

MỘT PHƯƠNG PHÁP PHÂN CỤM KHUÔN MẶT HIỆU QUẢ TRÊN MẠNG XÃ HỘI

Nguyễn Hữu Quỳnh*
Trường Đại học Điện lực

TÓM TẮT

Trong những năm gần đây, lượng thông tin trên mạng xã hội đang phát triển như vũ bão, chỉ tính riêng trên mạng facebook đã có hàng trăm tỷ bức hình. Do đó, xử lý các nguồn dữ liệu này để trợ giúp người dùng trong việc phát hiện tri thức và khai phá dữ liệu sẽ vô cùng cần thiết. Bài báo này trình bày phương pháp phân cụm các khuôn mặt trong một tập ảnh khuôn mặt đã có dựa vào đặc trưng là các thành phần chính được trích rút bằng thuật toán PCA. Sau đó sử dụng thuật toán phân cụm phân cấp (HAC) để phân cụm các khuôn mặt vào các cụm riêng biệt. Nghiên cứu đã thực nghiệm trên tập ảnh gồm 100 ảnh. Các kết quả thực nghiệm cho thấy phương pháp mới đề xuất cho kết quả với độ chính xác tốt.

Từ khóa: phân cụm phân cấp; phân tích thành phần chính; khai phá dữ liệu; khuôn mặt; phân cụm

GIỚI THIỆU

Hiện nay thế giới có hàng trăm mạng xã hội khác nhau như MySpace và Facebook nổi tiếng trong thị trường Bắc Mỹ và Tây Âu; Orkut và Hi5 tại Nam Mỹ; Friendster tại Châu Á và các đảo quốc Thái Bình Dương. Một số mạng xã hội khác đã gặt hái được thành công đáng kể theo vùng miền như Bebo tại Anh Quốc, CyWorld tại Hàn Quốc, Mixi tại Nhật Bản. Ở Việt Nam xuất hiện rất nhiều các mạng xã hội như: Facebook, Zing Me, YuMe, Tamtay. Với số lượng mạng xã hội đông đảo như thế, lượng thông tin dữ liệu thu được là khổng lồ. Trong lượng thông tin khổng lồ này, có một lượng lớn là hình ảnh. Một minh chứng rõ nhất là mạng xã hội facebook, cho đến nay đã có hàng trăm tỷ bức hình trong cơ sở dữ liệu. Việc tìm ra thông tin hữu ích trên lượng dữ liệu hình ảnh lớn như vậy sẽ rất cấp thiết.

Nhiều thông tin được chia sẻ trên mạng xã hội thể hiện bằng các hình ảnh cung cấp cho người dùng về thông tin của người, cảnh,... Tuy nhiên, mỗi khi một người dùng muốn tìm hiểu thông tin về một ai đó gặp phải vấn đề phải tìm thông tin về người đó rất khó khăn (tốn thời gian và nhiều khi không tìm được). Lý do của việc này là lượng ảnh trên mạng xã

hội quá nhiều và tăng nhanh hàng ngày. Do đó, một hệ thống có thể giúp gom các đối tượng ảnh khuôn mặt về cùng một cụm (theo một độ đo tương tự nào đó) trong một tập dữ liệu ảnh khổng lồ là vô cùng cần thiết.

Trong bài báo này tôi đề xuất phương pháp phân cụm khuôn mặt. Các nghiên cứu liên quan sẽ được tôi mô tả tại mục tiếp theo. Sau đó trình bày phương pháp phân cụm khuôn mặt của tôi. Tiếp đến là phần xây dựng tập dữ liệu và kết quả thực nghiệm. Cuối cùng là phần kết luận.

