

DÙNG ĐẶC TRƯNG GABOR KẾT HỢP ADABOOST VÀ K-MEANS TRONG BÀI TOÁN NHẬN DẠNG MẶT NGƯỜI

TRỊNH TẤN ĐẠT*, PHẠM THẾ BẢO**

TÓM TẮT

Trong bài báo này, chúng tôi đề xuất phương pháp nhận dạng mặt người dựa trên đặc trưng Gabor kết hợp AdaBoost và k-means. Chúng tôi sử dụng Gabor wavelet để rút trích các đặc trưng trên ảnh tạo vector đặc trưng. AdaBoost cải tiến được dùng như một phương pháp làm giảm số chiều của vector đặc trưng trong quá trình nhận dạng. Đồng thời, dùng thuật toán k-means để phân nhóm dữ liệu thành các nhóm riêng biệt khác nhau nhằm giảm thời gian xử lý và tăng hiệu quả nhận dạng. Cơ sở dữ liệu ảnh ORL của AT&T được dùng để thực nghiệm với 200 đặc trưng Gabor để nhận dạng thì tỉ lệ chính xác là 86,07% và thời gian nhận dạng trung bình là 0,06s.

Từ khóa: nhận dạng khuôn mặt, phương pháp Adaboost, đặc trưng Gabor, k-means.

ABSTRACT

Face recognition using Gabor feature combining with Adaboost and k-mean

In this article, we proposed a new hybrid method for face recognition based on Gabor feature combining with Adaboost and k-means. We made a Gabor's feature vector from a face image using Gabor wavelet method. Improving Adaboost was used as a method to reduce the dimensions of feature vectors during the recognition process. At the same time, we used k-mean method to divide face database into different groups so as to decrease the processing time and increase the efficiency of recognition process. ORL database of AT&T was used for testing with 200 Gabor features and the precision is 86,07 percent and the average amount of time for recognition is 0,06s.

Keywords: face recognition, adaboost, Gabor feature .

1. Giới thiệu

Nhận dạng mặt người là một trong những ứng dụng quan trọng của thị giác máy tính nói riêng cũng như khoa học máy tính nói chung. Bài toán nhận dạng khuôn mặt người vốn được nghiên cứu từ những năm 1970 và cho đến nay, rất nhiều nghiên cứu lần ứng dụng cho bài toán này đã ra đời [9]. Bài toán nhận dạng mặt người có thể áp dụng rộng rãi trong

nhiều lĩnh vực khác nhau. Các ứng dụng liên quan đến nhận dạng mặt người có thể kể như: hệ thống phát hiện tội phạm, hệ thống theo dõi nhân sự trong một đơn vị, hệ thống tìm kiếm thông tin trên ảnh, video dựa trên nội dung,...

Để giải quyết bài toán này, cần xác định hai vấn đề chính. Thứ nhất, dùng thông tin nào để nhận dạng, chẳng hạn như mắt, mũi, miệng, chân mày,... hay kết hợp tất cả các thông tin trên. Thứ hai, dùng phương pháp nào để nhận dạng nguồn thông tin đó. Nhiều phương pháp nhận dạng khuôn mặt người được đề xuất

* ThS, Trường Đại học Khoa học Tự nhiên ĐHQG TP HCM

** TS, Trường Đại học Khoa học Tự nhiên ĐHQG TP HCM

trong khoảng 30 năm gần đây, và có rất nhiều công trình và bài báo đã được công bố rộng rãi trên thế giới. Cụ thể có ba hướng tiếp cận [9] cho bài toán nhận dạng khuôn mặt người: phương pháp so khớp tổng thể, phương pháp so khớp dựa trên đặc trưng và phương pháp lai.

