

# ĐIỂM DANH BẰNG MẶT NGƯỜI VỚI ĐẶC TRƯNG GIST VÀ MÁY HỌC VÉCTƠ HỖ TRỢ

Phạm Nguyên Khang, Trần Nguyễn Minh Thư, Đỗ Thanh Nghị

Khoa CNTT-TT, Trường Đại học Cần Thơ  
Khu 2, Đường 3/2, Xuân Khánh, Ninh Kiều, TP. Cần Thơ

pnkhang@cit.ctu.edu.vn

**TÓM TẮT:** Trong bài viết này, chúng tôi đề xuất hệ thống điểm danh bằng mặt người với máy học véctơ hỗ trợ (Support Vector Machines - SVM) sử dụng đặc trưng GIST. Hệ thống điểm danh thực hiện hai bước chính là định vị khuôn mặt trong ảnh thu được từ camera và định danh đối tượng từ ảnh khuôn mặt. Bước định vị khuôn mặt được thực hiện dựa trên các đặc trưng Haar-like kết hợp với mô hình phân tầng (Cascade of Boosted Classifiers - CBC). Chúng tôi đề xuất huấn luyện mô hình máy học SVM sử dụng đặc trưng GIST để thực hiện định danh đối tượng từ ảnh khuôn mặt. Kết quả thực nghiệm trên tập dữ liệu gồm 6722 ảnh của 132 đối tượng là những sinh viên Khoa CNTT-TT, Trường Đại học Cần Thơ cho thấy máy học SVM sử dụng đặc trưng GIST đạt đến 99.29% độ chính xác trên tập kiểm tra, cao hơn khi so với mô hình mạng nơron tích chập (Convolutional Neural Network - CNN), máy học SVM sử dụng mô hình túi từ (Bag-of-Words - BoW) của đặc trưng SIFT (Scale-Invariant Feature Transform), Bayes thơ ngây với láng giềng gần nhất (Naive Bayes Nearest Neighbor - NBNN) sử dụng đặc trưng SIFT có độ chính xác lần lượt là 96.88%, 97.54% và 98.88%.

**Từ khóa:** Nhận dạng mặt người, đặc trưng GIST, máy học véctơ hỗ trợ (SVM).

## I. GIỚI THIỆU

Hệ thống điểm danh bằng mặt người thực hiện rút trích tự động khuôn mặt người trong ảnh thu được từ camera (webcam) và xác định danh tính của đối tượng trong hệ thống dựa vào nội dung của ảnh khuôn mặt rút trích được. Nhận dạng tự động khuôn mặt có thể ứng dụng nhiều trong thực tiễn như camera giám sát, hệ thống chấm công, chú thích ảnh tự động, lưu trữ thông tin khuôn mặt ở các máy ATM, tìm kiếm và xác minh tội phạm. Hệ thống điểm danh thực hiện nhận dạng khuôn mặt người qua 2 bước chính: định vị khuôn mặt trong ảnh thu được từ camera và định danh đối tượng từ ảnh khuôn mặt.

Trong bài báo của [Zhao et al., 2003] trình bày các nghiên cứu trước đây trong nhận dạng khuôn mặt người. Rút trích tự động khuôn mặt trong ảnh thu được từ camera dựa trên mô hình mạng nơron [Vaillant et al., 1993] hoặc đặc trưng haar-like kết hợp với mô hình phân tầng (Cascade of Boosted Classifiers - CBC) [Viola & Jones, 2001]. Để xác định danh tính của đối tượng trong hệ thống dựa vào nội dung của ảnh khuôn mặt, nghiên cứu trước đây sử dụng các đặc trưng như mắt, tai, màu tóc, độ dày môi để tự động nhận dạng. [Belhumeur et al., 1996], [Turk & Pentland, 1991] sử dụng phương pháp phân tích thành phần chính (Principal Component Analysis - PCA) và thuật toán eigenfaces để nhận dạng khuôn mặt.

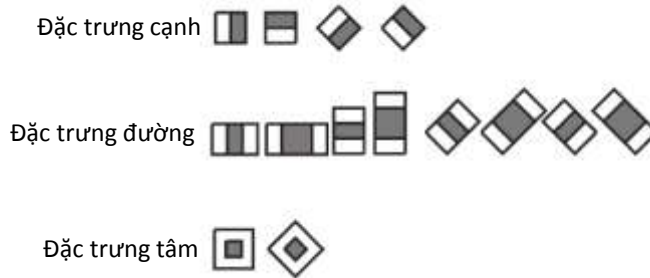
Các hướng nghiên cứu gần đây [Aly, 2006], [Bosch et al., 2006], [Geng & Jiang, 2009], [Do, 2011], [Châu et al., 2013, 2014] dựa vào phương pháp biểu diễn ảnh bằng các nét đặc trưng không đổi với những biến đổi tỉ lệ (Scale-Invariant Feature Transform - SIFT [Lowe, 2004]). Đặc trưng cục bộ SIFT không bị thay đổi trước những biến đổi tỉ 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, nhiều và sự che khuất. Để định danh đối tượng từ ảnh khuôn mặt người, [Aly, 2006] đề xuất so khớp các đặc trưng SIFT trong khi nghiên cứu của [Châu et al., 2013, 2014] kết hợp giải thuật Bayes thơ ngây với láng giềng gần nhất (Naive Bayes Nearest Neighbor - NBNN [Boiman et al., 2008]) và đặc trưng SIFT. [Bosch et al., 06], [Do, 2011] đề xuất huấn luyện mô hình máy học, chẳng hạn SVM [Vapnik, 1995] trên mô hình biểu diễn túi từ (Bag-of-Words - BoW được tạo từ các đặc trưng SIFT) cho phân lớp ảnh. Nghiên cứu của [Durr et al., 2015], [Parkhi et al., 2015] đề xuất xây dựng mô hình học sâu mạng nơron tích chập (Convolutional Neural Network - CNN [LeCun & Bengio, 1995], [LeCun et al., 2015]).

