

ỨNG DỤNG PHƯƠNG PHÁP PHÂN TÍCH THÀNH PHẦN CHÍNH VÀ MẠNG NEURON NHÂN TẠO ĐỂ NHẬN DẠNG NGÔN NGỮ KÝ HIỆU

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS METHOD AND ARTIFICIAL NEURAL NETWORK FOR IDENTIFYING SIGN LANGUAGE

Huỳnh Hữu Hưng, Tống Thị Hồng Ngọc, Nguyễn Trọng Nguyên

Trường Đại học Bách khoa, Đại học Đà Nẵng

hhhung@dut.udn.vn, hongngoc219.90@gmail.com, ntnguyen.dn@gmail.com

TÓM TẮT

Ngôn ngữ ký hiệu là ngôn ngữ chính được sử dụng trong cộng đồng người khiếm thính để thay thế cho ngôn ngữ nói truyền thống. Ngoài ra, ngôn ngữ ký hiệu còn được sử dụng trong việc tương tác giữa người và máy. Trong bài báo này, chúng tôi đề xuất một giải pháp nhận dạng ngôn ngữ ký hiệu dựa trên phương pháp phân tích thành phần chính và mạng neuron nhân tạo. Đầu tiên bàn tay được phát hiện, sau đó một số bước tiền xử lý được thực hiện để nâng cao chất lượng ảnh, tiếp đến chúng tôi tính toán một không gian mới để biểu diễn ảnh sao cho số chiều nhỏ hơn không gian ban đầu mà đặc trưng của ảnh vẫn được giữ lại tối đa, cuối cùng mạng neuron nhân tạo được dùng để huấn luyện và nhận dạng ảnh bằng cách ánh xạ ảnh đầu vào vào trong không gian tìm được. Giải pháp này có chi phí tính toán thấp và có thể thực thi việc nhận dạng theo thời gian thực. Kết quả thử nghiệm cho thấy độ ổn định và chính xác cao.

Từ khóa: ngôn ngữ ký hiệu; cử chỉ; màu da; phân tích thành phần chính; trị riêng; vector riêng

ABSTRACT

Sign language is used in the deaf community to replace the traditional spoken language. In addition, sign language is also used in the interaction between people and machines. In this paper, we propose a solution to recognize sign language using the principal component analysis method and artificial neural network. At first, the hand is detected, then some preprocessing steps are taken to improve the image quality, then we determine a new space to represent the hand image with fewer dimensions than the original space and characteristics of the image are kept, and the artificial neural network is used for training and testing with the mapped input image. This solution has low computational cost, and can execute in real-time. The proposed approach has been tested with high accuracy and stability.

Key words: sign language; gesture; skin color; PCA; eigenvalue; eigenvector

1. Đặt vấn đề

Ngôn ngữ ký hiệu là một trong những phương pháp giao tiếp được sử dụng bởi những người khiếm thính. Ngôn ngữ này sử dụng cử chỉ được thực hiện bằng cách cử động bàn tay kết hợp với nét mặt hay điệu bộ của cơ thể.

Lĩnh vực nghiên cứu về nhận dạng ngôn ngữ cử chỉ đang được đẩy mạnh để giúp người khiếm thính dễ dàng hòa nhập cộng đồng. Hầu hết các nhà nghiên cứu sử dụng thiết bị chuyên dụng như găng tay, cảm biến hoặc nhận dạng dựa trên kỹ thuật xử lý ảnh thông qua camera và máy tính. Các giải pháp xử lý ảnh thường dựa trên hai phương pháp chính: tập luật và máy học.

Trong bài báo này chúng tôi đề xuất một

hướng tiếp cận mới trong lĩnh vực máy học có thể khái quát những cử chỉ tay dạng tĩnh bằng phương pháp phân tích thành phần chính kết hợp với mạng neuron nhân tạo.

2. Các nghiên cứu liên quan

Một số đề tài đã đề xuất phương pháp nhận dạng cử chỉ bằng cách sử dụng máy ảnh và các kỹ thuật xử lý ảnh. Mục tiêu chung của các nghiên cứu này là để giúp người khuyết tật giao tiếp với nhau, thay thế ngôn ngữ truyền thống bằng ngôn ngữ cử chỉ. Một ứng dụng khác của ngôn ngữ cử chỉ là tương tác người – máy, trong đó cử chỉ được dùng như dữ liệu đầu vào, thông tin được truyền đến máy tính thông qua webcam. Fujisawa [1] đã phát triển một thiết bị truyền HID để thay thế cho chuột máy tính dành

cho người tàn tật. Bretzner [2] đã phát triển một hệ thống mà người dùng có thể điều khiển TV, đầu DVD dựa trên cử chỉ tay thông qua một camera. Marshall [3] thiết kế một hệ thống để hỗ trợ người dùng tương tác với các hệ thống đa phương tiện dựa trên những cử chỉ. Malima [4] đề xuất một thuật toán tự động xác định một số cử chỉ tay dùng cho mục đích điều khiển robot.