Chúng tôi sử dụng hướng tiếp cận dựa trên việc so khớp các đặc trưng để giải quyết bài toán. Với hướng tiếp cận này, nhiều tác giả đã sử dụng Gabor wavelet kết hợp với AdaBoost để giải quyết bài toán nhận dạng mặt người [2, 4, 8]. Trong bài báo này, chúng tôi cũng sử dụng lại hai phương pháp trên để giải quyết bài toán, bởi vì việc rút trích đặc trưng bằng Gabor wavelet thực hiện khá nhanh, Gabor wavelet không những phát hiện được những thay đổi nổi bật giữa các giá trị điểm ảnh mà hạn chế tác động của sự thay đổi độ sáng ảnh. Từ khi ra đời, AdaBoost đã được ứng dụng nhiều trong lĩnh vực phân loại mẫu và nhận dạng mẫu. AdaBoost là thuật toán học mạnh dựa trên nhiều thành phần yếu có nhiều thuận lợi [1, 8]. Tuy nhiên, đối với cơ sở dữ liệu lớn thì quá trình nhận dạng tăng thời gian xử lý nhiều khi tiến hành so khớp, đồng thời còn làm giảm tỉ lệ chính xác. Để khắc phục hạn chế này, chúng tôi sử dụng thuật toán k-means để phân nhóm tập dữ liệu thành các nhóm riêng biệt khác nhau nhằm giảm thời gian xử lý và tăng hiệu quả quá trình nhận dạng. Ngoài ra, chúng tôi cũng đề xuất ra một số hiệu chỉnh cho thuật toán AdaBoost để tăng tính hiệu quả.

2. Vectơ đặc trưng Gabor

2.1. Gabor wavelet

Các Gabor wavelet [2, 4, 8] $\psi_{\mu, \nu}$ được xác định như công thức 1. Trong đó: z là một điểm trong không gian hai chiều. Các tham số μ và ν lần lượt là hướng và tỉ lệ của nhân Gabor, σ là độ lệch chuẩn của phân phối Gauss, $\|\cdot\|$ là toán tử chuẩn.

$$\psi_{\mu, \nu}(z) = \frac{\|k_{\mu, \nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu, \nu}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu, \nu}z} - e^{-\sigma^2/2} \right] \quad (1)$$

Một vectơ sóng $k_{\mu, \nu}$ được xác định như công thức 2. Trong đó: $k_{\nu} = \frac{k_{\max}}{f^{\nu}}$ là tần số của sóng, $\phi_{\mu} = \frac{\pi\mu}{8}$ là hướng của sóng, với k_{\max} là tần số tối đa của sóng, f^{ν} là hệ số khoảng cách giữa các nhân trong miền tần số. Chúng tôi sử dụng 8 hướng và 5 tỉ lệ khác nhau để tạo ra 40 bộ lọc Gabor, cụ thể $\mu \in \{0, \dots, 7\}$ và $\nu \in \{0, \dots, 4\}$. Các thông số khác được chọn như sau: $\sigma = 2\pi$, $k_{\max} = \frac{\pi}{2}$, $f = \sqrt{2}$ [4, 7, 8].

$$k_{\mu, \nu} = k_{\nu} e^{i\phi_{\mu}} \quad (2)$$

Gọi $I(z)$ là giá trị mức xám tại điểm ảnh z , tích chập của $I(z)$ với một bộ lọc Gabor $\psi_{\mu, \nu}(z)$ được xác định như công thức 3.

$$O_{\mu, \nu}(z) = I(z) * \psi_{\mu, \nu}(z) \quad (3)$$

Giá trị $\|O_{\mu, \nu}(z)\|$ chính một đặc trưng Gabor tại vị trí z với hướng và tỉ lệ là μ và ν . Một vectơ đặc trưng Gabor đại diện cho toàn bộ bức ảnh I có được

bằng cách ghép nối theo dòng tất cả các giá trị

$$\|O_{\mu,\nu}(z)\|, \forall z \in I, \nu \in \{0, \dots, 4\}, \mu \in \{0, \dots, 7\}.$$

Ta gọi $G(I)$ là vector đặc trưng Gabor của ảnh I , công thức 4.

$$G(I) = (\|O_{0,0}\|, \|O_{0,1}\|, \dots, \|O_{0,7}\|, \|O_{1,0}\|, \dots, \|O_{4,7}\|) \quad (4)$$

Bài toán nhận dạng mặt người là bài toán nhận dạng cho đa lớp. Để dễ dàng, chúng tôi sử dụng AdaBoost cho hai lớp. Chúng tôi sử dụng phương pháp do Moghaddam, Wahid và Pentland [6] đề xuất để chuyển từ không gian đa lớp thành hai không gian intra-personal và không gian extra-personal.

