìn) và mạnh với những thay đổi về độ sáng, sự nhiễu và che khuất. Aly (2006) sử dụng các đặc trưng SIFT để nhận dạng mặt người. Chennamma, Lalitha Rangarajan, và Veerabhadrappe (2011) áp dụng thuật toán so khớp SIFT để nhận dạng khuôn mặt trên tập ảnh đã được chọn lọc trước, đồng thời so sánh kết quả nhận dạng với thuật toán phân tích thành phần chính. Shinfeng, Jia-Hong Lin, và Cheng-Chin Chiang (2011) kết hợp phương pháp phân tích thành phần chính và đặc trưng SIFT để nhận dạng khuôn mặt người. Kumar và Padmavati (2012) đã đề xuất các cách tính khoảng cách khác nhau khi so khớp các đặc trưng SIFT để nhận dạng mặt người, góp phần cải thiện tốc độ so khớp các đặc trưng SIFT.

Trong bài báo này, chúng tôi đề xuất sử dụng các đặc trưng Haar Like với thuật toán AdaBoost và mô hình phân tầng Cascade để định vị khuôn mặt kết hợp với phương pháp biểu diễn ảnh bằng các đặc trưng bất biến SIFT và phương pháp đối sánh SIFT dựa trên k láng giềng (kNN), kNN đảo ngược (Jegou, Schmid, Harzallah, & Verbeek, 2011) và thuật toán NBNN (Boiman, 2008) để nhận dạng mặt người trực tuyến. Mô hình hệ thống như sau (Hình 1):



Hình 1: Hệ thống nhận dạng mặt người

Hệ thống hoạt động như sau: Sử dụng giải thuật Haar Like – CBC để phát hiện khuôn mặt người trên ảnh. Rút trích ảnh khuôn mặt người vừa phát hiện được ta sẽ thu được ảnh đối tượng. Tính đặc trưng SIFT của ảnh đối tượng. Tiến hành nhận dạng bằng cách so khớp SIFT dựa vào kNN, kNN đảo ngược hoặc sử dụng thuật toán NBNN.

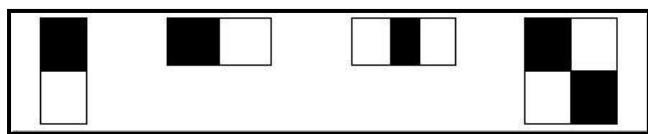
Phần tiếp theo của bài viết được trình bày như sau: phần 2 giới thiệu ngắn gọn về thuật toán phát hiện khuôn mặt Haar Like Features - Cascade of Boosted Classifiers, biểu diễn ảnh bằng các đặc trưng cục bộ không đổi SIFT được trình bày ngắn

gọn trong phần 3, phần 4 trình bày phương pháp định danh đối tượng dựa vào các đặc trưng SIFT, phần 5 trình bày các kết quả thực nghiệm trước khi vào phần kết luận và hướng phát triển.

2. ĐỊNH VỊ KHUÔN MẶT

2.1 Đặc trưng Haar Like

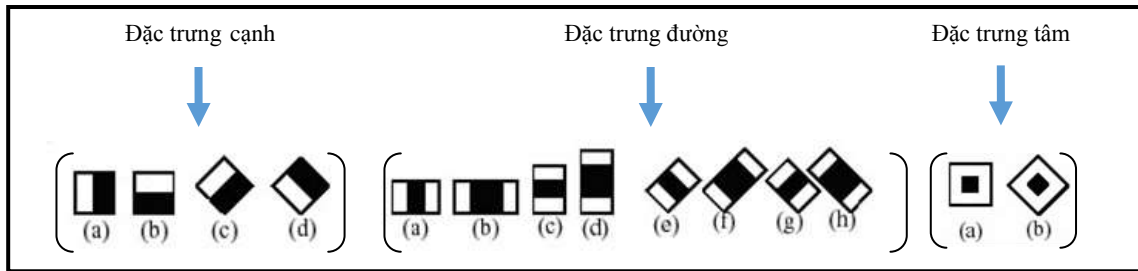
Đặc trưng Haar Like được tạo thành bằng việc kết hợp các hình chữ nhật đen, trắng với nhau theo một trật tự, một kích thước nào đó. Hình dưới đây mô tả 4 đặc trưng Haar Like cơ bản như sau:



Hình 2: Các đặc trưng Haar Like cơ bản

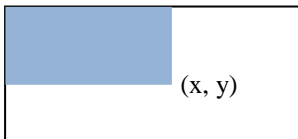
Để phát hiện khuôn mặt các đặc trưng Haar Like cơ bản trên được mở rộng (Lienhart, Kuranov, & Pisarevsky, 2002; Lienhart & Maydt, 2002) thành

nhóm các đặc trưng cạnh, đặc trưng đường và đặc trưng tâm (Hình 3).



