

HỆ THỐNG NHẬN DẠNG ẢNH MẶT NGƯỜI DÙNG MẠNG NEURAL VÀ BIẾN ĐỔI GABOR WAVELET

Bùi Ngọc Tuấn - Phùng Trung Nghĩa - Đỗ Huy Khôi (*Khoa Công nghệ thông tin- ĐHTN*)

1. Giới thiệu chung

Nhận dạng mặt người là một lĩnh vực nghiên cứu hấp dẫn được nhiều người quan tâm trong vài năm gần đây. Nói chung có hai phương pháp để nhận dạng mặt người, phương pháp dựa vào mẫu (template based) và phương pháp dựa vào các đặc trưng hình học (geometric features based). Trong phương pháp dựa vào mẫu, các phương pháp thống kê được sử dụng để biểu diễn toàn bộ ảnh. Phương pháp dựa vào mẫu điển hình đầu tiên là phương pháp “eigenface” [1], và được phát triển thành phương pháp “fisherface” [2]. Trong khi đó, phương pháp nhận dạng theo đặc trưng mô tả khuôn mặt theo ý tưởng đối tượng ảnh có thể được biểu diễn như một tập hợp các đặc trưng trừu tượng. Một số đặc trưng đơn giản như các đường biên, đường thẳng, đoạn thẳng, điểm. Các đặc trưng phức tạp hơn có thể được xây dựng từ các đặc trưng đơn giản này. Các phương pháp nhận dạng dựa vào mẫu hiện đại được biết đến như phương pháp sử dụng mô hình HMM [3], phương pháp phân tích các đặc trưng cục bộ LFA [4]. Trong hầu hết các phương pháp nhận dạng theo đặc trưng, sự lựa chọn các đặc trưng và mô tả chúng là rất quan trọng.

Wavelet là các hàm toán học chia dữ liệu thành các thành phần tần số khác nhau và xem xét từng thành phần với độ phân giải thích hợp. Wavelet có rất nhiều ưu điểm so với biến đổi Fourier truyền thống để phát hiện sự gián đoạn và thay đổi đột biến của tín hiệu. Phương pháp nhận dạng dùng biến đổi Gabor là một phương pháp nhận dạng dựa trên đặc trưng. Biến đổi Gabor biểu diễn ảnh mô phỏng theo hệ thống thị giác của con người HVS (Human Visual System), các đặc trưng Gabor của ảnh có thể được sử dụng hiệu quả trong các ứng dụng thị giác máy tính (Computer Vision), robotic.

Nền tảng chung để biểu diễn ảnh không gian ở đây dựa trên đồ thị có thứ tự, các điểm đặc trưng không gian được định vị để biểu diễn mẫu trong ảnh. Ở mỗi đặc trưng, một băng lọc đa phân giải ứng với các hàm Gabor wavelet đa hướng được sử dụng. Có rất nhiều cách biểu diễn ảnh dùng biến đổi Gabor wavelet. Ảnh có thể biểu diễn bằng cách tính Gabor wavelet cho mỗi điểm ảnh. Như vậy, thông thường mỗi điểm ảnh sẽ sử dụng 60 Gabor wavelet (6 tỉ lệ, 10 hướng). Nếu kích cỡ của ảnh là 100x100, số chiều của vector đặc trưng khi đó sẽ là 600.000, vector đặc trưng nhiều chiều như vậy sẽ dẫn tới chi phí tính toán và lưu trữ rất lớn. Một giải pháp khác là xác định một đồ thị cho khuôn mặt trong đó các node của đồ thị là các đặc trưng không gian trên khuôn mặt như mắt, tai, mũi, miệng. Phương pháp này yêu cầu phải định vị chính xác các điểm đặc trưng không gian. Trong bài báo này, phương pháp sử dụng là phương pháp kết hợp hai phương pháp trên sử dụng các điểm đặc trưng trong một mạng wavelet ứng với hình chữ nhật lớn nhất xác định trên khuôn mặt.

Hai dạng ứng dụng chính của nhận dạng mặt người là tìm ra người trong tập mẫu cho trước có chứa ảnh của người cần nhận dạng (identification) và xác thực xem người trong ảnh có trong tập mẫu hay không (verification). Trong bài báo này chúng tôi tập trung xây dựng hệ thống nhận dạng ảnh mặt người (với giả thiết người cần nhận dạng đã có ảnh trong tập mẫu). Các kết quả thực nghiệm đã chứng tỏ rằng phương pháp đề xuất cho kết quả tốt hơn các phương pháp truyền thống cả về sự hiệu quả và độ chính xác.