Để nhận dạng cử chỉ tay, việc lựa chọn các đặc trưng là rất quan trọng bởi các cử chỉ tay rất đa dạng trong hình dạng, sự biến đổi, sự chuyển động và kết cấu. Hầu hết các tính năng được sử dụng trong các nghiên cứu trước đây được trích xuất từ ba phương pháp sau đây.

Phương pháp tiếp cận dựa trên mô hình: phương pháp này cố gắng suy ra tư thế của lòng bàn tay và các góc khớp, có thể tạo ra một mô hình 3D của một bàn tay, lý tưởng cho các tương tác trong môi trường thực tế ảo. Ueda [5] áp dụng phương pháp này để tương tác với một đối tượng trong không gian ảo, khu vực bàn tay được xác định dựa trên một hệ thống nhiều camera cùng hoạt động. Utsumi [6] dựa vào hình dạng và chuyển động của bàn tay để điều khiển các đối tượng trong thế giới ảo. Bettio [7] trình bày một cách phát triển môi trường tương tác cho phép con người tương tác với các mô hình 3D phức tạp mà không cần phải tự vận hành thiết bị đầu vào. Trong phương pháp tiếp cận dựa trên mô hình, nhiều là một vấn đề có ảnh hưởng lớn đến kết quả thực hiện, đồng thời đòi hỏi nhiều thời gian hơn để thiết kế các hệ thống.

Một phương pháp khác được thực hiện dựa trên chuỗi những ảnh 2D thu được tại những góc nhìn khác nhau ở cùng một thời điểm, cử chỉ được mô hình hóa như là một chuỗi các góc nhìn. Một sự kết hợp tuyến tính của những hình ảnh này có thể được sử dụng để khoảng tái tạo lại cử chỉ thực tế. Cách tiếp cận này được sử dụng trong rất nhiều nghiên cứu nhận dạng cử chỉ tay, chẳng hạn như [8]. Dù vậy, tỉ lệ lỗi của phương pháp này lên đến hơn 7%, đồng thời việc thu thập ảnh huấn luyện cũng là một vấn đề.

Đặc trưng cấp thấp: Một số nhà nghiên cứu trình bày các phương pháp nhận dạng dựa trên những đặc trưng đơn giản, cho rằng thông tin chi tiết về hình dạng tay là không cần thiết.

Họ nhận thấy rằng tất cả các bàn tay con người có màu sắc nằm trong cùng một miền giá trị và sự khác nhau chủ yếu là ở độ sáng. Để nhận dạng, bàn tay được xác định dựa trên màu sắc và các đặc trưng được trích xuất dựa vào tọa độ tâm bàn tay, độ lệch tâm, hay góc của trục khung ellipse bao bàn tay [4].

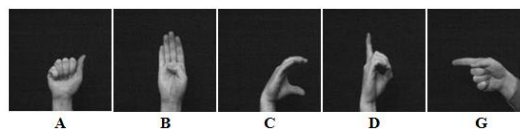
3. Phương pháp đề xuất

3.1. Dữ liệu vào và dữ liệu thử nghiệm

Dữ liệu có thể là một hình ảnh hoặc một chuỗi các hình ảnh (video), được thực hiện bởi một camera duy nhất hướng về phía bàn tay thực hiện cử chỉ. Một số hệ thống sử dụng hai hoặc nhiều camera để có được thêm thông tin về bàn tay, ưu điểm là mỗi hệ thống có thể nhận ra những cử chỉ ngay cả khi tay bị che khuất nhưng bù lại, chi phí tính toán cao, khó áp dụng vào hệ thống nhận dạng theo thời gian thực. Các giai đoạn tiếp theo của quá trình nhận dạng sẽ ít phức tạp hơn nếu ảnh thu được có nền đơn giản và độ tương phản cao so với bàn tay. Vì vậy, ảnh thường được thu trong một môi trường nền đồng nhất, và hạn chế bóng xuất hiện trong ảnh thu được. Dữ liệu chúng tôi dùng để thực hiện đề tài được thu thập từ nhiều nguồn dữ liệu mở.