Hình 3: Các đặc trưng Haar Like mở rộng

Giá trị của đặc trưng Haar Like là sự chênh lệch giữa tổng các điểm ảnh của các vùng đen và các vùng trắng. Để có thể tính nhanh các đặc trưng này, Viola và Jones (2001; 2004) giới thiệu khái niệm ảnh tích phân (Integral Image). Integral Image là một mảng hai chiều với kích thước bằng kích thước của ảnh cần tính giá trị đặc trưng Haar Like. Dưới đây là mô tả cách tính ảnh tích phân:

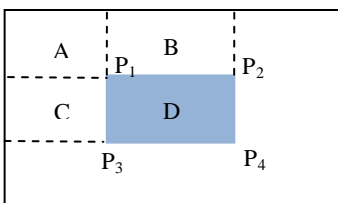


Hình 4: Tính giá trị ảnh tích phân tại điểm có tọa độ (x, y)

Giá trị của ảnh tích phân tại điểm P có tọa độ (x, y) được tính như sau:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (1)$$

Sau khi đã tính được ảnh tích phân, việc tính tổng các giá trị mức xám của một vùng ảnh bất kỳ nào đó trên ảnh thực hiện theo cách sau, ví dụ tính giá trị của vùng D trong hình 5 như sau:
 $D = A + B + C + D - (A + B) - (A + C) + A$



Hình 5: Tính nhanh giá trị của vùng ảnh D

Tiếp theo, sử dụng phương pháp máy học AdaBoost để xây dựng bộ phân loại mạnh với độ chính xác cao.

2.2 Thuật toán AdaBoost

AdaBoost (Freund & Schapire, 1995) là một bộ phân loại mạnh phi tuyến phức, hoạt động trên nguyên tắc kết hợp tuyến tính các bộ phân loại yếu để tạo nên một bộ phân loại mạnh. AdaBoost sử dụng trọng số để đánh dấu các mẫu khó nhận dạng. Trong quá trình huấn luyện cứ mỗi bộ phân loại yếu được xây dựng thì thuật toán sẽ tiến hành cập nhật lại trọng số để chuẩn bị cho việc xây dựng bộ phân loại tiếp theo. Cập nhật bằng cách tăng trọng số của các mẫu nhận dạng sai và giảm trọng số của các mẫu được nhận dạng đúng bởi bộ phân loại yếu vừa xây dựng. Bằng cách này thì bộ phân loại sau có thể tập trung vào các mẫu mà bộ phân loại trước nó làm chưa tốt. Cuối cùng các bộ phân loại yếu sẽ được kết hợp lại tùy theo mức độ tốt của chúng để tạo nên một bộ phân loại mạnh.

Bộ phân loại yếu h_k được biểu diễn như sau:

$$h_k(x) = \begin{cases} 1 & \text{nếu } p_k f_k(x) < p_k \theta_k \\ 0 & \text{nếu ngược lại} \end{cases} \quad (2)$$

Với x là cửa sổ con cần xét, θ_k là ngưỡng, f_k là giá trị đặc trưng Haar Like và p_k là hệ số quyết định chiều của phương trình.

2.3 Mô hình phân tầng Cascade

Cascade of Boosted Classifiers là mô hình phân tầng với mỗi tầng là một mô hình AdaBoost sử dụng bộ phân lớp yếu là cây quyết định với các đặc trưng Haar-Like.

Trong quá trình huấn luyện, bộ phân lớp phải duyệt qua tất cả các đặc trưng của mẫu trong tập huấn luyện. Việc này tốn rất nhiều thời gian. Tuy nhiên, trong các mẫu đưa vào, không phải mẫu nào cũng thuộc loại khó nhận dạng, có những mẫu

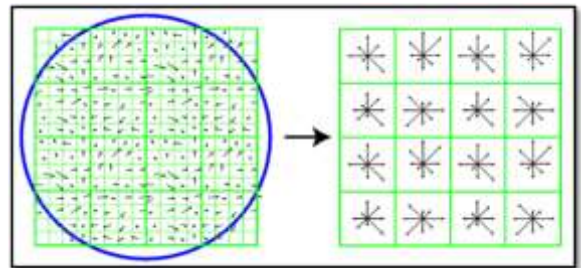
background rất dễ nhận ra (gọi đây những mẫu background đơn giản). Đối với những mẫu này, chỉ cần xét một hay một vài đặc trưng đơn giản là có thể nhận dạng được chứ không cần xét tất cả các đặc trưng. Nhưng đối với các bộ phân loại thông thường thì cho dù mẫu cần nhận dạng là dễ hay khó nó vẫn phải xét tất cả các đặc trưng mà nó rút ra được trong quá trình học. Do đó, chúng tốn thời gian xử lý một cách không cần thiết.