2. Phương pháp và thuật toán

Hệ thống được chia thành 3 modul: thu nhận và phân tích ảnh, trích đặc trưng, huấn luyện và nhận dạng ảnh

2.1. Khối thu nhận và phân tích ảnh

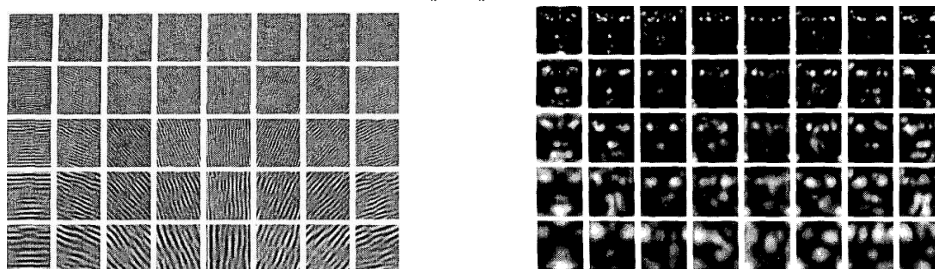
Đầu vào của khối nhận dạng ảnh là ảnh của người cần nhận dạng. Ảnh sau đó được thực hiện tiền xử lý để triệt nhiễu và nâng cao chất lượng ảnh nếu cần thiết. Ảnh sau tiền xử lý chỉ được nâng cao chất lượng (chứ không thay đổi tư thế và biểu cảm trên khuôn mặt) và đưa sang khối trích đặc trưng. Trong nghiên cứu này chúng tôi thực nghiệm với các ảnh sạch không nhiễu nên không áp dụng khối tiền xử lý nhiễu. Bước tiếp theo trước khi trích đặc trưng ảnh là phải xác định một vùng thông tin quan trọng trong ảnh. Trong nghiên cứu này, chúng tôi xây dựng vùng thông tin quan trọng này là một lưới hình chữ nhật đặt bên trong vùng hình ảnh khuôn mặt. Dựa vào lưới chữ nhật này, mạng wavelet được xác định cho chúng ta một tập hợp các điểm đặc trưng. Tăng số lượng điểm đặc trưng sẽ biểu diễn ảnh chính xác hơn nhưng cũng tăng chi phí tính toán và ngược lại. Do vậy, xác định số lượng đặc trưng dung hòa được hai yếu tố độ chính xác và phức tạp tính toán là điều rất quan trọng.

2.2. Khối trích đặc trưng

Khối này là khối rất quan trọng của hệ thống nhận dạng. Trong nghiên cứu này, chúng tôi áp dụng phương pháp trích đặc trưng dùng biến đổi Gabor wavelet trong [5]. Theo đó, ảnh khuôn mặt được xử lý bằng biến đổi Gabor wavelet để trích ra các đặc trưng. Một tập các tần số và hướng của các điểm đặc trưng xác định bởi mạng wavelet sẽ là thông tin đặc trưng để biểu diễn ảnh. Chúng ta chọn hai tần số chặn trên và chặn dưới dựa vào phổ của ảnh f_{LB} và f_{UB} . Một tập P tần số f_i được sử dụng cho mỗi điểm đặc trưng bắt đầu từ f_{LB} và nhân liên tiếp với 2 cho tới f_{UB} . Với mỗi tần số một tập Q hướng θ_j được chọn từ $-\pi$ đến π . Các điểm đặc trưng có thể là tất cả các điểm ảnh nhưng cũng có thể chỉ chọn một số điểm ảnh “quan trọng” trên khuôn mặt được xác định bằng (c_x, c_y) trên trục tọa độ Đề các. Như vậy, một tập P tần số và Q hướng xác định mỗi điểm đặc trưng. Số lượng các điểm đặc trưng là R, số lượng wavelet sẽ là $N = P*Q*R$. Với $I(x,y)$ biểu diễn một điểm ảnh, $0 < x < h, 0 < y < w$ với h, w là chiều cao và độ rộng của ảnh. Một họ các hàm N Gabor wavelet $\psi = \{ \psi_{1,1,1}, \dots, \psi_{P,Q,R} \}$ được xác định như sau:

$$\psi_{i,j,k}(x, y) = \frac{f_i^2}{2\pi} \exp\{-0.5f_i^2[(x-c_{x_k})^2 - (y-c_{y_k})^2]\} * \sin\{2\pi f_i[(x-c_{x_k})\cos\theta_j - (y-c_{y_k})\sin\theta_j]\} \quad (1)$$

Trong đó f_i là tần số, θ_j là hướng và c_{x_k}, c_{y_k} chỉ ra vị trí của điểm đặc trưng.