CÁC NGHIÊN CỨU LIÊN QUAN

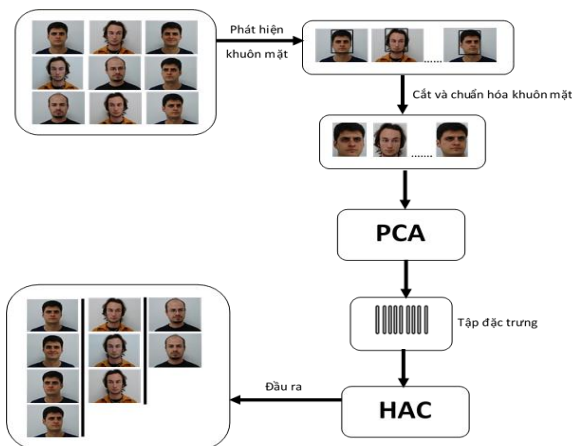
Phân cụm [1,2] có thể được coi như một hình thức nhận dạng không giám sát trên một tập dữ liệu hạn các đối tượng dựa trên một số độ đo tương tự hay độ đo khoảng cách [7,9,12,13]. Phương pháp phân cụm khuôn mặt dựa trên những đặc trưng xuất hiện khuôn mặt [4,5,6] được nghiên cứu rộng rãi và có sự tiến bộ đáng kể đã đạt được trong hai thập kỷ vừa qua. Các phương pháp phân cụm khuôn mặt khác nhau cũng được sự quan tâm của nhiều tác giả. Ở trong tài liệu [10] tác giả đề xuất một phương pháp phân cụm khuôn mặt sử dụng đặc trưng SIFT và phân cụm phân cấp tích tụ, cho ta thấy sự hiệu quả của việc phân cụm với đặc trưng mô tả bậc thấp. Một cách tiếp cận phân cụm khác, Fitzgibbon and Zisserman [3] có đề xuất một cách tiếp cận có

* Email: quynhnh@epu.edu.vn

liên quan đến khoảng cách Joint Manifold (JMD). Trong phương pháp đề xuất mỗi không gian con đại diện cho một tập hợp các khuôn mặt của cùng một người. Mặt khác, Eickeler[11] đã đề xuất một phương pháp phân cụm khuôn mặt, được gọi là Hidden Markov Models-clustering (HMM-clustering), tức là một phân cụm K-means sử dụng mô hình Markov ẩn để đại diện cho một mẫu cụm.

Trong bài báo tôi kết hợp việc sử dụng phương pháp phân tích thành phần chính PCA để trích rút đặc trưng và phân cụm phân cấp tích tụ để phân cụm khuôn mặt vào những nhóm tương đồng.

PHÂN CỤM KHUÔN MẶT



Hình 1. Sơ đồ tổng quan của hệ thống

Trong Hình 1, với tập dữ liệu ảnh khuôn mặt đầu vào tôi sử dụng đặc trưng Haarlike để phát hiện ra khuôn mặt trong mỗi bức ảnh. Sau đó tôi sử dụng thuật toán PCA để giảm số chiều của dữ liệu đồng thời trích rút những thành phần chính đảm bảo được đầy đủ thông tin của khuôn mặt trong ảnh. Cuối cùng dựa trên tập đặc trưng tôi áp dụng thuật toán phân cụm phân cấp tích tụ (HAC) để thu được kết quả cuối cùng là các khuôn mặt tương tự nhau được gom cùng một cụm. Chi tiết thuật toán của tôi được mô tả như Hình 2.

Thuật toán **PCA_HAC** có các tham số đầu vào là một tập ảnh các khuôn mặt cho trước. Hàm **detect_all_face()** nếu phát hiện được

khuôn mặt trong mỗi bức ảnh sẽ trả về giá trị **flag = true**. **boxFaces** là hình chữ nhật bao khuôn mặt được lưu trữ dưới dạng hai điểm **Top-left** và **Bottom-Right**. Hàm **crop()** cắt lấy ảnh con chứa khuôn mặt từ hình chữ nhật bao khuôn mặt đồng thời đưa các ảnh mặt thu được về cùng một kích cỡ. Hàm **push_back()** tạo nên tập khuôn mặt **LF**. Hàm **Hierarchical()** trả về kết quả các cụm khuôn mặt tương tự nhau **C**.