2.2. Không gian intra-personal và không gian extra-personal

Không gian intra-personal (IS) chính là không gian chỉ sự khác biệt giữa các bức ảnh khác nhau của cùng một đối tượng. Không gian extra-personal (ES) chính là không gian chỉ sự khác biệt giữa các bức ảnh khác nhau của các đối tượng khác nhau. Dựa vào các vector đặc trưng Gabor, hai không gian này được xác định như công thức 5 và 6. Trong đó I_p và I_q là ảnh khuôn mặt của đối tượng p và q .

$$IS = \left\{ \|G(I_p) - G(I_q)\|, p = q \right\} \quad (5)$$

$$ES = \left\{ \|G(I_p) - G(I_q)\|, p \neq q \right\} \quad (6)$$

Các vector (các mẫu) trong không gian IS được gán nhãn là +1 (mẫu dương tính) và các mẫu trong không gian ES được gán nhãn là -1 (mẫu âm tính).

3. Chọn lựa đặc trưng và phân lớp bằng AdaBoost

Một tập huấn luyện gồm N mẫu được chọn từ không gian $IE = IS \cup ES$ được biểu diễn thành một bộ các vector

$$\left[\{(x_1, y_1), \dots, (x_N, y_N)\}, x_i = (x_{i1}, \dots, x_{im}), y_i \in \{-1, +1\} \right].$$

Trong đó: x_i là các vector đặc trưng Gabor, $x_{ik} = \|G(I_p) - G(I_q)\|_k$ là đặc trưng Gabor tại thành phần thứ k của vector x_i .

Chúng tôi xây dựng một phân loại AdaBoost để xác định lớp của vector đặc trưng Gabor có được từ ảnh đầu vào. Phân loại này sẽ dựa trên một bộ các phân loại yếu, mỗi phân loại yếu sẽ sử dụng một thành phần trong vector đặc trưng, biểu thức 7 và 8. Vì thế việc chọn các bộ phân loại yếu để hình thành một phân loại mạnh sẽ tương ứng việc lựa chọn các đặc trưng Gabor quan trọng trong vector đặc trưng Gabor ban đầu, vector đặc trưng ban đầu sẽ có số chiều lớn. Số lượng đặc trưng cần dùng trong vector đặc trưng chính là số lượng T phân loại yếu chúng ta có trong phân loại AdaBoost trong thuật toán huấn luyện dữ liệu.

$$h_k = \begin{cases} +1, & f_k(I) < \lambda_k \\ -1, & f_k(I) \geq \lambda_k \end{cases} \quad (7)$$

Với $f_k(I) = x_{ik} = \|G(I_p) - G(I_q)\|_k$ và ngưỡng λ_k được tính như công thức 8.

$$\lambda_k = \frac{1}{2} \left(\frac{1}{m} \sum_{p=1}^m (x_{pk} | y_p = 1) + \frac{1}{l} \sum_{q=1}^l (x_{qk} | y_q = -1) \right) \quad (8)$$

Trong đó m là số lượng mẫu trong không gian IS và l là số lượng mẫu trong không gian ES.

Ngoài ra, chúng tôi cũng đề xuất một sự thay đổi cho thuật toán AdaBoost. Trong thuật toán AdaBoost gốc, sau mỗi vòng lặp, một bộ phân loại yếu tương ứng vị trí của một đặc trưng được chọn dựa vào sai số nhỏ nhất để giảm số chiều của vector đặc trưng Gabor. Tuy nhiên, dựa vào thực nghiệm chúng tôi nhận thấy

rằng, một vài đặc trưng được chọn lại nhiều lần. Điều này không hợp lý theo khái niệm chọn lựa đặc trưng và làm giảm hiệu quả của hệ thống nhận dạng. Để khắc phục hạn chế này, chúng đưa ra một cải tiến trong bước lặp của thuật toán AdaBoost. Sau mỗi bước lặp, đổi với vị

trí của đặc trưng đã được chọn trong bước lặp trước đó, chúng tôi sẽ không xem xét lại đặc trưng đó trong bước lặp hiện hành. Do đó, những đặc trưng chỉ được chọn duy nhất một lần qua sự điều chỉnh này.