Mô hình Cascade of Classifiers được xây dựng nhằm rút ngắn thời gian xử lý, giảm thiểu nhận dạng lầm (false alarm) cho bộ phân loại. Cascade trees gồm nhiều tầng (stage hay còn gọi là layer), mỗi tầng là một mô hình AdaBoost với bộ phân lớp yếu là các cây quyết định. Một mẫu để được phân loại là đối tượng thì nó cần phải đi qua hết tất cả các tầng. Các tầng sau được huấn luyện bằng những mẫu âm negative (không phải mặt người) mà tầng trước nó nhận dạng sai, tức là nó sẽ tập trung học từ các mẫu background khó hơn, do đó sự kết hợp các tầng AdaBoost này lại sẽ giúp bộ phân loại giảm thiểu nhận dạng lầm. Với cấu trúc này, những mẫu background dễ nhận dạng sẽ bị loại ngay từ những tầng đầu tiên, giúp đáp ứng tốt nhất thời gian xử lý và vẫn duy trì được hiệu quả phát hiện khuôn mặt.

3. BIỂU DIỄN ĐẶC TRƯNG KHÔNG ĐỔI

Rút trích các đặc trưng ảnh là một bước quan trọng trong nhận dạng ảnh. Bước này giúp biểu diễn ảnh bằng các đặc trưng quan trọng mà giải thuật có thể thực hiện nhận dạng ảnh từ các đặc trưng này. Hai tiếp cận chính cho biểu diễn ảnh hiện nay là: sử dụng nét đặc trưng toàn cục (global features) như véc-tơ bitmap, tổ chức đồ màu (color histogram) và sử dụng nét đặc trưng cục bộ (local features) như điểm đặc trưng, vùng đặc trưng để biểu diễn ảnh. Tiếp cận thứ nhất đơn giản nhưng lại không thật sự hiệu quả vì cách biểu diễn này không thích hợp với những biến đổi về góc nhìn, biến đổi tỷ lệ, phép quay, độ sáng, sự che khuất, sự biến dạng, sự xáo trộn của hình nền và sự biến đổi trong nội bộ lớp. Ngược lại, tiếp cận thứ hai, đặc trưng cục bộ SIFT (Lowe, 2004) lại rất mạnh với những thách thức này và đạt được hiệu quả cao trong nhận dạng và tìm kiếm ảnh. Chính vì lý do đó, chúng tôi đề xuất sử dụng các nét đặc trưng cục bộ SIFT để biểu diễn ảnh phục vụ cho quá trình nhận dạng.

Các bước thực hiện rút trích đặc trưng SIFT được mô tả tóm tắt như sau. Ảnh được đưa về dạng mức xám. Các điểm đặc trưng được tính trên ảnh này bằng cách sử dụng các giải thuật phát hiện điểm đặc trưng cục bộ (local feature detector) như là Harris-Affine, Hessian-Affine. Những điểm đặc trưng này có thể là cực trị cục bộ của phép toán DoG (Difference of Gaussian) hoặc là cực đại của phép toán LoG (Laplace of Gaussian). Sau đó, vùng xung quanh các điểm đặc trưng được xác định và mô tả bằng các véc-tơ mô tả cục bộ. Véc-tơ mô tả SIFT được đánh giá rất cao bởi các chuyên gia trong việc biểu diễn các vùng xung quanh điểm đặc trưng bởi vì nó không đối đối với những biến đổi tỷ lệ, tịnh tiến, phép quay, và không đối một phần đối với những thay đổi về góc nhìn, đồng thời nó cũng rất mạnh với những thay đổi về độ sáng, sự che khuất và nhiễu.



Hình 6: Đặc trưng cục bộ SIFT được tính toán từ vùng xung quanh điểm đặc biệt (vòng tròn): gradient của ảnh (trái), véc-tơ mô tả (phải)

Hình 6 minh họa một ví dụ của véc-tơ mô tả SIFT được xây dựng từ vùng cục bộ xung quanh một điểm đặc trưng. Mỗi véc-tơ mô tả là một ma trận 4x4 các tổ chức đồ. Mỗi tổ chức đồ có 8 khoảng tương ứng với 8 hướng. Do đó, mỗi véc-tơ mô tả SIFT là một véc-tơ $4 \times 4 \times 8 = 128$ chiều. Lúc này, mỗi ảnh được biểu diễn bởi một tập các véc-tơ mô tả SIFT.

4. PHƯƠNG PHÁP ĐỊNH DANH

Trong bài báo này, chúng tôi đề xuất sử dụng thuật toán so khớp các véc-tơ mô tả SIFT của ảnh khuôn mặt dựa trên kNN, kNN đảo ngược và thuật toán NBNN.