Trong bài viết này, chúng tôi đề xuất hệ thống điểm danh bằng mặt người sử dụng đặc trưng GIST [Oliva & Torralba, 2001] và máy học SVM [Vapnik, 1995]. Bước định vị khuôn mặt trong ảnh thu được từ camera được thực hiện tự động bằng mô hình phân tầng CBC của [Viola & Jones, 2001] sử dụng các đặc trưng Haar-like. Tiếp theo, hệ thống thực hiện rút trích đặc trưng GIST từ ảnh khuôn mặt và huấn luyện mô hình máy học SVM để định danh đối tượng từ ảnh khuôn mặt. Kết quả thực nghiệm trên tập dữ liệu gồm 6722 ảnh của 132 đối tượng là những sinh viên Khoa CNTT-TT, Trường Đại học Cần Thơ cho thấy hệ thống chúng tôi đề xuất (SVM-GIST) đạt đến 99.29% độ chính xác trên tập kiểm tra, cao hơn khi so với mô hình mạng nơron tích chập (CNN), máy học SVM sử dụng mô hình túi từ của đặc trưng SIFT (SVM-BoW-SIFT), Bayes thơ ngây với láng giềng gần nhất sử dụng đặc trưng SIFT (NBNN-SIFT) có độ chính xác lần lượt là 96.88%, 97.54% và 98.88%.

Phần còn lại của bài viết được tổ chức như sau. Chúng tôi sẽ trình bày hệ thống điểm danh bằng mặt người với đặc trưng GIST và giải thuật máy học SVM trong phần 2. Kết quả thực nghiệm sẽ được trình bày trong phần 3 trước khi kết luận và hướng phát triển được trình bày trong phần 4.



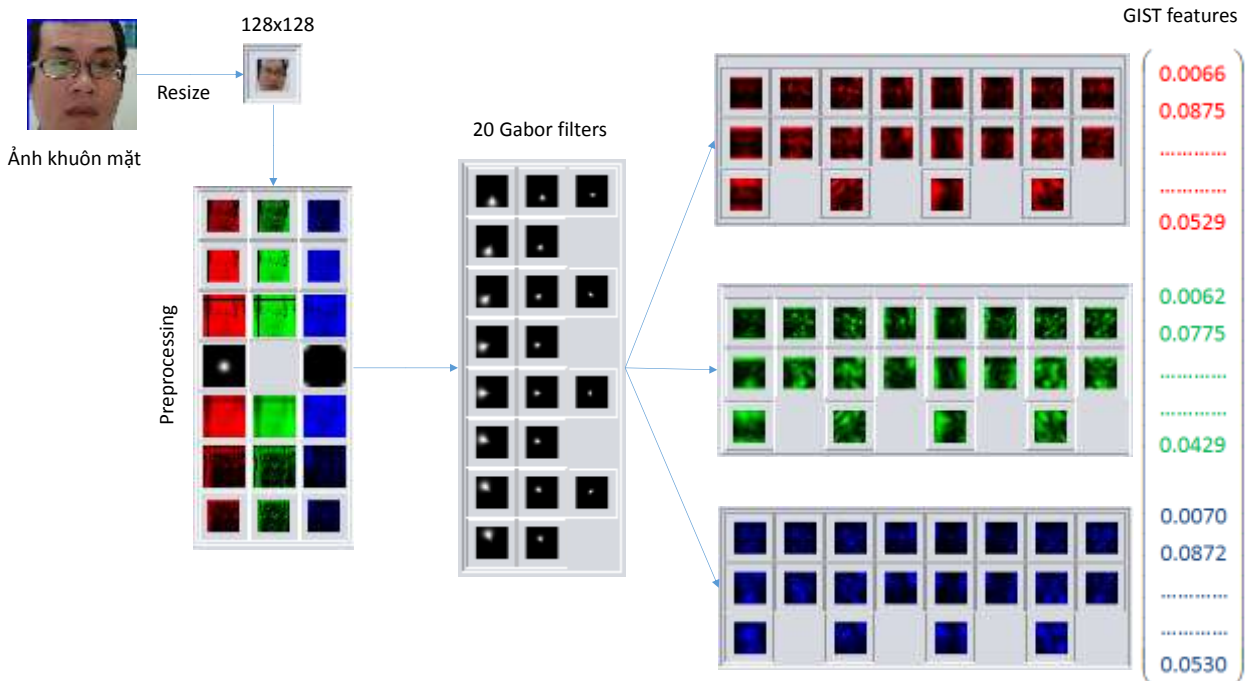
CBC, nếu ở tầng thứ  $t$  mô hình  $h_t$  phân lớp vùng ứng viên không phải là khuôn mặt người thì cửa sổ trượt tiếp tục đến vị trí tiếp theo trên ảnh, nếu vùng ứng viên được mô hình  $h_t$  ở tầng thứ  $t$  phân lớp là khuôn mặt người thì vùng ứng viên tiếp tục chuyển đến tầng thứ  $t+1$  để xét vùng ứng viên có phải mặt người hay không, quá trình tiếp tục cho đến khi tầng  $N$  mô hình  $h_N$  phân lớp là khuôn mặt người thì vùng ứng viên được xác định là khuôn mặt người.