Thuật toán huấn luyện dữ liệu bằng AdaBoost cải tiến

Dữ liệu vào

1. Cho một tập huấn luyện gồm N mẫu có đánh dấu $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ với x_i thuộc không gian IE và y_i thuộc $Y = \{-1, 1\}$.
2. Số nguyên T – số lần lặp.

Huấn luyện

Khởi tạo dữ liệu ban đầu

Khởi tạo trọng số ban đầu $w_{1,i}$ cho tất cả các mẫu: với $i = 1, \dots, N$ theo công thức 9.

$$w_{1,i} = \begin{cases} \frac{1}{2m}, & x_i \in IS \\ \frac{1}{2l}, & x_i \in ES \end{cases} \tag{9}$$

Xây dựng T bộ phân loại yếu

Duyệt $t = 1, \dots, T$

1. Chuẩn hóa lại các trọng số $w_{t,i} = \frac{w_{t,i}}{\sum_{i=1}^N w_{t,i}}$, với $i = 1, \dots, N$

2. Huấn luyện các bộ phân loại yếu bằng phân phối w_t . Với mỗi đặc trưng j trong mỗi vectơ đặc trưng Gabor (xét trên từng mẫu), ta xây dựng một bộ phân loại yếu h_j với ngưỡng λ_j và lỗi ε_j , sai số (lỗi) được tính theo công thức 10.

$$\varepsilon_j = \sum_i w_{t,i} |(h_j(x_i) - y_i) / 2| \tag{10}$$

3. Chọn ra h_t với ε_j nhỏ nhất, nghĩa là ta chọn được đặc trưng tại vị trí j và các đặc trưng nào đã được chọn thì ta không chọn lại. Khi đó, ta được h_t tương ứng với ε_t để có $h_t: X \rightarrow \{-1, 1\}$.

4. Cập nhật lại trọng số $w_{t+1,i}$ theo công thức 11.

$$w_{t+1,i} = w_{t,i} \times \begin{cases} e^{-\alpha_t}, & h_t(x_i) = y_i \\ e^{\alpha_t}, & h_t(x_i) \neq y_i \end{cases} \text{ với } i = 1, \dots, N \tag{11}$$

với $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$

Dữ liệu ra

Một phân loại mạnh
$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right)$$

4. Phân nhóm dữ liệu

Nếu số lượng ảnh khuôn mặt người nhiều cũng như số lượng người trong bộ dữ liệu nhận dạng của hệ thống nhận dạng nhiều thì chi phí xử lý để biết một người cần nhận dạng là ai sẽ tăng đáng kể. Để giải quyết vấn đề này chúng tôi đề xuất sử dụng phương pháp k-means [3, 5] với mong muốn để đưa tất cả ảnh của một người thành một nhóm và chọn phần tử đại diện cho quá trình nhận dạng. Do đó, giá trị của k được chọn chính là số lượng người có khuôn mặt trong cơ sở dữ liệu ảnh với khoảng cách Euclide bình phương.

5. Thực nghiệm

Chúng tôi cài đặt chương trình ứng dụng bằng Matlab và sử dụng cơ sở dữ liệu ảnh mặt người ORL của AT&T, Trường Đại học Cambridge. Chúng tôi chọn 400 ảnh khuôn mặt của 40 người, mỗi người sẽ có 10 ảnh. Trong các ảnh này, khuôn mặt người được chụp chính diện, mỗi khuôn mặt đều có đầy đủ các thông tin đặc trưng như chân mày, mắt, mũi, miệng, cằm,... Một số khuôn mặt hơi nghiêng với góc quay không đáng kể và có nhiều biểu cảm khác nhau trên khuôn mặt: bình thường, vui, buồn, cười,... Một vài đối tượng có đeo kính hoặc có râu. Các ảnh được chụp trong các điều kiện ánh sáng khác nhau. Điều này cho thấy tập dữ liệu mà chúng tôi chọn khá đa dạng. Để huấn luyện cũng như kiểm tra, chúng tôi tiến hành chuẩn hóa các ảnh trong tập dữ liệu thu thập trên.