THUẬT TOÁN PCA_HAC

Input: LI – tập ảnh gồm n ảnh

Output: C – các cụm khuôn mặt

1. Khởi tạo

$nImages \leftarrow n$;

$totalFaces \leftarrow 0$;

2. Phát hiện khuôn mặt

For $i = 0$ **to** $nImages$ **do**

$Bool\ flag \leftarrow detect_all_faces(\&boxFaces, \&nFace, LI_i)$;

If ($flag$)

For $j = 0$ **to** $nFaces()$ **do**

$IF \leftarrow crop(boxFaces_j)$;

$LF.push_back(IF)$;

3. Trích rút đặc trưng

$S \leftarrow asRowMatrix(LF)$;

$D \leftarrow PCA(S)$;

4. Phân cụm khuôn mặt

$C \leftarrow Hierarchical(D)$;

5. Return C ;

Hình 2. Thuật toán PCA_HAC

Phần tiếp theo tôi trình bày về thuật toán PCA, sau đó là phần trích rút đặc trưng, cuối cùng là phần thuật toán phân cụm phân cấp.

Phân tích thành phần chính

Phân tích thành phần chính (Principal Component Analysis- PCA), còn được gọi là chuyển đổi Karhunen-Loève, là một biến đổi tuyến tính có thể nắm bắt sự thay đổi của dữ liệu đầu vào. PCA tìm ra một không gian mới theo hướng biến thiên mạnh nhất của một tập hợp các vector trong không gian cho trước giúp giảm số chiều của dữ liệu. Trong không gian mới ít chiều hơn, nhưng lại có khả năng biểu diễn dữ liệu tốt tương đương với không gian cũ đảm bảo được tối đa thông tin quan trọng nhất.

Giả sử ta cần xem xét tập dữ liệu

$$X = [x_1, x_2, \dots, x_n] \quad (1)$$

Trong đó n là số mẫu dữ liệu, x_i là mẫu dữ liệu thứ i có kích thước là d . Đầu tiên ta tính giá trị trung bình của X trên mỗi chiều

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

Trừ các giá trị trung bình ta thu được

$$\tilde{X} = x_i - \bar{x} \quad (3)$$

Tính ma trận hiệp phương sai (covariance) C :

$$C = \frac{1}{N} A A^T \quad (4)$$

Ma trận hiệp phương sai C có vector riêng $\Phi_{\tilde{X}}$ với giá trị riêng $\Lambda_{\tilde{X}}$.

$$C \Phi_{\tilde{X}} = \Phi_{\tilde{X}} \Lambda_{\tilde{X}} \quad (5)$$

$$\text{trong đó } \Lambda_{\tilde{X}} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_d \end{bmatrix} \quad (6)$$

là ma trận chéo của giá trị riêng tương ứng với vector riêng của

$$\Phi_{\tilde{X}} = [\Phi_1, \Phi_2, \dots, \Phi_d] \quad (7)$$

Các vector riêng tương ứng với giá trị riêng cao nhất đại diện cho các thành phần chính đầu tiên.

$$\Phi_{PCA} = [\Phi_1, \Phi_2, \dots, \Phi_k, k < d] \quad (8)$$

Trích rút đặc trưng

Mỗi ảnh đưa ở tập ảnh đầu vào có cùng kích thước $N \times N$ tương đương với N^2 vector đặc trưng khuôn mặt, như vậy các đặc trưng khuôn mặt này là rất lớn, để giảm số đặc trưng khuôn mặt ta áp dụng thuật toán PCA đã trình bày ở phần trên (chỉ còn K vector đặc trưng được giữ lại, $K \ll N^2$). Vector đặc trưng khuôn mặt được giữ lại chính là K vector riêng tương ứng với giá trị riêng lớn nhất. Có 2 cách để xác định K sao cho hiệu quả.

Cách đầu tiên ta sắp xếp theo thứ tự giảm dần các giá trị riêng đã tìm được. Thứ tự này vẫn đảm bảo được thứ tự của các vector đặc trưng tương ứng. Theo dõi sự biến thiên của dãy trên, khi không còn biến thiên (hoặc xấp xỉ bằng không) thì lúc đó ta đã chọn đủ K vector đặc trưng.

Cách thứ hai thì ta chọn K theo tiêu chuẩn sau:

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} > \text{Threshold (eg 0.9 or 0.95)} \quad (9)$$

Phân cụm khuôn mặt

Tôi biểu diễn tập dữ liệu đặc trưng của các khuôn mặt dưới dạng ma trận. Nếu ta có n khuôn mặt, mỗi khuôn mặt có p đặc trưng thì sẽ có một ma trận với n dòng, p cột.