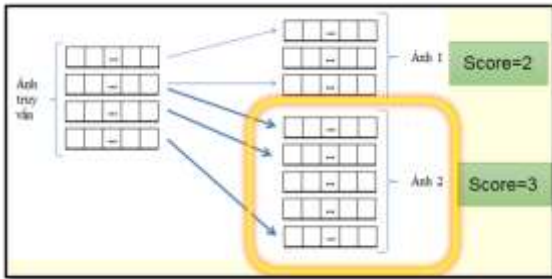
4.1 So khớp SIFT dựa trên kNN

Để đối sánh ảnh truy vấn với các ảnh khác trong cơ sở dữ liệu, trước tiên cần trích xuất tập đặc trưng từ ảnh truy vấn tương ứng, sau đó tiến hành so khớp các đặc trưng SIFT của ảnh truy vấn với tất cả các đặc trưng SIFT của các ảnh trong cơ sở

dữ liệu. Bước chính trong kỹ thuật đối sánh sẽ thực hiện tìm tập con đặc trưng so khớp nhau ở hai ảnh, để thực hiện việc này sẽ tìm các cặp đặc trưng trùng nhau lần lượt ở hai ảnh. Tập con các đặc trưng so khớp chính là vùng ảnh tương đồng.

Quá trình so khớp thực hiện như sau: với véc-tơ SIFT A, ta sử dụng giải thuật kNN để tìm B và C là hai véc-tơ SIFT gần A nhất và nhì. Nếu tỷ lệ khoảng cách của (A, B) và khoảng cách của (A, C) nhỏ hơn hoặc bằng 0.8 thì SIFT B được gọi là khớp với SIFT A (Lowe, 2004).

Thuật toán nhận dạng sẽ tìm ảnh của đối tượng có số lượng SIFT khớp với các SIFT của ảnh truy vấn nhiều nhất.

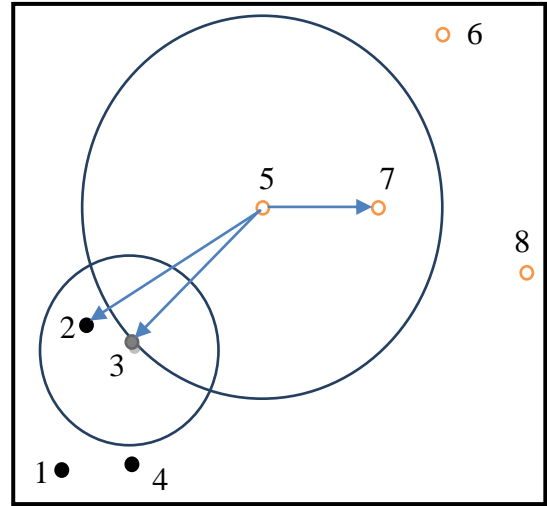


Hình 7: Minh họa so khớp SIFT

4.2 So khớp SIFT dựa trên kNN đảo ngược

Ý tưởng kNN đảo ngược đề xuất bởi Jegou và cs. (2011) có thể được tóm tắt như sau. B được gọi là láng giềng của A khi A cũng phải là láng giềng của B. Xét ví dụ minh họa trong hình 8. Cho véc-

tơ 5, 3 láng giềng của nó là véc-tơ 7, 3, 2. Trong khi đó, xét véc-tơ 3, 3 láng giềng của nó là 2, 4, 1. Hay nói cách khác, véc-tơ 3 thuộc 3 láng giềng của véc-tơ 5, trong khi véc-tơ 5 không thuộc 3 láng giềng của véc-tơ 3. Điều này chứng tỏ véc-tơ 3 không thật sự là láng giềng của véc-tơ 5.



Hình 8: Ví dụ về láng giềng không đảo ngược

Ở đây chúng tôi đề xuất thực hiện tìm láng giềng đảo ngược khi thực hiện so khớp SIFT để định danh. Thuật toán so khớp với kNN đảo ngược từ ảnh truy vấn Q đến ảnh I như sau:

- Bước 1: Tính toán tất cả các mô tả d_1, d_2, \dots, d_n của ảnh truy vấn Q.
- Bước 2: $\forall d_i, \forall I$, tìm láng giềng gần nhất và nhì của d_i trong I: $NN_{I_1}(d_i), NN_{I_2}(d_i)$
- Bước 3: $\forall NN_{I_j}(d_i), \forall Q$, tìm láng giềng gần nhất và nhì của $NN_{I_j}(d_i)$ trong Q: $d_1^*(NN_{I_j}(d_i)), d_2^*(NN_{I_j}(d_i))$
- Bước 4: $\forall NN_{I_j}(d_i)$, nếu $d_1^*(NN_{I_j}(d_i)) \equiv d_i$ hoặc $d_2^*(NN_{I_j}(d_i)) \equiv d_i$, thực hiện tiếp bước 5.
- Bước 5: nếu $\left(\frac{\|d_i - NN_{I_1}(d_i)\|}{\|d_i - NN_{I_2}(d_i)\|} \leq 0.8 \right)$, thì $NN_{I_1}(d_i)$ được xem là khớp với d_i .