3.1.1. Ảnh

Ảnh được thu thập từ một số bộ dữ liệu mở [14][15] hoặc trực tiếp từ webcam.



Hình 1. Một số ảnh từ tập dữ liệu [14]

3.1.2. Video

Video được ghi lại từ một webcam cố định, với nền đơn giản và ánh sáng ổn định. Một người thực hiện một số cử chỉ trước webcam. Để việc xác định bàn tay đơn giản, cánh tay được che lại và người thực hiện không lộ khuôn mặt trên webcam. Các video được thử nghiệm ở định dạng AVI (Audio Video Interleave).

3.2. Tiền xử lý

Dưới đây là các bước cần thiết để trích lọc đối tượng, nâng cao chất lượng ảnh và chuẩn hóa nhằm nâng cao hiệu quả nhận dạng.

3.2.1. Phát hiện tay

Để xác định các cử chỉ tay, bước đầu tiên là phát hiện tay trong ảnh đầu vào. Trong giải pháp được đề xuất, chúng tôi sử dụng phương pháp lọc màu da. Theo Fleck và Forsyth [16], màu da của con người được tạo thành bởi hai cực màu sắc: màu đỏ (máu) và màu vàng (melanin), với độ bão hòa trung bình. Fleck cũng nhận thấy rằng màu da là cấu trúc biên độ thấp. Màu sắc đặc trưng của da có thể được sử dụng để phát hiện tay người.



Hình 2. Kết quả lọc màu da

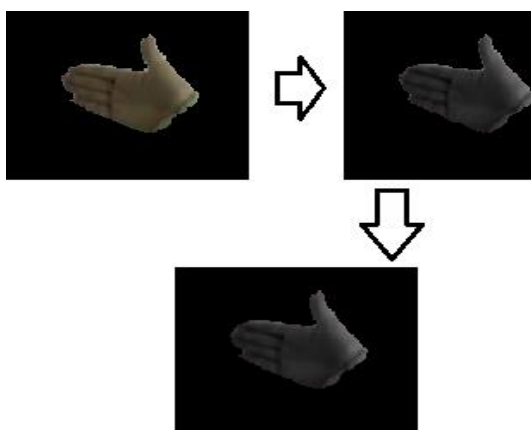
3.2.2. Chọn đối tượng lớn nhất

Bước này giúp giữ lại một đối tượng duy nhất trên hình ảnh. Đối với hệ thống xác định cử chỉ tay, đối tượng lớn nhất xuất hiện trong ảnh là bàn tay. Vì vậy, sau bước này, các đối tượng không liên quan sẽ được loại bỏ.



Hình 3. Giữ lại đối tượng lớn nhất

3.2.3. Sử dụng bộ lọc trung vị cho ảnh xám



Hình 4. Lọc trung vị ảnh xám

Trong xử lý ảnh số, khử nhiễu là bước cần thiết để mang lại kết quả tốt như mong muốn. Lọc trung vị là kỹ thuật lọc phi tuyến, thường

được sử dụng để loại bỏ nhiễu. Khử nhiễu là một bước tiền xử lý điển hình để cải thiện kết quả xử lý ảnh sau này (ví dụ, phát hiện cạnh đối tượng trên ảnh). Lọc trung vị được sử dụng rất rộng rãi vì, theo một số điều kiện nhất định, nó giữ lại được các cạnh trong khi loại bỏ nhiễu.

3.2.4. Loại bỏ cánh tay

Trong một số trường hợp, kết quả lọc màu da chứa cả cánh tay. Vì cánh tay không được sử dụng trong quá trình nhận dạng nên chúng cần được loại bỏ khỏi ảnh. Bước này không chỉ giúp nâng cao độ chính xác mà còn giảm chi phí tính toán khi xử lý – một yếu tố quan trọng trong các ứng dụng thời gian thực.