Khoảng cách giữa hai khuôn mặt x, y hay độ đo phi tương tự giữa hai khuôn mặt được xác định bằng độ đo khoảng cách Euclidean

$$d(i, j) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (12)$$

HAC dựa theo đặc thù của thuật toán phân cụm đệ quy và coi mỗi đối tượng như một điểm dữ liệu trong không gian Euclide. Bằng cách đi lên từ lớp dưới cùng lên nút trên đầu, sơ đồ cây phân cấp cho chúng ta thấy các bước kết hợp đôi một từng nhóm. Việc tính toán độ tương tự giữa các cụm dựa vào cách tính khoảng cách trong không gian Euclide. Với r, s : hai cụm. i, j : hai đối tượng bất kỳ thuộc hai cụm ta có một số phương pháp tính khoảng cách giữa các cụm

Single link: Với 2 cụm, ta tính tất cả các khoảng cách giữa 2 phần tử bất kỳ thuộc 2 cụm đó và khoảng cách nhỏ nhất tìm được chính là khoảng cách giữa 2 cụm đó. Tại mỗi bước, 2 cụm gần nhau nhất sẽ được chọn để ghép lại với nhau

$$D(r, s) = \text{Min}(d(i, j)) \quad (13)$$

Complete linkage: Phương pháp này đối ngược với single link. Với 2 cụm, ta lấy khoảng cách lớn nhất giữa các phần tử làm khoảng cách giữa 2 cụm. Khoảng cách giữa các cụm được định nghĩa:

$$D(r, s) = \text{Max}(d(i, j)) \quad (14)$$

Average-linkage: đánh giá ghép cụm dựa vào toàn bộ độ tương tự giữa tất cả các đối tượng trong cụm vì vậy mà nó tránh được những thiếu sót của hai phương pháp single-link và complete-link – chỉ đánh giá được một phần các cụm.

$$D(r, s) = \text{Mean}(d(i, j)) \quad (15)$$

Hình dưới đây là thuật toán phân cụm phân cụm HAC.

```

Input: D- Ma trận đặc trưng cỡ NxN
Output: A - Các cụm khuôn mặt

For n ← 1 to N do
  For i ← 1 to N do
    C[n][i] ← SIM(Dn,i);
  I[n] ← 1;
  A ← [];
  Fork k ← 1 to N - 1 do
    <i, m> ← argAve { <i, m>: i ≠ m ∧ I[i] = 1 ∧ I[m] = 1 }
    C[i][m]
    A.APPEND(<i, m>)
  For j ← 1 to N do
    C[i][j] ← SIM(i, m, j);
    C[j][i] ← SIM(i, m, j);
  I[m] ← 0;
Return A;

```

Hình 3. Thuật toán HAC

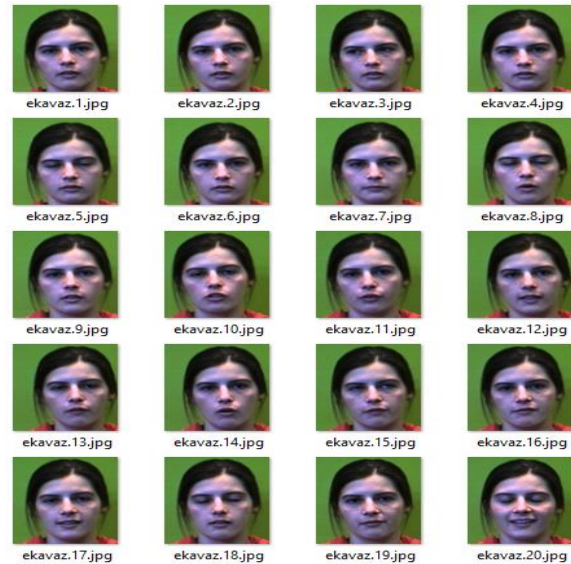
Trong thuật toán HAC trong Hình 3, đầu tiên chúng ta tính ma trận khoảng cách C cỡ NxN. Sau đó thực hiện N- 1 bước để sáp nhập các cụm hiện tại có độ tương tự nhỏ hơn một ngưỡng cho trước. Trong mỗi lần lặp, ta thu được hai cụm tương tự được sáp nhập và các hàng và cột của cụm i đã sáp nhập trong C được cập nhật. Các cụm được lưu trữ như một danh sách của việc sáp nhập trong A. I cho ta biết được trạng thái của cụm có thể được sáp nhập. Hàm SIM(i, j, m) tính độ tương tự của cụm j và cụm mới sau khi sáp nhập cụm i với cụm m, ở đây chúng ta dùng độ đo Average-linkage.