Kết quả của thuật toán nhận dạng là ảnh của đối tượng có số lượng SIFT khớp với các SIFT của ảnh truy vấn nhiều nhất.

4.3 Phương pháp NBN (Naive Bayes Nearest Neighbours)

NBN được đề xuất bởi Boiman (2008) là phương pháp phân lớp ảnh bằng tính toán trực tiếp khoảng cách từ “ảnh truy vấn đến lớp” (từ ảnh truy vấn Q đến lớp C). Trước hết cần tính các

mô tả d_1, d_2, \dots, d_n của ảnh truy vấn Q. Tiếp đến, cần xác định láng giềng gần nhất của d_i trong toàn bộ các mô tả của lớp C là $NN_C(d_i)$. Ảnh Q được gán cho lớp C có khoảng cách từ d_i đến $NN_C(d_i)$ là nhỏ nhất. Thuật toán NBN được tóm tắt như sau:

- Bước 1: Tính toán tất cả các mô tả d_1, d_2, \dots, d_n của ảnh truy vấn Q.
- Bước 2: $\forall d_i \forall C$ tìm láng giềng gần nhất của d_i trong C: $NN_C(d_i)$.
- Bước 3: $\hat{C} = \arg \min_C \sum_{i=1}^n \|d_i - NN_C(d_i)\|^2$

5. THỰC NGHIỆM

Trong phần này, chúng tôi tiến hành đánh giá hiệu năng của hệ thống nhận dạng mặt người như đã đề xuất (sử dụng thuật toán Haar Like Feature – Cascade of Boosted Classifiers và các đặc trưng SIFT). Hệ thống nhận dạng này được cài đặt bằng ngôn ngữ lập trình C/C++, sử dụng thư viện mã nguồn mở OpenCV của Intel (Bradski & Kaehler, 2012; Laganière, 2011), trên một máy tính cá nhân chạy hệ điều hành Linux với bản phân phối Ubuntu.

Bước phát hiện mặt người thu được từ camera (webcam) sẽ được thực hiện thông qua việc huấn luyện mô hình phân tầng với mỗi tầng là một mô hình AdaBoost sử dụng bộ phân lớp yếu là cây quyết định với các đặc trưng Haar-Like (hỗ trợ bởi `opencv_createsamples` và `opencv_haartraining` của OpenCV) trên tập ảnh (mặt người và không phải mặt người).

Phương pháp của chúng tôi được đánh giá trên các tập dữ liệu kiểm chuẩn AT&T (ORL) được giới thiệu bởi AT&T Laboratories Cambridge (1994); Face94, Face95, Face96, Grimace được giới thiệu bởi Spacek (2007a; 2007b; 2007c; 2007d); Jaffe được giới thiệu bởi Lyons, Miyuki Kamachi, và Jiro Gyoba (1998). Đầu tiên, chương trình sử dụng mô hình phân tầng đã huấn luyện để phát hiện mặt người, rút trích ra khuôn mặt. Sau đó, chúng tôi sử dụng lớp `SiftFeatureDetector` và `SiftDescriptorExtractor` từ thư viện OpenCV để rút trích các đặc trưng SIFT của tất cả các ảnh

khuôn mặt (không phải ảnh gốc) và lưu vào cơ sở dữ liệu SIFT. Giải thuật kNN được sử dụng để tìm hai láng giềng gần nhất và láng giềng đảo ngược hoặc một láng giềng gần nhất tương ứng với thuật toán nhận dạng dựa trên so khớp SIFT và NBN.

Nghi thức kiểm tra trong thực nghiệm của chúng tôi là repeated hold-out (Đỗ Thanh Nghị, 2011; Đỗ Thanh Nghị & Phạm Nguyên Khang, 2012), lấy ngẫu nhiên 2/3 tập dữ liệu làm tập học (hay cơ sở dữ liệu đối tượng), 1/3 tập dữ liệu còn lại làm tập kiểm tra. Chúng tôi thực hiện việc kiểm thử 5 lần, sau đó tính trung bình cộng để xác định giá trị lỗi tổng thể. Tiếp theo, so sánh các kết quả thu được từ việc sử dụng giải thuật kNN, kNN đảo ngược, NBN. Việc rút trích được khuôn mặt người từ ảnh gốc đã giảm lược khá nhiều số lượng SIFT của đối tượng, nhờ vậy hệ thống đã tăng tốc đáng kể quá trình nhận dạng và đạt được độ chính xác cao hơn. Với ảnh có kích thước 500x500 pixels thì có khoảng 2000 SIFT. Nhưng khi trích xuất được khuôn mặt người trong ảnh thì số lượng SIFT trung bình còn khoảng 200 SIFT. Số lượng SIFT giảm đi không những không làm giảm độ chính xác của chương trình mà còn làm cho độ chính xác được tăng lên, vì các đặc trưng không hữu ích hoặc làm ảnh hưởng xấu đến kết quả nhận dạng đối tượng đã được loại bỏ. Thực nghiệm cho thấy số lượng SIFT không hữu ích là lớn hơn rất nhiều so với SIFT có ý nghĩa trong nhận dạng. Kết quả nhận dạng trên các tập dữ liệu kiểm thử được trình bày như trong Bảng 1.