**Hình 3.** Đặc trưng Haar-like

Ưu điểm của phương pháp đề xuất bởi [Viola & Jones, 2001] là việc rút trích đặc trưng Haar-like nhanh và đơn giản về tính toán so với các phương pháp rút trích đặc trưng khác. Ngoài ra, mô hình phân tầng CBC đã loại bỏ hầu hết các vùng ứng viên không phải là khuôn mặt từ các tầng đầu tiên, chỉ xét rất ít vùng ứng viên ở các tầng cuối. Chính vì vậy mô hình phân tầng CBC sử dụng đặc trưng Haar-like định vị rất nhanh và chính xác khuôn mặt trong ảnh thu được từ camera.

**B. Rút trích đặc trưng GIST từ ảnh khuôn mặt**



**Hình 4.** Rút trích đặc trưng GIST từ ảnh khuôn mặt

Một khi đã định vị được khuôn mặt trong ảnh, bước tiếp theo là rút trích đặc trưng quan trọng để có thể xây dựng mô hình nhận dạng (định danh) đối tượng. Chúng tôi đề xuất sử dụng đặc trưng GIST [Oliva & Torralba, 2001] do sự đơn giản và hiệu quả của nó được minh chứng trong các vấn đề tìm kiếm ảnh [Douze et al., 2009], phân lớp ảnh, nhận dạng chữ viết tay [Do & Pham, 2015]. Đặc trưng GIST được trình bày theo định dạng véctơ có 960 thành phần (chiều) được tính toán từ kết quả của việc áp dụng 20 bộ lọc Gabor lên ảnh. GIST có thể rút trích tập hợp các đặc trưng quan trọng như: tính tự nhiên (degree of naturalness), tính cởi mở (degree of openness), tính mở rộng (degree of expansion), độ nhám (degree of roughness) và độ chắc chắn (degree of ruggedness), cho phép trình bày cấu trúc không gian của một cảnh. Hơn nữa, đặc trưng GIST được trích chọn theo các hướng của các tần số khác nhau từ đó làm nổi bật thêm cấu trúc của ảnh.

Hình 4 mô tả các bước rút trích đặc trưng GIST từ ảnh khuôn mặt. Trước tiên, ảnh khuôn mặt cần chuẩn hóa về kích thước 128x128 và chuyển sang bước:

- Tiền xử lý: tách 3 kênh màu (đỏ, xanh lá cây, xanh dương), biến đổi tỷ lệ lôgarit (giảm chênh lệch tỷ lệ các điểm ảnh tối so với các điểm ảnh sáng), thêm các điểm ảnh biên, lọc trắng ảnh (cân bằng năng lượng phổ), lọc chuẩn hóa độ tương phản, xóa các điểm ảnh biên,

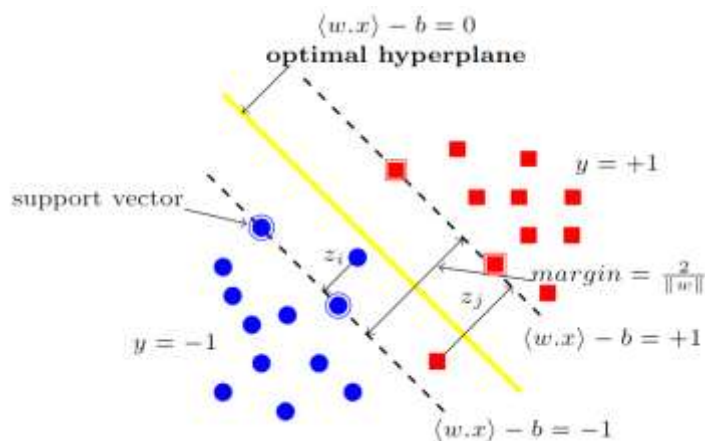
- Sinh 20 bộ lọc Gabor để áp dụng lên từng ảnh (đỏ, xanh lá cây, xanh dương, đã được tiền xử lý và chuyển về miền tần số),

- Chia ảnh thành 16 vùng riêng biệt bằng nhau, tính giá trị trên mỗi vùng bằng cách lấy tổng giá trị của các điểm ảnh trên mỗi vùng chia cho số điểm ảnh của vùng, thực hiện lần lượt 20 bộ lọc trên 16 vùng của ảnh đỏ, xanh lá cây, xanh dương, thu được vectơ có 960 thành phần (chiều).

### C. Máy học vectơ hỗ trợ SVM

Bước rút trích đặc trưng GIST từ ảnh khuôn mặt tạo ra tập dữ liệu có  $m$  dòng (ảnh khuôn mặt), mỗi dòng có 960 chiều (đặc trưng GIST) và nhãn (đối tượng trong hệ thống). Vì máy học SVM [Vapnik, 1995] là một trong top 10 giải thuật phân lớp hiệu quả, phổ biến trong cộng đồng khám phá tri thức và khai thác dữ liệu [Wu & Kumar, 2009], chính vì vậy chúng tôi đề xuất huấn luyện mô hình máy học SVM để nhận dạng (định danh) các đối tượng trong hệ thống.