(a) (b)
Hình 1. Biến đổi Gabor wavelet của một ảnh mặt người mẫu
a, Phần thực của wavelet; b, Độ lớn wavelet

Hàm wavelet $\psi_{i,j,k}(x, y)$ được chuẩn hóa thành $\hat{\psi}_{i,j,k}(x, y)$ với

$$\hat{\psi}_{i,j,k}(x, y) = \frac{\psi_{i,j,k}(x, y)}{\|\psi_{i,j,k}(x, y)\|} \quad (2)$$

Hàm wavelet ψ là một hàm phức và toán tử $\|\cdot\|$ là toán tử xác định độ lớn hàm phức.

Tập trọng số $\omega_1, \dots, \omega_u$ ứng với wavelet ψ_1, \dots, ψ_u , với $2 \leq u \leq N$.

$$\omega_1 = I * \hat{\psi}_1 \quad (3)$$

$$\omega_u = I_{diff_u} * \hat{\psi}_u \quad (4)$$

$$I_{diff_1} = I(x, y) \quad (5)$$

$$I_{diff_u} = I_{diff_{u-1}} - \hat{I}_{u-1} \quad (6)$$

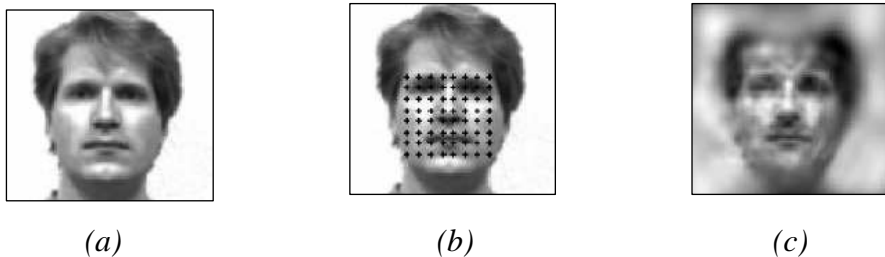
Trong đó I_{diff} biểu diễn sự sai khác giữa ảnh gốc và ảnh tái tạo ứng với các wavelet, \hat{I}_u là ảnh tái tạo từ ứng với wavelet u

$$\hat{I}_u = \omega_u * \hat{\psi}_u \quad (7)$$

Ảnh tái tạo cuối cùng

$$\hat{I} = \sum_{u=0}^{N-1} \hat{I}_u \quad (8)$$

Ảnh gốc I , mạng wavelet và ảnh tái tạo \hat{I} được trình bày trên hình 1.



Hình 2. Ảnh gốc (a), mạng wavelet (b) và ảnh tái tạo (c) với $N=100$

2.3 Khối huấn luyện và nhận dạng ảnh dùng mạng neural

Loại mạng neural điển hình được dùng trong xử lý ảnh là mạng neural perception nhiều lớp MLP (multi-layer perception). Mỗi neural trong một lớp kết nối trực tiếp với các neural ở lớp dưới.

Mạng neural sử dụng rộng rãi để huấn luyện, học mẫu là mạng lan truyền ngược (back propagation). Ở đây, giá trị đầu ra được so sánh với kết quả đúng để tính giá trị của hàm sai số xác định trước. Giá trị hàm sai số này sau đó được truyền ngược lại mạng để điều chỉnh trọng số của mỗi kết nối để cực tiểu hóa lỗi. Sau khi lặp lại quá trình này với số lần huấn luyện đủ lớn sẽ hội tụ đến các trạng thái ít lỗi nhất. Để điều chỉnh trọng số một cách chính xác, phương pháp tối ưu phi tuyến gọi là kỹ thuật *giảm gradient* được áp dụng. Theo đó, đạo hàm của hàm sai số tương ứng với các trọng số của mạng được tính và trọng số sẽ thay đổi để giảm giá trị hàm sai số.