Đầu tiên, tay được tách riêng bởi khung chữ nhật bao quanh. Sau đó vị trí cổ tay được xác định, bàn tay và cánh tay sẽ được tách nhau tại vị trí đó. Giải thuật xác định vị trí cổ tay chúng tôi đề xuất được trình bày dưới đây:

Gọi m_i là bề rộng của đối tượng tại dòng i

- B_1 : tính m_i ở dòng cuối cùng
- B_2 : tính m_{i-1} ở dòng bên trên
- B_3 : nếu m không tăng thì quay lại bước B_2 , ngược lại thì cổ tay là dòng ngay dưới dòng hiện tại



Hình 5. Xác định cổ tay và tách bàn tay

3.2.5. Chuẩn hóa độ sáng ảnh

Bước này cải thiện độ tương phản, giúp cân bằng độ sáng của ảnh. Điều này giúp độ chính xác không bị ảnh hưởng khi ảnh được thu trong các điều kiện ánh sáng khác nhau.



Hình 6. Chuẩn hóa độ sáng ảnh

3.2.6. Biến đổi kích thước ảnh

Bước này chuẩn hóa kích thước bàn tay trên ảnh, chuẩn bị cho quá trình huấn luyện và

nhận dạng, đồng thời tăng độ chính xác. Nếu sử dụng các phương pháp biến đổi kích thước thông thường, bàn tay có thể bị thu nhỏ hoặc kéo giãn về một tỷ lệ (ngang/dọc) khác so với ban đầu, khi đó đặc trưng của bàn tay và kết quả nhận dạng sẽ bị ảnh hưởng. Vì vậy chúng tôi đề xuất một bước bổ sung trước khi thực hiện chuẩn hóa kích thước ảnh.

Đầu tiên bàn tay được cắt riêng khỏi ảnh bởi khung bao chữ nhật. Giả sử chúng ta có một hình ảnh tay với kích thước $w \times h$, trong đó h là chiều cao và w là chiều rộng của hình ảnh. α được định nghĩa là độ chênh lệch giữa w và h . Chúng tôi đề xuất một phương pháp để điều chỉnh kích thước bàn tay như sau:

- Nếu $h > w$ thì $\alpha = h - w$
 - Chèn $\alpha/2$ cột vào bên trái ảnh
 - Chèn $\alpha/2$ cột vào bên phải ảnh
- Nếu $h < w$ thì $\alpha = w - h$
 - Chèn $\alpha/2$ dòng vào bên trên ảnh
 - Chèn $\alpha/2$ dòng vào bên dưới ảnh

Sử dụng các phương pháp hiệu chỉnh kích thước thông dụng để đưa ảnh thu được về kích thước 50×50 .

Một ví dụ về hiệu chỉnh kích thước ảnh với $h < w$ được trình bày trong hình dưới đây:



Hình 7. Hiệu chỉnh kích thước ảnh với $h < w$

3.3. Nhận dạng cử chỉ

3.3.1. Trích xuất đặc trưng

Để trích xuất đặc trưng cho các ảnh trong tập mẫu, chúng tôi sử dụng phương pháp phân tích thành phần chính [17] như sau:

Bước 1: ta có tập S với M ảnh mẫu, mỗi ảnh được chuyển sang dạng vector kích thước $N = 2500$ phần tử (50×50)

$$S = \{\Gamma_1, \Gamma_2, \Gamma_3 \dots \Gamma_M\} \quad (1)$$

Bước 2: tính toán ảnh trung bình

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (2)$$

Bước 3: tính độ sai khác giữa mỗi vector ảnh mẫu với vector ảnh trung bình

$$\Phi_i = \Gamma_i - \Psi \quad (3)$$

Bước 4: tính ma trận hiệp phương sai C theo công thức sau:

$$C = \frac{1}{M} \sum_{i=1}^M \Phi_i \Phi_i^T = \frac{1}{M} A A^T \quad (4)$$

Trong đó $A = [\Phi_1, \Phi_2, \Phi_3 \dots \Phi_M]$

Bước 5: xác định M vector trực giao u_i có thể mô tả tốt nhất sự phân bố của tập dữ liệu. Tuy nhiên kích thước của ma trận C quá lớn (2500×2500). Vì vậy để tìm u_i , chúng ta xác định các vector riêng v_i của ma trận $L = A^T A$ có kích thước $M \times M$.

Bước 6: tính các vector riêng của ma trận C dựa trên các vector riêng của ma trận L

$$u_i = A v_i \quad (5)$$

Bước 7: mỗi ảnh mẫu Γ được ánh xạ sang không gian mới và biểu diễn dưới dạng vector Ω

$$\omega = u_i^T (\Gamma - \Psi) \quad (6)$$

Trong đó $\Omega^T = [\omega_1, \omega_2, \omega_3 \dots \omega_M]$

Các vector thu được là vector đặc trưng sử dụng cho việc huấn luyện. Để nhận dạng một cử chỉ, ảnh được ánh xạ lên không gian mới bằng công thức (6) và vector thu được là đầu vào của mạng neuron.