Xét ví dụ phân lớp nhị phân tuyến tính đơn giản được mô tả như hình 5, với  $m$  phần tử  $x_1, x_2, \dots, x_m$  trong không gian  $n$  chiều với nhãn (lớp) của các phần tử tương ứng là  $y_1, y_2, \dots, y_m$  có giá trị  $+1$  (lớp dương) hay giá trị  $-1$  (lớp âm). Giải thuật máy học SVM [Vapnik, 1995] tìm siêu phẳng tối ưu (xác định bởi vectơ pháp tuyến  $w$  và độ lệch của siêu phẳng với gốc tọa độ  $b$ ) để tách dữ liệu ra 2 lớp. Máy học SVM tìm siêu phẳng tách 2 lớp ra xa nhất có thể (siêu phẳng tối ưu) dựa trên 2 siêu phẳng hỗ trợ song song của 2 lớp. Siêu phẳng hỗ trợ của lớp  $+1$  ( $w \cdot x - b = +1$ ) là siêu phẳng mà các phần tử  $x_p$  thuộc lớp  $y_p = +1$  nằm về phía bên phải của nó, tức là:  $w \cdot x_p - b \geq +1$ . Tương tự, siêu phẳng hỗ trợ của lớp  $-1$  ( $w \cdot x - b = -1$ ) là siêu phẳng mà các phần tử  $x_n$  thuộc lớp  $y_n = -1$  nằm về phía bên trái siêu phẳng hỗ trợ lớp  $-1$ , tức là:  $w \cdot x_n - b \leq -1$ . Những phần tử nằm ngược phía với siêu phẳng hỗ trợ được coi như lỗi. Khoảng cách lỗi được biểu diễn bởi  $z_i \geq 0$  (với  $x_i$  nằm đúng phía của siêu phẳng hỗ trợ của nó thì khoảng cách lỗi tương ứng  $z_i = 0$ , còn ngược lại thì  $z_i > 0$  là khoảng cách từ điểm  $x_i$  đến siêu phẳng hỗ trợ tương ứng của nó). Khoảng cách giữa 2 siêu phẳng hỗ trợ được gọi là lề  $= 2/\|w\|$ , trong đó  $\|w\|$  là độ lớn (2-norm) của pháp vectơ  $w$ . Siêu phẳng tối ưu (nằm giữa 2 siêu phẳng hỗ trợ) cần tìm phải cực đại hóa lề (lề càng lớn, mô hình phân lớp càng an toàn) và cực tiểu hóa lỗi. Vấn đề tìm siêu phẳng tối ưu của giải thuật SVM dẫn đến việc giải bài toán quy hoạch toàn phương (1):



Hình 5. Phân lớp tuyến tính với máy học SVM

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m y_i y_j \alpha_i \alpha_j K \langle x_i, x_j \rangle - \sum_{i=1}^m \alpha_i \quad (1)$$

với ràng buộc:

$$\begin{cases} \sum_{i=1}^m \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C \quad " i = 1, 2, \dots, m \end{cases}$$

trong đó  $C$  là hằng số dương dùng để điều chỉnh độ rộng của lề phân hoạch và tổng khoảng cách lỗi;  $K \langle x_i, x_j \rangle$  là hàm nhân tuyến tính  $K \langle x_i, x_j \rangle = \langle x_i \bullet x_j \rangle$ .

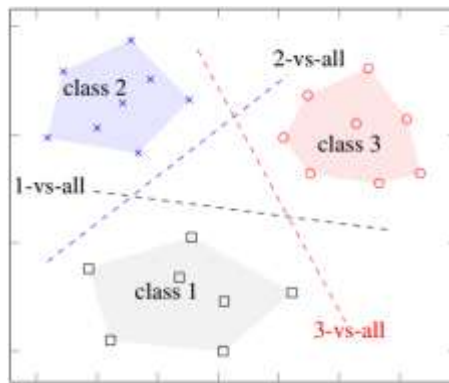
Giải bài toán quy hoạch toàn phương (1) thu được #SV phần tử  $x_i$  tương ứng với  $\alpha_i > 0$ , được gọi là các véctơ hỗ trợ. Chỉ cần #SV véctơ hỗ trợ này ta có thể dựng lại được siêu phẳng phân lớp. Mô hình SVM thực hiện phân lớp phần tử mới  $x$  bằng (2):

$$\text{predict}(x) = \text{sign}\left(\sum_{i=1}^{\#SV} y_i \alpha_i K\langle x, x_i \rangle - b\right) \quad (2)$$

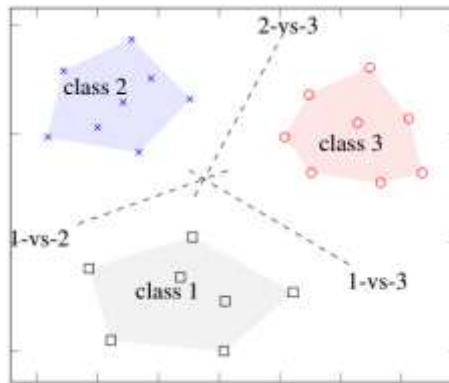
Máy học SVM có thể sử dụng các hàm nhân khác nhau để giải quyết lớp các bài toán phân lớp phi tuyến [Cristianini & Shawe-Taylor, 2000]. Để xử lý các vấn đề phân lớp phi tuyến, không cần bất kỳ thay đổi nào hơn từ giải thuật mà chỉ cần thay thế hàm nhân tuyến tính trong (1) và (2) bằng các hàm nhân khác. Hàm nhân phi tuyến phổ biến là hàm cơ sở bán kính (Radial Basic Function – RBF):  $K\langle x_i, x_j \rangle = e^{-\gamma \|x_i - x_j\|^2}$ .

Mô hình máy học SVM cho kết quả cao, ổn định, chịu đựng nhiễu tốt và phù hợp với các bài toán phân lớp, hồi quy. Nhiều ứng dụng thành công của SVM đã được công bố trong nhiều lĩnh vực như nhận dạng ảnh, phân loại văn bản và sinh-tin học [Guyon, 1999].

Để giải quyết vấn đề phân lớp đa lớp (số lớp  $c \geq 3$ ), giải thuật SVM thường mở rộng theo 2 phương pháp đơn giản là 1-vs-all [Vapnik, 95] và 1-vs-1 [Krebel, 99].