Quá trình chuẩn hóa bao gồm các thao tác sau: chuyển tất cả ảnh về thành ảnh xám, loại bỏ phần dư thừa, ảnh chỉ chứa các đặc trưng quan trọng trên khuôn mặt như: chân mày, mắt, mũi, miệng, gò má, cằm và chuẩn hóa các ảnh về kích thước 32x32 pixel, hình 1.

Đối với từng đối tượng, chúng tôi chọn ra 03 ảnh để huấn luyện và 07 ảnh để kiểm tra. Tập huấn luyện của chúng tôi sẽ có 120 ảnh và tập kiểm tra sẽ có 280 ảnh. Mỗi vectơ đặc trưng Gabor có số chiều là $32.32.40 = 40960$ (chiều). Số lượng mẫu trong không gian IS mà chúng tôi chọn là $40x\left(\frac{3}{2}\right) = 120$ mẫu và số lượng mẫu trong không gian ES là $\left(\frac{40x3}{2}\right) - 40x\left(\frac{3}{2}\right) = 7020$ mẫu. Chúng tôi chọn ngẫu nhiên 1600 mẫu từ 7020 mẫu trong không gian ES để huấn luyện. Như vậy, chúng tôi dùng 1720 mẫu, trong đó có 120 mẫu dương tính và còn lại 1600 mẫu âm tính, huấn luyện. Số lượng đặc trưng cần chọn là (T =) 200 thì có kết quả huấn luyện như bảng 1.

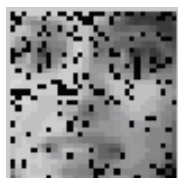


Hình 1. Một phần tập dữ liệu thực nghiệm

Bảng 1. Kết quả huấn luyện

Mẫu IS	Mẫu ES	Mẫu huấn luyện	Số đặc trưng được chọn (T)	Thời gian huấn luyện	Mẫu bị phân sai lớp		Lỗi huấn luyện
					Mẫu IS	Mẫu ES	
120	1600	1720	200	10,741 giây	0	115	6,686%

Các vị trí được chọn trên ảnh phần lớn tập trung ở vùng mắt, chân mày, mũi,... các vị trí trên đều là những vị trí quan trọng để phân biệt các khuôn mặt với nhau, hình 2.



Hình 2. Các vị trí trên ảnh được lựa chọn

Việc sử dụng k-means để phân nhóm tập huấn luyện phụ thuộc vào giá trị k. Thông thường giá trị k được chọn một cách ngẫu nhiên hoặc dựa vào thực nghiệm. Trong bài báo này, chúng tôi chọn lựa giá trị này dựa vào tính chất của tập dữ liệu sử dụng và thực nghiệm. Tập dữ liệu ORL mà chúng tôi sử dụng được thu thập từ 40 người. Chúng tôi mong muốn rằng khi áp dụng k-means, số lượng ảnh của cùng một đối tượng được phân vào một nhóm riêng biệt. Điều này có thể tăng hiệu quả của hệ thống nhận dạng và giảm thời gian tính toán đáng kể. Nếu chúng ta chọn giá trị k nhỏ hơn số lượng người có trong tập dữ liệu, khi đó

những ảnh của các đối tượng khác nhau có thể được gom vào chung một nhóm. Do đó, điều này sẽ làm giảm hiệu quả của hệ thống nhận dạng. Ngược lại, nếu chúng ta chọn giá trị k lớn hơn số lượng người có trong tập dữ liệu, các ảnh của cùng một đối tượng có thể được gom vào các nhóm khác nhau. Điều này làm tăng thời gian tính toán và giảm hiệu quả nhận dạng. Dựa vào những lí do này, chúng tôi quyết định chọn giá trị k bằng với số người có trong tập dữ liệu thực nghiệm.

Sử dụng k-means để phân nhóm tập ảnh huấn luyện gồm 120 ảnh với k = 40, kết quả là có 11 ảnh bị gom sai nhóm. Với 280 ảnh cần nhận dạng, hệ thống cho 241 ảnh được nhận dạng chính xác và 39 ảnh được nhận dạng không chính xác. Tỷ lệ nhận dạng chính xác là 86,07%. Nếu không sử dụng k-means, với 280 ảnh cần nhận dạng thì có 238 ảnh được nhận dạng chính xác và 42 ảnh được nhận dạng không chính xác. Tỷ lệ nhận dạng khoảng 85%. Tuy nhiên, thời gian xử lí sẽ nhiều hơn, bảng 2.