Bảng 1: Tỷ lệ nhận dạng trên các tập dữ liệu

Tập dữ liệu	Số lượng phân lớp	Tổng số ảnh	Tỷ lệ nhận dạng		
			kNN (%)	kNN đảo ngược (%)	NBNN (%)
AT&T	40	400	97.62	93.86	99.37
Face94	152	3040	100.00	99.37	100.00
Face95	72	1440	96.21	92.25	98.83
Face96	151	3016	94.35	86.05	99.20
Grimace	18	360	99.00	99.17	100.00
Jaffe	10	213	97.18	90.00	100.00

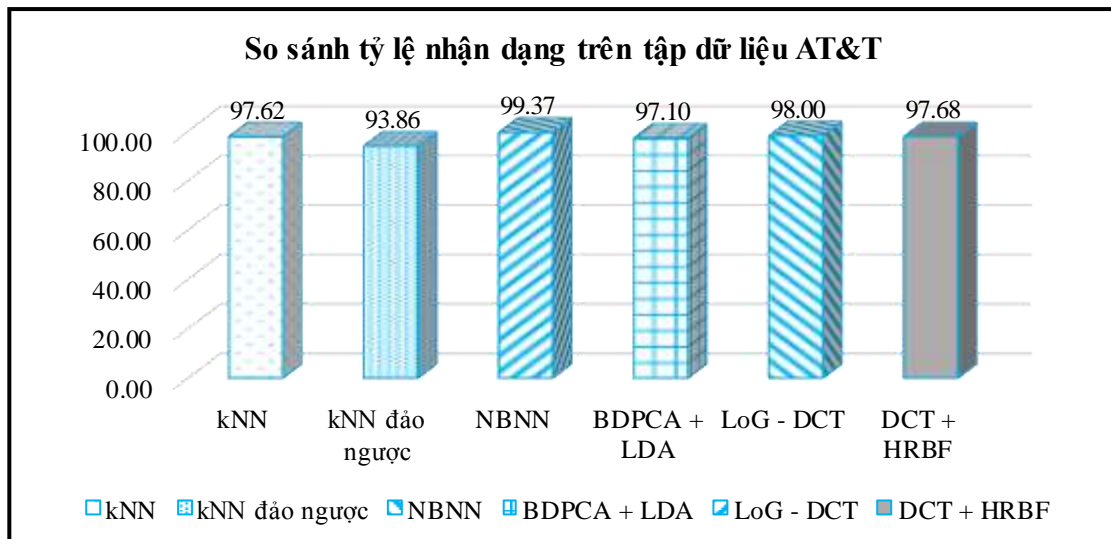
Để việc đánh giá được thuyết phục hơn, chúng tôi so sánh kết quả của ba phương pháp đề xuất với các phương pháp tốt nhất mà chúng tôi biết hiện nay (state-of-the-art) trong lĩnh vực nhận dạng mặt người, phương pháp BDPCA + LDA (Zuo & cs., 2006), LoG - DCT (Sharif & cs., 2012), DCT

+ HRBF (Chen & Yaou Zhao, 2006); trên cùng tập dữ liệu.

Kết quả nhận dạng khi tiến hành thực nghiệm trên tập dữ liệu AT&T được trình bày như trong Bảng 2 và Hình 9.

Bảng 2: Tỷ lệ nhận dạng trên tập dữ liệu AT&T

Phương pháp	Tỷ lệ nhận dạng (%)
kNN	97.62
kNN đảo ngược	93.86
NBNN	99.37
BDPCA + LDA	97.10
LoG - DCT	98.00
DCT + HRBF	97.68



Hình 9: Tỷ lệ nhận dạng trên tập dữ liệu AT&T

Từ bảng kết quả nhận dạng trên tập dữ liệu AT&T cho thấy độ chính xác của thuật toán so khớp SIFT dựa trên kNN và kNN đảo ngược cho kết quả khá cao, lần lượt là 97.62% và 93.86%. Trong

đó, hệ thống nhận dạng với thuật toán NBNN cho kết quả cao nhất là 99.37%.