Hình 6. Phương pháp 1-vs-all cho máy học SVM đa lớp



Hình 7. Phương pháp 1-vs-1 cho máy học SVM đa lớp

Phương pháp 1-vs-all (hình 6) xây dựng  $c$  mô hình SVM nhị phân, mô hình thứ  $c_i$  tách lớp  $c_i$  (lớp dương) ra khỏi các lớp khác (âm). Phương pháp 1-vs-1 (hình 7) xây dựng  $c(c-1)/2$  mô hình SVM nhị phân, mỗi mô hình tách một cặp 2 lớp. Việc phân lớp dựa vào bình chọn khoảng cách đến các siêu phẳng thu được từ SVM nhị phân. LibSVM [Chang & Lin, 2011] cài đặt phương pháp 1-vs-1 trong khi LIBLINEAR [Fan et al., 2008] cài đặt phương pháp 1-vs-all cho SVM đa lớp.

### III. KẾT QUẢ THỰC NGHIỆM

#### A. Chương trình

Để đánh giá hiệu quả của hệ thống điểm danh được đề xuất, chúng tôi cài đặt chương trình để định vị khuôn mặt trong ảnh thu được từ camera với mô hình phân tầng CBC sử dụng đặc trưng Haar-like bằng ngôn ngữ lập trình C/C++, thư viện OpenCV [Bradski & Kaehler, 2012]. Rút trích đặc trưng GIST từ ảnh khuôn mặt được cài đặt bằng

ngôn ngữ lập trình C/C++. Thư viện LibSVM [Chang & Lin, 2011] được sử dụng để huấn luyện mô hình nhận dạng đối tượng. Chúng tôi cũng muốn so sánh phương pháp đề xuất (SVM-GIST) với mô hình mạng nơron tích chập CNN [LeCun & Bengio, 1995], [LeCun et al., 2015], máy học SVM sử dụng mô hình túi từ (Bag-of-Words - BoW) của đặc trưng SIFT (SVM-BoW-SIFT) và Bayes thơ ngây với láng giềng gần nhất sử dụng đặc trưng SIFT (NBNN-SIFT). Chương trình huấn luyện và nhận dạng của mạng tích chập CNN được cài đặt bằng ngôn ngữ lập trình Python sử dụng thư viện Tensorflow [Abadi et al., 2015]. Chương trình SVM-BoW-SIFT, NBNN-SIFT cũng được cài đặt bằng ngôn ngữ lập trình C/C++ sử dụng thư viện OpenCV.

Tất cả các thí nghiệm được chạy trên máy tính cá nhân, cài hệ điều hành Linux Fedora 20, bộ vi xử lý Intel® Core i7-4790, 3.6 GHz, 4 nhân và bộ nhớ RAM 32 GB.

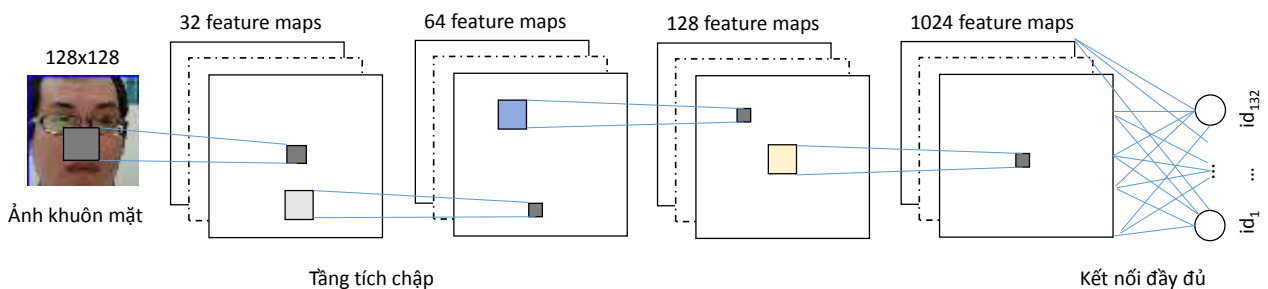
## B. Chuẩn bị tập dữ liệu

Thí nghiệm được thực hiện trên tập dữ liệu gồm 6722 ảnh khuôn mặt của 132 đối tượng là những sinh viên Khoa CNTT-TT, Trường Đại học Cần Thơ. Chúng tôi tiến hành huấn luyện mô hình định vị khuôn mặt bằng cách tạo chỉ mục, vị trí khuôn mặt trong các ảnh, tạo tập dữ liệu không có khuôn mặt, huấn luyện mô hình phân tầng CBC có 20 tầng, tỷ lệ nhận dạng khuôn mặt tối thiểu là 99.90% và tỷ lệ nhận dạng lầm khuôn mặt tối đa không quá 50%. Thời gian huấn luyện mô hình mất khoảng 7 giờ, tuy nhiên thời gian định vị khuôn mặt chỉ khoảng 15 mili giây với độ chính xác 98%. Chương trình định vị khuôn mặt trong ảnh thu được từ camera để thu thập tập dữ liệu ảnh dùng trong thực nghiệm định danh đối tượng trong hệ thống. Ảnh khuôn mặt được chuẩn hóa về kích thước 128x128 để có thể dễ dàng xây dựng mô hình nhận dạng đối tượng trong hệ thống.

Tập dữ liệu khuôn mặt được chia ngẫu nhiên thành hai tập: huấn luyện gồm 4482 ảnh và kiểm tra gồm 2240 ảnh. Chúng tôi sử dụng tập huấn luyện để huấn luyện các mô hình nhận dạng SVM-GIST, CNN, SVM-BoW-SIFT, NBNN-SIFT. Sau đó, sử dụng các mô hình thu được để nhận dạng các đối tượng trong tập kiểm tra.