Trong nghiên cứu này, chúng tôi sử dụng tập đặc trưng Gabor wavelet của ảnh mặt người được xác định như trong [5]. Tuy nhiên chúng tôi kết hợp với huấn luyện và nhận dạng bằng mạng neural nhiều lớp truyền thẳng MLP sử dụng thủ tục huấn luyện lan truyền ngược. Mạng neural có thể được cấu hình ở chế độ huấn luyện hoặc nhận dạng. Ban đầu mạng được khởi tạo ở chế độ huấn luyện và tất cả các ảnh có trong tập mẫu sẽ được huấn luyện. Các thông số như số lượng neural ẩn, momentum, tỉ lệ học, số lần lặp và đầu ra mạng được biến đổi trong

quá trình huấn luyện để tối ưu quá trình huấn luyện. Trong chế độ nhận dạng, đầu ra của mạng neural là mẫu tìm thấy phù hợp nhất. Các tham số thực nghiệm của mạng wavelet và neural được cài đặt cụ thể trong hệ thống nhận dạng thử nghiệm trình bày chi tiết trong phần sau.

3. Xây dựng hệ thống nhận dạng ảnh mặt người thử nghiệm

Chúng tôi chọn dữ liệu ảnh thử nghiệm của 8 người khác nhau từ thư viện ảnh của Yale và Olivetti, ảnh được cắt về kích cỡ thống nhất là 92x112, mỗi người chọn 4 ảnh khác nhau, 1 ảnh dùng cho huấn luyện và 3 ảnh để kiểm tra kết quả nhận dạng. Ảnh để huấn luyện chọn là ảnh có độ sáng đồng nhất, khuôn mặt bình thường không biểu cảm, tư thế đầu thẳng ngay ngắn. Các ảnh chọn để nhận dạng kiểm nghiệm có các biểu cảm khác nhau trên khuôn mặt, độ sáng trên ảnh không đồng nhất.



Hình 3. Ảnh huấn luyện



Hình 4. Các ảnh dùng để nhận dạng thử nghiệm với độ sáng, biểu cảm trên khuôn mặt, tư thế khuôn mặt khác nhau

Trong thực nghiệm chúng tôi xác định mạng wavelet dựa trên hình chữ nhật nội tiếp lớn nhất trong ảnh, đó có thể coi là hình chữ nhật với bốn điểm góc là bốn điểm biên sát với 2 mắt và 2 góc quai hàm trên khuôn mặt. Mạng wavelet sau đó được chia ra thành 8x8 điểm xác định 64 điểm đặc trưng trên khuôn mặt ($R=64$).

Ở mỗi điểm đặc trưng, 6 tần số ($P=6$) và 10 hướng ($Q=10$) được chọn. Số chiều của vector đặc trưng khi đó sẽ bằng $P*Q*R = 64*6*10 = 3840$ là một con số khá lớn, chúng ta sẽ chọn 50 giá trị đặc trưng lớn nhất trong số 3840 giá trị này làm đầu vào cho mạng neural. Các tham số của mạng neural trong quá trình huấn luyện được đặt như sau: số lượng neural ẩn bằng 20% tổng số neural vào, hệ số học và momentum đều được đặt là 0.5. Chúng tôi dùng hai phương án thử nghiệm. Ở phương án thử nghiệm đầu tiên, mức ngưỡng lỗi được đặt là 0.0001, số bước lặp là 250001, mức lỗi được kiểm tra sau mỗi 10000 bước lặp. Phương án thử nghiệm thứ 2, mức ngưỡng lỗi được đặt là 0.001, số bước lặp là 25001, mức lỗi được kiểm tra sau mỗi 1000 bước lặp. Quá trình huấn luyện sẽ dừng lại khi mức lỗi đạt tới ngưỡng hoặc thực hiện hết số lần lặp cho ra kết quả đầu ra của mạng. Tương tự, quá trình nhận dạng sẽ thực hiện cho đến khi cho ra kết quả đầu ra của mạng neural là mẫu tìm thấy phù hợp nhất.

Hệ thống nhận dạng thử nghiệm được xây dựng trên MATLAB 7.01 để tận dụng khả năng hỗ trợ các hàm xử lý Neural và Wavelet của MATLAB trong hai toolbox Neural Network và Wavelet. Hệ thống được kết hợp xây dựng ở chế độ GUI để dễ dàng cho người sử dụng.

4. Đánh giá chương trình thử nghiệm

Máy tính sử dụng để đánh giá hệ thống là PC Intel Pentium M 1.6 GHz, 512 RAM, card tăng tốc đồ họa NVIDIA GeForce 4 MX 4000 128 MB. Thời gian nhận dạng thử nghiệm với tập mẫu ảnh của 8 người với phương án một là xấp xỉ 40 giây, phương án 2 là xấp xỉ 5 giây.