THỰC NGHIỆM

Xây dựng kho dữ liệu

Tôi thực nghiệm với tập ảnh khuôn mặt chỉ có một khuôn mặt với góc nhìn thẳng và điều kiện đủ ánh sáng. Kho dữ liệu được xây dựng từ 100 ảnh khuôn mặt khác nhau ở định dạng JPEG được trích rút từ cơ sở dữ liệu khuôn mặt UOF[8]. Cơ sở dữ liệu khuôn mặt UOF là

một cơ sở dữ liệu khuôn mặt của Dr Libor Spacek, trường University of Essex, UK. Ảnh trong cơ sở dữ liệu là ảnh màu 24 bit định dạng JPEG kích cỡ 180x200. Tập dữ liệu chứa một tập hợp các hình ảnh khuôn mặt gồm 395 cá nhân (cả nam và nữ), 20 ảnh cho mỗi cá nhân, tổng cộng có 7900 hình ảnh. Tất cả khuôn mặt chủ yếu được thực hiện bởi các sinh viên đại học năm đầu tiên độ tuổi từ 18 đến 20 tuổi và một số người lớn tuổi. Một số cá nhân đeo kính và có râu. Tập dữ liệu được lưu trữ trong bốn thư mục (faces94, faces95, faces96, grimace). Ví dụ về 20 hình ảnh của một cá nhân trong tập faces94 xem trong Hình 4.



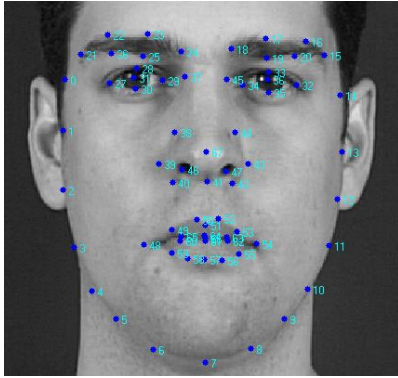
Hình 4. Tập 20 ảnh của một cá nhân trong faces94

Tôi xây dựng một kho cơ sở dữ liệu gồm 4 tập chứa 100 ảnh. Bằng cách thay đổi các hình ảnh khuôn mặt khác nhau trên tập dữ liệu, cho chúng ta thấy hiệu quả của hệ thống.

Kết quả thực nghiệm

Hệ thống được xây dựng bằng ngôn ngữ lập trình C/C++ với bộ thư viện hỗ trợ OpenCV trên Visual Studio 2012. Với kho dữ liệu là tập ảnh thử nghiệm đã nêu ở phần trước, tôi áp dụng thuật toán PCA để lấy các đặc trưng bằng cách giảm chiều các vector ban đầu của mỗi khuôn mặt. Bên cạnh đó tôi cũng trích rút đặc trưng dùng active shape model (ASM)

[14] để so sánh độ chính xác với phương pháp PCA. Độ chính xác là tổng các phân lớp đúng trong tổng số các ảnh. Phương pháp ASM so khớp mô hình hóa khuôn mặt cho phù hợp với hình dạng mong muốn ta thu được 68 điểm mốc.



Hình 5. Các điểm mốc của phương pháp ASM

Từ những điểm mốc đó ta trích rút được các đặc trưng nổi bật trên khuôn mặt đang xét

Tôi tiến hành 4 thực nghiệm đều gồm 100 hình ảnh khác nhau của 5 cá nhân bất kỳ thuộc bốn thư mục:

Thực nghiệm 1: faces94. Nền ảnh là màu xanh lá cây, chụp thẳng, độ nghiêng và vị trí của khuôn mặt biến đổi rất nhỏ.