## C. Điều chỉnh tham số

Trong đề xuất SVM-GIST của chúng tôi, bước rút trích đặc trưng GIST tạo ra tập dữ liệu dạng bảng, mỗi dòng tương ứng là ảnh khuôn mặt có 960 chiều và nhãn là đối tượng tương ứng trong hệ thống. Chúng tôi sử dụng thư viện LibSVM [Chang & Lin, 2011] để huấn luyện mô hình SVM nhận dạng đối tượng. Mô hình SVM sử dụng hàm nhân RBF vì tính tổng quát và tính hiệu quả. Chúng tôi cũng điều chỉnh siêu tham số  $\gamma$  của hàm nhân RBF và hằng số  $C$  (tham số dung hòa lỗi và độ rộng của lề SVM) để có được kết quả cao nhất dựa trên nghi thức kiểm tra chéo từ tập dữ liệu huấn luyện. Bộ tham số tối ưu tìm được là  $\gamma = 0.01$  và  $C = 100000$ .



**Hình 8.** Kiến trúc mạng tích chập CNN để nhận dạng đối tượng

Mô hình mạng CNN [LeCun & Bengio, 1995], [LeCun et al., 2015] để nhận dạng đối tượng được huấn luyện từ tập ảnh khuôn mặt cũng được điều chỉnh để đạt độ chính xác cao trên tập huấn luyện bằng nghi thức kiểm tra chéo. Kiến trúc mạng CNN như hình 8. Bộ tham số như tốc độ học  $\eta=0.001$ ,  $batch\_size = 128$ , số lượng  $epochs = 100$  được điều chỉnh để đạt kết quả nhận dạng tối ưu.

Phương pháp SVM-BoW-SIFT rút trích đặc trưng SIFT, xây dựng mô hình túi từ BoW với 5000 từ dựa trên nghiên cứu [Bosch et al., 2006], [Do, 2011]. Huấn luyện mô hình nhận dạng SVM sử dụng hàm nhân RBF với bộ tham số tối ưu là  $\gamma = 0.005$  và  $C = 100000$ , tìm được bằng nghi thức kiểm tra chéo trên tập dữ liệu huấn luyện.

Phương pháp NBNN [Boiman et al., 2008] sử dụng đặc trưng SIFT thực hiện nhận dạng ảnh đối tượng bằng tính toán trực tiếp khoảng cách (dựa trên các vectơ mô tả đặc trưng) từ ảnh truy vấn đến lớp (đối tượng). Phương pháp NBNN-SIFT là mô hình phi tham số.

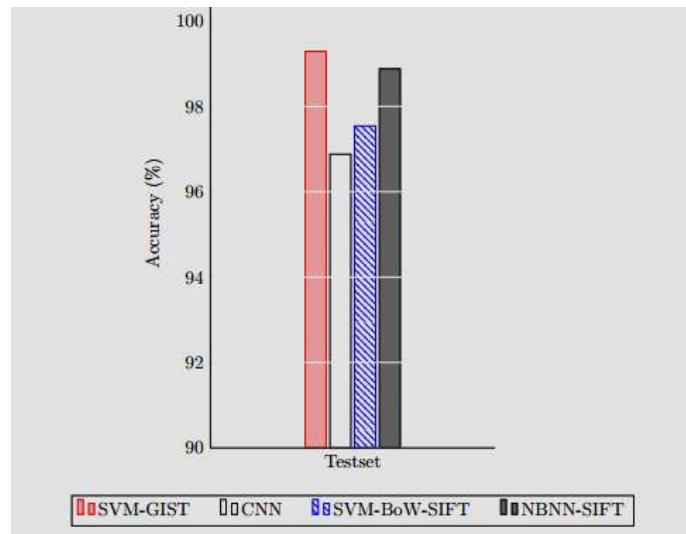
## D. Kết quả thực nghiệm

Các mô hình được xây dựng dựa trên tập huấn luyện có 4482 ảnh và kết quả được ghi nhận dựa trên tập kiểm tra gồm 2240 ảnh của 132 đối tượng trong hệ thống. Kết quả thực nghiệm thu được như trình bày trong bảng 1, hình 9, hình 10.

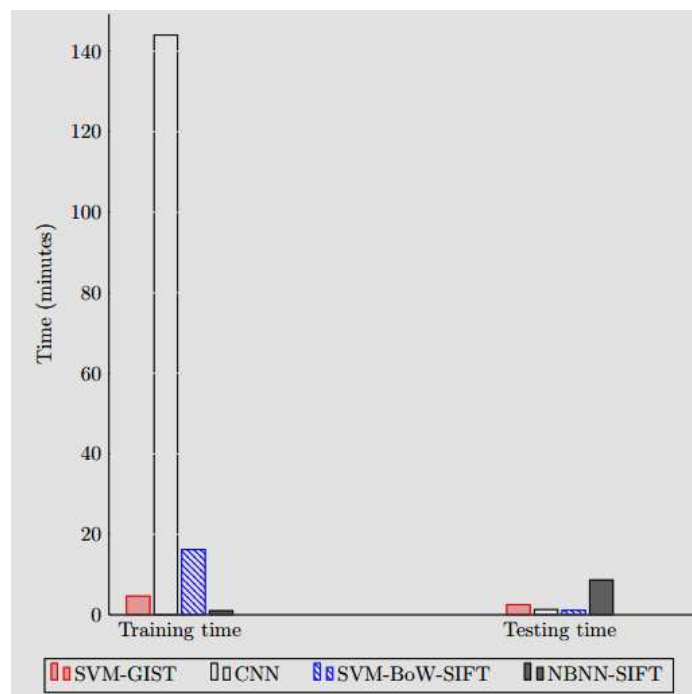
Phương pháp đề xuất SVM-GIST cho kết quả nhận dạng (định danh) khuôn mặt với độ chính xác 99.29%, là phương pháp cho độ chính xác cao nhất, xếp thứ hai là phương pháp NBNN-SIFT với độ chính xác 98.88%, xếp thứ ba là SVM-BoW-SIFT cho độ chính xác là 97.54%, xếp sau cùng là mạng tích chập CNN cho độ chính xác 96.88%. So với mạng tích chập CNN, phương pháp SVM-GIST cải thiện 2.41% độ chính xác trong nhận dạng đối tượng trong hệ thống.