3.3.2. Huấn luyện mạng và nhận dạng

Chúng tôi sử dụng mạng neuron nhân tạo vì những ưu điểm sau: khả năng tổng quát hóa, có thể tính toán phi tuyến, ít bị ảnh hưởng bởi nhiễu và mạng có khả năng xử lý song song.

Chúng tôi thiết kế một mạng Perceptron có ba lớp: lớp vào có số neuron tương ứng với kích thước của vector đặc trưng (số lượng ảnh huấn luyện), số neuron lớp ẩn được xác định bằng phương pháp thử-sai, và lớp ra có 23 neuron tương ứng với 23 cử chỉ cần nhận dạng.

Hàm chuyển được sử dụng là hàm tansig:

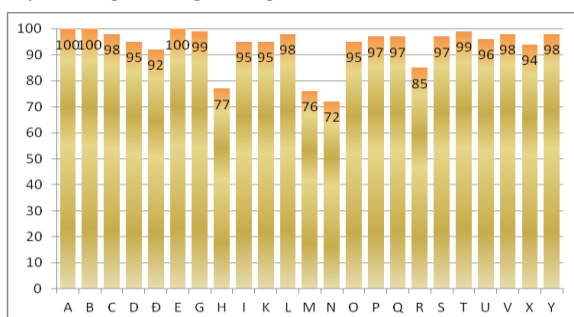
$$\text{tansig}(n) = \frac{2}{1 + e^{-2n}} - 1 \quad (7)$$

4. Kết quả thử nghiệm

Chúng tôi sử dụng webcam Logitech 9000, khoảng cách từ webcam đến bàn tay nằm trong khoảng 0.8 đến 1.2m. Hệ thống được xây dựng bằng ngôn ngữ C# sử dụng thư viện OpenCVSharp. Chúng tôi chọn 23 ký tự trong bảng chữ cái để nhận dạng, với 40 ảnh mẫu huấn luyện cho mỗi ký tự. Các ảnh này bao gồm những ảnh mẫu đã đề cập và các ảnh được chúng tôi thu thập trong thực tế.

Các cử chỉ được thử nghiệm bằng cách thực hiện trực tiếp trước webcam. Mỗi cử chỉ được nhận dạng 100 lần, tương ứng với tập 2300 cử chỉ nhận dạng. Tỷ lệ nhận dạng chính xác trung bình đạt 93.6% với 920 ảnh mẫu huấn luyện.

Kết quả cụ thể đối với từng ký tự được trình bày trong hình 8. Nhiều ký tự được nhận dạng chuẩn xác như A, B, E hoặc tỷ lệ chính xác cao (trên 90%) như C, D, Đ, G, I, K, L, O, P, Q, S, T, U, V, X, Y. Dù vậy, một số cử chỉ thường bị nhận dạng nhầm như H, M, N và R. Nguyên nhân là do độ tương đồng cao giữa đặc trưng của các cử chỉ (ví dụ như M và N). Một số lý do khác là độ nghiêng của bàn tay lớn và lòng bàn tay không hướng thẳng vào webcam.



Hình 8. Kết quả nhận dạng với 23 ký tự

5. Kết luận

Trong bài báo này, một hướng tiếp cận mới đã được đề xuất để nhận dạng ngôn ngữ ký hiệu. Hệ thống nhận dạng bao gồm các bước sau: phân đoạn ảnh, tiền xử lý, trích xuất đặc trưng, huấn luyện và nhận dạng. Trong bước phân đoạn, thông tin màu sắc được sử dụng để xác định các điểm ảnh có màu tương tự màu da. Quá trình tiền xử lý nâng cao chất lượng ảnh để phục vụ cho việc nhận dạng. Các đặc trưng của bàn tay được trích xuất bằng phương pháp phân tích thành phần chính, và mạng neuron nhân tạo được sử dụng để nhận dạng. Chúng tôi cũng đã đề xuất một phương pháp đơn giản để loại bỏ cánh tay, giúp nâng cao hiệu quả nhận dạng. Bên cạnh đó, kích thước ảnh bàn tay được chuẩn hóa đồng thời vẫn giữ nguyên tỉ lệ, qua đó giữ được đặc trưng quan trọng của cử chỉ, trong khi các giải pháp thường thấy đều điều chỉnh kích thước trực tiếp làm ảnh bị co giãn không đều và mất đi tỉ lệ ngang-dọc của đối tượng ban đầu. Ngoài ra hệ thống có thể thực thi theo thời gian thực.