Bảng kết quả tổng hợp khi thực nghiệm trên các tập dữ liệu kiểm chuẩn, đồng thời so sánh với các

phương pháp PCA + DCT (Sunil, 2013), SVM based on LDA (Linear and RBF kernel) (Bouzalmat & cs., 2013), PCA + LDA on Holistic Approach (Suhass & cs., 2012); được trình bày như Bảng 3.

Bảng 3: Tỷ lệ nhận dạng tổng hợp

Phương pháp	AT&T	Face94	Face95	Face96	Grimace	Jaffe
KNN	97.62	100	96.21	94.35	99.00	97.18
KNN đảo ngược	93.86	99.37	92.25	86.05	99.17	90.00
NBNN	99.37	100	98.83	99.20	100	100
BDPCA+LDA	97.10	-	-	-	-	-
LoG - DCT	98.00	-	-	-	-	-
DCT + HRBF	97.68	-	-	-	-	-
PCA + DCT	91.30	99.90	87.00	94.00	-	92.60
SVM based on LDA (Linear)	-	90.00	-	90.00	91.42	-
SVM based on LDA (RBF kernel)	-	97.14	-	95.10	100	-
PCA + LDA on Holistic Approach	-	-	100	-	100	-

Từ bảng kết quả nhận dạng tổng hợp cho thấy thuật toán so khớp dựa trên KNN và KNN đảo ngược cho kết quả khá cao, từ 94% và 86%. Thuật toán NBNN cho kết quả cao nhất trong hầu hết các tập dữ liệu (trên tập dữ liệu Face95, thuật toán NBNN cho kết quả cao thứ hai, sau thuật toán PCA + LDA on Holistic Approach). Thêm vào đó mô hình NBNN rất đơn giản vì không sử dụng bất kỳ tham số và cũng không cần quá trình huấn luyện (tương tự KNN, KNN đảo ngược). Điều này giúp cho hệ thống nhận dạng trở nên đơn giản hơn, dễ cài đặt nhưng lại cho kết quả với độ chính xác cao.

6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Chúng tôi vừa trình bày sự kết hợp phương pháp phát hiện mặt người với Haar Like Feature - Cascade of Boosted Classifiers và sự so khớp các đặc trưng cục bộ không đổi SIFT, cho nhận dạng mặt người. Kết quả thử nghiệm trên nhiều tập dữ liệu ảnh AT&T, Fae94, Face95, Face96, Grimace, Jaffe cho thấy hệ thống do chúng tôi đề xuất cho phép nhận dạng khuôn mặt người một cách hiệu quả về độ chính xác. Hơn nữa, việc rút trích khuôn mặt trước khi biểu diễn ảnh bằng các đặc trưng không đổi SIFT là rất cần thiết và hiệu quả: vì nếu chỉ xử lý ảnh thô thì chương trình sẽ hoàn toàn không nhận thấy được các điểm đặc biệt của khuôn mặt và do đó khả năng nhận dạng sai là rất cao. Điều này đã giúp bước phát hiện mặt người cải thiện đáng kể tốc độ cũng như độ chính xác của thuật toán nhận dạng.

Công việc kế tiếp của chúng tôi sẽ là thực hiện đánh giá thời gian thực hiện của các thuật toán nhận dạng trên các tập dữ liệu khác nhau. Đồng thời có thể kết hợp với các hệ thống điều khiển

Robot để nhận dạng đối tượng như: Nhận dạng chủ nhân của Robot, nhận dạng và định danh tội phạm, v.v.

TÀI LIỆU THAM KHẢO

- Aly, M. (2006). Face Recognition Using SIFT Features. *Technical Report*, Caltech.
- AT&T Laboratories Cambridge as Olivetti Research Laboratory. (1994). UK.
- Boiman, O. (2008). In Defense of Nearest-Neighbor Based Image Classification. In CVPR.
- Bouzalmat, A., Jamal Kharroubi & Arsalane Zarghili. (2013). Face Recognition Using SVM Based on LDA. *IJCSI International Journal of Computer Science Issues*, 10(4). ISSN (Print) : 1694-0814 | ISSN (Online) : 1694-0784.
- Bradski, A., & Kaehler, A. (2012). *Learning OpenCV*. O'Reilly Media.
- Chen, Y., & Yaou Zhao. (2006). *Face Recognition Using DCT and Hierarchical RBF Model*. School of Information Science and Engineering Jinan University, Jinan 250022, P.R. China.
- Chennamma H. R., Lalitha Rangarajan., & Veerabhadrapa. (2011). *Face Identification from Manipulated Facial Images using SIFT*. Department of Studies in Computer Science University of Mysore, Mysore, India.
- Đỗ Thanh Nghị. (2011). *Khai mở dữ liệu*. Cần Thơ: Nhà xuất bản Đại học Cần Thơ.
- Đỗ Thanh Nghị., & Phạm Nguyên Khang. (2012). *Nguyên lý máy học*. Cần Thơ: Nhà xuất bản Đại học Cần Thơ.
- Freund, Y., & Robert E.Schapire. (1995). *A decision-theoretic generalization of on-line learning and an application to boosting*. In Computational Learning Theory: Eurocolt '95. Springer-Verlag.
- Jegou, H., Schmid, C., Harzallah, H., & Verbeek, J.. (2011). *Accurate image search using the contextual dissimilarity measure*. Inria Grenoble, St. Ismier, France.