**Bảng 1.** So sánh hiệu quả của các phương pháp nhận dạng khuôn mặt theo độ chính xác (%), thời gian huấn luyện (phút), thời gian nhận dạng (phút)

ID	Phương pháp	Độ chính xác (%)	Thời gian huấn luyện (phút)	Thời gian nhận dạng (phút)
1	SVM-GIST	99.29	4.65	1.95
2	Mạng tích chập CNN	96.88	144.00	1.30
3	SVM-BoW-SIFT	97.54	16.20	1.10
4	NBNN-SIFT	98.88	1.00	8.65



**Hình 9.** So sánh độ chính xác



**Hình 10.** So sánh thời gian huấn luyện và nhận dạng

Xét về thời gian huấn luyện mô hình nhận dạng (tổng thời gian rút trích đặc trưng và huấn luyện mô hình nhận dạng) và thời gian nhận dạng (tổng thời gian rút trích đặc trưng và nhận dạng). Phương pháp NBNN-SIFT có thời gian huấn luyện 1 phút, nhanh nhất do không phải huấn luyện mô hình mà chỉ rút trích đặc trưng SIFT tuy nhiên khi nhận dạng mất 8.65 phút. SVM-GIST cần 4.65 phút cho huấn luyện và 1.95 phút cho nhận dạng. SVM-BoW-SIFT cần 16.20 phút để huấn luyện mô hình nhận dạng trong khi cần chỉ 1.10 phút để nhận dạng. Mạng tích chập hoàn thành huấn luyện mô hình nhận dạng trong 144 phút và quá trình nhận dạng mất 1.30 phút.

Kết quả thực nghiệm cho thấy SVM-GIST cho kết quả nhận dạng nhanh, chính xác khi so sánh với các phương pháp CNN, SVM-BoW-SIFT, NBNN-SIFT.

#### IV. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Chúng tôi vừa trình bày một hệ thống điểm danh bằng mặt người với máy học SVM sử dụng đặc trưng GIST. Hệ thống điểm danh thực hiện định vị khuôn mặt trong ảnh thu được từ camera bằng mô hình phân tầng CBC sử dụng đặc trưng Haar-like, đáp ứng được tiêu chí về tốc độ và độ chính xác trong định vị. Hệ thống thực hiện nhận dạng đối tượng trong hệ thống bằng ảnh khuôn mặt với mô hình máy học SVM và đặc trưng GIST (SVM-GIST). Ngoài ra để đánh giá hiệu quả của sự kết hợp của đặc trưng GIST và SVM, chúng tôi cũng đã nghiên cứu và làm thực nghiệm với các đặc trưng (SIFT-BoW) và các phương pháp phân lớp khác (NBNN, CNN).

Kết quả thực nghiệm trên tập dữ liệu gồm 6722 ảnh của 132 đối tượng là những sinh viên Khoa CNTT-TT, Trường Đại học Cần Thơ cho thấy phương pháp đề xuất SVM-GIST đạt đến 99.29% độ chính xác trên tập kiểm tra, cao hơn khi so với mô hình mạng nơron tích chập CNN, máy học SVM sử dụng mô hình túi từ BoW của đặc trưng SIFT (SVM-BoW-SIFT), Bayes thơ ngây với láng giềng gần nhất sử dụng đặc trưng SIFT (NBNN-SIFT). Kết quả này cho thấy rằng sự kết hợp của đặc trưng GIST và SVM cho kết quả tốt nhất trong các phương pháp. Có thể lý giải kết quả này như sau: (i) ở giai đoạn phân lớp, ảnh đầu vào là ảnh mặt người đã được trích ra khỏi ảnh gốc vì thế đặc trưng cục bộ hay toàn cục có tác dụng như nhau; (ii) khi sử dụng đặc trưng SIFT, ta cần phải có một bước lượng hoá. Lỗi lượng hoá sẽ góp phần làm giảm hiệu quả phân lớp. Có lẽ vì thế mà khi áp dụng phương pháp NBNN, kết quả được cải thiện do NBNN không cần bước lượng hoá. Nếu so sánh riêng SVM-GIST và NBNN-SIFT, cả hai cho kết quả phân lớp tương đương tuy nhiên thời gian phân lớp của NBNN nhiều hơn 4 lần.

Trong tương lai gần, chúng tôi dự định cung cấp thêm kết quả thực nghiệm trên những tập dữ liệu lớn hơn và so sánh hiệu quả của SVM-GIST với các phương pháp khác. Có thể mở rộng SVM-GIST cho các vấn đề như dự đoán tuổi, nhận dạng logo, ngôn ngữ ký hiệu.