Giải pháp của chúng tôi và một số phương pháp khác khi áp dụng trên cùng một bộ dữ liệu huấn luyện và thử nghiệm được so sánh về độ chính xác trong bảng dưới đây:

Bảng 1. So sánh độ chính xác các phương pháp

Chúng tôi	[9]	[10]	[11]	[12]	[13]
93.6%	84%	92.78%	90.45%	92.2%	90.45%

Trong các nghiên cứu tới, giải pháp này sẽ được cải tiến để nhận dạng cử chỉ tay động, hiệu quả ngay cả khi góc nghiêng bàn tay lớn và lòng bàn tay không hướng thẳng về camera.

TÀI LIỆU THAM KHẢO

- [1] Fujisawa, S. et al, "Fundamental research on human interface devices for physically handicapped persons", 23rd Int. Conf. IECON, New Orleans, 1997.
- [2] Soren Lenman, Lars Bretzner, Bjorn Thuresson, "Computer Vision Based Hand Gesture Interfaces for Human – Computer Interaction", Department of Numerical Analysis and Computer Science, June 2002.
- [3] M. Marshall, "Virtual Sculpture-Gesture Controlled System for Artistic Expression", Proceedings of the AISB 2004 COST287 - ConGAS Symposium on Gesture, Interfaces for Multimedia Systems, Leeds, UK, 2004, pp 58-63.

- [4] A. Malima, E. Ozgur, M. Cetin, "A fast algorithm for vision-based hand gesture recognition for robot control", *IEEE Conference on Signal Processing and Communications 2006*, 2006, pp.1-4.
- [5] Ueda E., "A Hand Pose Estimation for Vision-Based Human Interfaces", *IEEE Transactions on Industrial Electronics*, Vol. 50, No. 4, 2003, pp. 676–684.
- [6] Utsumi A., Ohya J., "Multiple Hand Gesture Tracking using Multiple Cameras", *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, 1999, pp.473–478.
- [7] Bettio, F. et al, "A Practical Vision-Based Approach to Unencumbered Direct Spatial Manipulation in Virtual Worlds", *Eurographics Italian Chapter Conf.*, 2007.
- [8] Gupta N. et al, "Developing a gesture based inter-face", *IETE, Journal of Research: Special Issue on Visual Media Processing*, 2002.
- [9] Mokhtar M. Hasan, Pramoud K. Mirsa, "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", *International Journal of Computer Science & Information Technology (IJCSIT)*, Vol. 3(2), 2011.
- [10] V. S. Kulkarni, S. D. Lokhande, "Appearance Based Recognition of American Sign Language Using Gesture Segmentation", *International Journal on Computer Science and Engineering (IJCSE)*, Vol. 2(3), 2010, pp. 560-565.
- [11] Shuying Zhao, Wenjun Tan, Shiguang Wen, Yuanyuan Liu, "An Improved Algorithm of Hand Gesture Recognition under Intricate Background", *Springer the First International Conference on Intelligent Robotics and Applications (ICIRA 2008)*, Part I., 2008, pp. 786–794.
- [12] Byung-Woo Min, Ho-Sub Yoon, Jung Soh, Yun-Mo Yang, Toshiaki Ejima, "Hand Gesture Recognition Using Hidden Markov Models", *IEEE International Conference on computational cybernetics and simulation*, Vol.5, 1997.
- [13] E. Stergiopoulou, N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique", *Elsevier Engineering Applications of Artificial Intelligence*, Vol. 22(8), 2009, pp. 1141 – 1158.
- [14] Sébastien Marcel - Hand Posture and Gesture Datasets: <http://www.idiap.ch/resource/gestures>
- [15] T.M's Gesture Recognition: <http://www-prima.inrialpes.fr/FGnet/data/12-MoeslundGesture>
- [16] Fleck M., Forsyth D., Bregler C., "Finding Naked People", *European Conference on Computer Vision*, 1996.
- [17] Lindsay I Smith – A tutorial on Principal Component Analysis, Feb. 2002.

(BBT nhận bài: 10/08/2013, phản biện xong: 10/09/2013)