Các kết quả thử nghiệm cho thấy hệ thống đặc biệt hiệu quả khi nhận dạng các ảnh mang kính, mang các kiểu tóc khác nhau, tư thế khuôn mặt thay đổi chút ít (nhưng không thay đổi góc chụp của camera). Với số lượng người trong tập huấn luyện nhỏ hơn 5, tỉ lệ nhận dạng đúng trong cả hai phương án thử nghiệm là 100%.

Khi số lượng người trong tập mẫu lớn hơn 5, hiệu quả nhận dạng ảnh khi thay đổi tư thế khuôn mặt và biểu cảm giảm xuống. Với phương án một, tỷ lệ nhận dạng đúng là xấp xỉ 90% với tập mẫu 10 người, kết quả tỉ lệ nhận dạng đúng của hệ thống với 8 người trong tập mẫu là xấp xỉ 92%. Với phương án hai, tỷ lệ nhận dạng đúng là xấp xỉ 88% với tập mẫu 10 người, kết quả tỉ lệ nhận dạng đúng của hệ thống với 8 người trong tập mẫu là xấp xỉ 90%

So sánh với các nghiên cứu gần đây trong [1], [2], [3] và [5], nếu sử dụng phương án một, tỉ lệ nhận dạng có cao hơn, nhưng xét về thời gian thực thi thì lại kém hiệu quả. Trong phương án hai, tỉ lệ nhận dạng đúng là cao hơn [1], [2], [3] và tương đương [5]. Như vậy thông qua các kết quả thực nghiệm có thể kết luận: phương pháp sử dụng trong nghiên cứu hiệu quả hơn các phương pháp gần đây, đặc biệt là khi tối ưu được các tham số của mạng neural.

Các hạn chế cần phải khắc phục của hệ thống là hệ thống không thật sự hiệu quả với các khuôn mặt thay đổi biểu cảm rõ rệt, tư thế đầu quay với góc quay lớn, không sử dụng được để nhận dạng các ảnh với góc chụp camera rất khác nhau. Thời gian nhận dạng là thích hợp với nhận dạng ảnh offline nhưng chưa đủ thích hợp cho các ứng dụng nhận dạng online.

5. Kết luận

Trong bài báo này, chúng tôi đã trình bày về hệ thống nhận dạng ảnh mặt người dùng biến đổi Gabor wavelet và mạng neural nhiều lớp MLP. Hệ thống đã tỏ ra hiệu quả hơn các phương pháp nhận dạng mặt người truyền thống cả về độ chính xác và sự hiệu quả. Kết quả nghiên cứu này có thể được sử dụng để phát triển các ứng dụng liên quan đến nhận dạng ảnh mặt người đặc biệt là các ứng dụng trong an ninh như tìm kiếm, xác thực tội phạm qua ảnh 

Tóm tắt

Nhận dạng mặt người với sự thay đổi tư thế, độ sáng, biểu cảm của khuôn mặt đang là một thách thức. Hiện có nhiều phương pháp nhận dạng mặt người, mỗi phương pháp có ưu điểm và hạn chế nhất định. Bài báo này trình bày một phương pháp nhận dạng mặt người dùng mạng Neural và biến đổi Gabor wavelet. Các kết quả thực nghiệm chứng tỏ phương pháp đề xuất hiệu quả và chính xác hơn các phương pháp truyền thống.

Summary

Face recognition with variant pose, illumination and expression is a challenging problem. Today there exist many well-known techniques for face recognition, each with its own inherent limitations. In this paper, we present a novel approach to face recognition using Neural network and Gabor wavelets. Experimental results show that the proposed method performs better than traditional approaches in terms of both efficiency and accuracy.

Tài liệu tham khảo

- [1]. M. Turk and A. Pentland (1991), "Face recognition using Eigenfaces", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586-591.
- [2]. N. B. Peter, P. H. Joao and J. K. David (1997), "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711-720.
- [3]. F. S. Samaria (1994) "Face recognition using hidden Markov models," PhD thesis, University of Cambridge, UK,.
- [4]. P. S. Penev and J. J. Atick (1996) "Local feature analysis: A statistical theory for object representation," Network: Computation in Neural Systems, Vol. 7, pp. 477-500.
- [5]. Vinay Kumar and B. Shreyas (2007), "Face Recognition Using Gabor Wavelets", IEEE - ICSCN 2007, MIT Campus, Anna University, Chennai, India. Feb. 22-24, 2007. pp.433-437.