#### TÀI LIỆU THAM KHẢO

- [1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, R. Jozefowicz, Y. Jia, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, M. Schuster, R. Monga, S. Moore, D. Murray, C. Olah, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng.: TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org
- [2] M. Aly.: Face Recognition Using SIFT Features. Technical Report Caltech, 2006.
- [3] P. Belhumeur, J. Hespanha and D. Kriegman.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions On Pattern Analysis And Machine Intelligence* Vol.19(7):711-720, 1996.
- [4] O. Boiman, E. Shechtman and M. Irani.: In Defense of Nearest-Neighbor Based Image Classification. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition - CVPR, 2008, pp. 1-8.
- [5] A. Bosch, A. Zisserman and X. Munoz.: Scene classification via pLSA. Proceedings of the European Conference on Computer Vision, 2006, pp. 517-530.
- [6] G. Bradski and A. Kaehler.: *Learning OpenCV*. O'Reilly Media, 2012.
- [7] C. Chang, and C. J. Lin.: LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* Vol.2(27):1-27, 2011.
- [8] N. K. Châu, T. N. Đỗ, T. T. Võ and N. K. Phạm.: Nhận dạng mặt người với giải thuật Haar like feature cascade of boosted classifiers và các đặc trưng SIFT. Kỳ yếu hội thảo toàn quốc về CNTT-TT, Đại học Cần thơ, 2013, trang 1-10.
- [9] N. K. Châu and T. N. Đoàn: Nhận dạng mặt người với giải thuật Haar Like Feature - Cascade of Boosted Classifiers và đặc trưng SIFT. *Tạp chí khoa học AGU* Vol.3(2):15-24, 2014.
- [10] N. Cristianini and J. Shawe-Taylor.: *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA, 2000.
- [11] T. N. Do. Detection of Pornographic Images Using Bag-of-Visual-Words and Arcx4 of Random Multinomial Naive Bayes. in *Journal of Science and Technology* Vol.49(5): 13-24, 2011.



- [12] T. N. Do and N. K. Pham.: Handwritten Digit Recognition Using GIST Descriptors and Random Oblique Decision Trees. in *Advances in Intelligent Systems and Computing* Vol.341:1-15, Springer, 2015.
- [13] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid.: Evaluation of GIST descriptors for web-scale image search. *Proceedings of the ACM International Conference on Image and Video Retrieval*, 2009, pp. 1-8.
- [14] O. Dürr, Y. Pauchard, D. Browarnik, R. Axthelm and M. Loeser.: Deep Learning on a Raspberry Pi for Real Time Face Recognition. *Proceedings of the Eurographics 2015*, pp. 11-12.
- [15] R. Fan, K. Chang, C. Hsieh, X. Wang and C. Lin.: LIBLINEAR: a library for large linear classification. *Journal of Machine Learning Research* Vol.9(4):1871-1874, 2008.
- [16] Y. Freund and R. Schapire.: A decision-theoretic generalization of on-line learning and an application to boosting. *Proceedings of the Second European Conference on Computational Learning Theory*. 1995, pp. 23-37.
- [17] C. Geng and X. Jiang.: Face Recognition Using SIFT Features. *Proceedings of the 16th IEEE International Conference on Image Processing*, 2009, pp. 3277-3280.
- [18] I. Guyon.: Web page on svm applications. 1999. <http://www.clopinet.com/isabelle/Projects/SVM/applist.html>.
- [19] U. Krebel.: Pairwise classification and support vector machines. *Advances in Kernel Methods: Support Vector Learning*, 1999, pp. 255-268.
- [20] A. Krizhevsky, I. Sutskever and G. Hinton.: ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097-1105.
- [21] Y. LeCun and Y. Bengio.: Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361 (10), 1995.
- [22] Y. LeCun, Y. Bengio and G. Hinton.: Deep learning. *Nature* 521(7553): 436-444, 2015.
- [23] R. Lienhart and J. Maydt.: An Extended Set of Haar-like Features for Rapid Object Detection. *Proceedings of International Conference on Image Processing*, 2002, pp. 900-903.
- [24] D. Lowe.: Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 2004, pp. 91-110.
- [25] A. Oliva and A. Torralba.: Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42, 145-175, 2001.
- [26] O. Parkhi, A. Vedaldi and A. Zisserman.: Deep Face Recognition. *Proceedings of the British Machine Vision Conference*, 2015.
- [27] M. Turk and A. Pentland.: Face recognition using eigenfaces. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1991.
- [28] R Vaillant, C Monrocq and Y LeCun.: Localizing faces in images. *Revue technique-Thomson-CSF*, 1993.
- [29] V. Vapnik.: *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.
- [30] P. Viola and M. Jones.: Robust Real-time Object Detection. *International Journal of Computer Vision*, 2001.
- [31] X. Wu and V. Kumar.: *Top 10 Algorithms in Data Mining*. Chapman & Hall/CRC, 2009.
- [32] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld.: Face recognition: A literature survey. *ACM Computing Surveys* Vol.35(4):399-458, 2003.

## HUMAN FACE RECOGNITION WITH GIST FEATURES AND SUPPORT VECTOR MACHINE ALGORITHM

Nguyen Khang Pham, Minh Thu Tran Nguyen, Thanh Nghi Do

**ABSTRACT:** In this paper, we present the new human face recognition system using GIST features and Support Vector Machines (SVM). Our system performs the human face recognition task via two main steps: detecting the human face from the camera and identifying the corresponding individual in the database. The system detects the human face with Haar Like Feature - Cascade of Boosted Classifiers (CBC); followed which, it extracts GIST features from the human face region and learns the SVM model from GIST features to identify who is the individual corresponding to the human face region. The numerical test results on a real dataset with 6722 images collected from 132 students at the Can Tho University show that our system achieves an accuracy 99.29%, more accurate than the Convolutional Neural Network (CNN), the SVM using Bag-of-Words (BoW) and the Scale-Invariant Feature Transform (SIFT), the Naïve Bayes Nearest Neighbor (NBNN) using the SIFT features with accuracies 96.88%, 97.54% and 98.88% respectively.