- Kirby, M., & Sirovich, L. (1988). A Low Dimensional Procedure for the Characterization of Human Faces. *J. Optical Soc. Am. A*, 4(3), 519-524.
- Kumar, H., & Padmavati. (2012). Face Recognition using SIFT by varying Distance Calculation Matching Method. *International Journal of Computer Application*, 47(3).
- Laganière, R. (2011). *OpenCV 2 Computer Vision Application Programming Cookbook*. Packt Publishing Ltd.
- Lienhart, R., & Maydt, R. (2002). *An Extended Set of Haar-like Features for Rapid Object Detection*. Intel Labs, Intel Corporation, Santa Clara, CA 95052, USA.
- Lienhart, R., Kuranov, A., & Pisarevsky, V. (2002). *Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection*. Microprocessor Research Lab, Intel Labs, Intel Corporation, Santa Clara, CA 95052, USA.
- Lowe, D. G. (1999). *Object Recognition from Local Scale-Invariant Features*. *International Conference on Computer Vision*, Corfu, Greece.
- Lowe, D. G. (2004). Distinctive image features from Scale-Invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- Lu Boun Vinh., & Hoàng Phương Anh. (2004). *Nghiên cứu và xây dựng hệ thống nhận dạng mặt người dựa trên FSVM và AdaBoost*. Thành phố Hồ Chí Minh, Việt Nam.
- Lyons, M., Miyuki Kamachi., & Jiro Gyoba. (1997). *Japanese Female Facial Expression (JAFFE)*, Database of digital images.
- Marqués, I. (2010). Face Recognition Algorithms, Universidad del País Vasco.
- Sharif, M., Sajjad Mohsin., Muhammad Younas Javed., & Muhammad Atif Ali. (2012). Single Image Face Recognition Using Laplacian of Gaussian and Discrete Cosine Transforms. *The International Arab Journal of Information Technology*, 9(6).
- Shinfeng D. Lin., Jia-Hong Lin., & Cheng-Chin Chiang. (2011). Using Gradient Features from Scale-Invariant Keypoints on Face Recognition. *International Journal of Innovative Computing, Information and Control*, 7(4).
- Spacek, L. (2007a). Computer Vision Science Research ml. Projects, <http://cswww.essex.ac.uk/mv/allfaces/faces94.html>.
- Spacek, L. (2007b). Computer Vision Science Research ml. Projects, <http://cswww.essex.ac.uk/mv/allfaces/faces95.html>.
- Spacek, L. (2007c). Computer Vision Science Research ml. Projects, <http://cswww.essex.ac.uk/mv/allfaces/faces96.html>.
- Spacek, L. (2007d). Computer Vision Science Research ml. Projects, <http://cswww.essex.ac.uk/mv/allfaces/grimace.html>.
- Suhas, S. Satonkar., Kurhe Ajay B., Dr. Prakash Khanale B. (2012). Face Recognition Using Principal Component Analysis and Linear Discriminant Analysis on Holistic Approach in Facial Images Database. *IOSR Journal of Engineering*, 2(12), 15-23.
- Sunil, Renke. Pradnya. (2013). Automatic Face Recognition using Principal Component Analysis with DCT. *Journal of Electronics and Communication Engineering (IOSR-JECE)*, ISSN: 2278-2834-, ISBN: 2278-8735, PP: 01-07. India.
- Trần Phước Long., & Nguyễn Văn Lượng. (2003). *Nhận dạng người dựa vào thông tin khuôn mặt xuất hiện trên ảnh*. Thành phố Hồ Chí Minh, Việt Nam.
- Viola, P., & Jones, M. (2001). Robust Real-time Object Detection. *International Journal of Computer Vision*.
- Viola, P., & Jones, M. (2004). Robust Real-time Face Detection. *International Journal of Computer Vision*, Kluwer Academic Publishers, Netherlands.
- Zuo, W., Zhang, D., Senior Member, IEEE, Jian Yang, & Kuanquan Wang. BDPCA Plus LDA. (2006). A Novel Fast Feature Extraction Technique for Face Recognition. *IEEE Transactions on Systems, Man, And Cybernetics – Part B: Cybernetics*, 36(4).