Bảng 2. Kết quả của AdaBoost kết hợp k-means và AdaBoost

	Số ảnh cần nhận dạng	Số ảnh nhận dạng đúng	Số ảnh nhận dạng sai	Tỷ lệ chính xác (%)	Thời gian trung bình
AdaBoost kết hợp k-means	280	241	39	86,07	0,06 giây
AdaBoost	280	238	42	85,00	0,9 giây

6. Kết luận

Qua kết quả thực nghiệm thấy rằng phương pháp cải tiến của chúng tôi đề xuất đã cho kết quả khả quan hơn so với phương pháp ban đầu với cơ sở dữ liệu ORL của AT&T; cơ sở dữ liệu đa dạng. Việc rút trích đặc trưng bằng Gabor wavelet thực hiện khá nhanh, lỗi huấn luyện khi kết hợp AdaBoost và k-means đạt tỉ lệ thấp. Trong trường hợp không sử dụng k-means thì thời gian xử lí trung bình gấp gần 15 lần và tỉ lệ nhận dạng lại thấp hơn so với phương pháp chúng tôi đề xuất.

Tuy nhiên, khi sử dụng phương pháp AdaBoost, một vấn đề khó khăn đặt ra là việc lựa chọn số lượng mẫu để huấn luyện. Phương pháp AdaBoost đòi hỏi số lượng mẫu lớn để huấn luyện thì bộ phân loại đạt độ chính xác cao. Vấn đề thứ hai là việc mất cân đối về số lượng mẫu trong hai không gian IS và ES. Do đó, để khắc phục hạn chế này chúng tôi sẽ chọn cấu trúc phân tầng. Đồng thời, chúng tôi cải tiến thuật toán AdaBoost cho bài toán khuôn mặt người bị nghiêng; đây chính là vấn đề làm giảm tỉ lệ chính xác của nghiên cứu này.

TÀI LIỆU THAM KHẢO

1. Lu Boun Vinh, Hoàng Phương Anh (2004), *Nghiên cứu và xây dựng hệ thống nhận dạng mặt người dựa trên FSVM và AdaBoost*, Đề tài Luận văn tốt nghiệp Cử nhân Tin học, Trường Đại học Khoa học Tự nhiên TPHCM.
2. Bagci U., Bai L. (2008), “Parallel AdaBoost Algorithm for Gabor Wavelet Selection in Face Recognition”, *Proc of IEEE ICIP*, pp. 1640–1643.
3. Jain A. K., Dubes R. C. (1988), *Algorithms for Clustering Data*, Prentice Hall Advanced Reference Series, Prentice Hall.
4. Kepenekci B.(2001), *Face Recognition Using Gabor Wavelet Transform*, Msc thesis, Dept. of electrical and electronic engineering, The Middle East Technical University.
5. Lee Y., Choi S. (2004), “Minimum entropy, k-means, spectral clustering”, *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 117-122.
6. Moghaddam B., Wahid W., Pentland A. (1988), “Beyond Eigenfaces: Probabilistic Matching for Face Recognition”, *The 3rd IEEE Int'l Conference on Automatic Face & Gesture Recognition*, pp. 30-35.
7. Shen L. L., Bai L., Bardsley D., Wang Y. (2005), “Gabor Feature Selection for Face Recognition using Improved AdaBoost Learning”, *Proceedings of International Workshop on Biometric Recognition System in conjunction with ICCV*, pp. 39-49.
8. Yang P., Shan S., Gao W., Li S. Z. and Zhang D. (2004), “Face Recognition Using Ada-Boosted Gabor Features”, *Proceeding of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 356-361.
9. Zhao W., Chellappa R., Phillips P. J., Rosenfeld A. (2003), “Face Recognition: A Literature Survey”, *In ACM Computing Surveys*, 35(4), pp. 399-458.

(Ngày Tòa soạn nhận được bài: 02-01-2013; ngày phản biện đánh giá: 28-01-2013;
ngày chấp nhận đăng: 18-02-2013)