Thực nghiệm 2: faces95. Nền ảnh là màu đỏ, chụp thẳng, đầu biến đổi lớn dần, vị trí của khuôn mặt thay đổi nhỏ.

Thực nghiệm 3: faces96. Nền ảnh phức tạp (dán áp phích), chụp thẳng, đầu biến đổi lớn dần, vị trí của khuôn mặt thay đổi, ánh sáng thay đổi đáng kể.

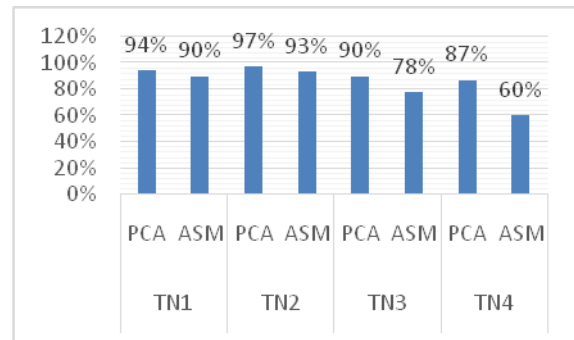
Bảng 1. Độ chính xác của các thực nghiệm

STT	Cỡ tập ảnh	SL cụm	Phương pháp	Tỷ lệ đúng
TN1	100	5	PCA	94%
			ASM	90%
TN2	100	5	PCA	97%
			ASM	93%
TN3	100	5	PCA	90%
			ASM	78%
TN4	100	5	PCA	87%
			ASM	70%

Thực nghiệm 4: grimace. Nền ảnh màu xám, chụp thẳng, độ nghiêng và vị trí của khuôn

mặt không thay đổi, chủ yếu thay đổi về biểu hiện cảm xúc của khuôn mặt.

Từ các thực nghiệm trên tôi thu được các kết quả như Bảng 1.



Hình 5. Đồ thị so sánh độ chính xác PCA với phương pháp ASM

Với những thực nghiệm khác nhau trên, tôi thấy phương pháp PCA cho những đặc trưng tốt hơn ASM khi phân cụm. Độ chính xác cả hai có kém đi với những ảnh khác nhau về điều kiện ánh sáng, điệu bộ (nghiêng đầu, xoay mặt,...) hay cảm xúc (cười to, há miệng...). Còn phương pháp ASM cho độ chính xác kém đi nhiều với những ảnh khác nhau về điệu bộ, cảm xúc và cử chỉ khuôn mặt.

KẾT LUẬN

Các nghiên cứu hướng ứng dụng mang tới nhiều cơ hội phát triển những hệ thống công nghệ thông tin trong thực tế, đáp ứng yêu cầu khai phá thông tin ngày càng lớn hiện nay.

Tôi đã phát triển PCA_HAC, một phương pháp phân cụm khuôn mặt người ứng dụng trong mạng xã hội. Phương pháp PCA_HAC có ưu điểm: phân tích được các thành phần quan trọng trên khuôn mặt thuận lợi cho việc trích rút đặc trưng, giảm chiều véc tơ đặc trưng.

Các kết quả thực nghiệm trên cơ sở dữ liệu gồm 100 ảnh trong nhiều tập ảnh khác nhau chỉ ra độ chính xác của phương pháp được đề xuất. Kết quả phân cụm khi dùng đặc trưng PCA tốt hơn khi dùng đặc trưng ASM.

TÀI LIỆU THAM KHẢO

1. A. K. Jain and R. C. Dubes (1988), *Algorithms for Clustering Data*, Prentice-Hall.

2. A. K. Jain, M.N. Murthy and P.J. Flynn (1999), *Data Clustering: A Review*, ACM Computing Reviews.
3. A. W. Fitzgibbon, A. Zisserman (2003), "Joint Manifold Distance: a new approach to appearance based clustering", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '03)* - Vol. 1, pp. 26-36
4. C. Otto, B. Klare, and A. Jain (2015), "An efficient approach for clustering face images," in *ICB*.
5. D. Jacobs and A. Biswas (2012), "Active image clustering: Seeking constraints from humans to complement algorithms", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2152-2159
6. D. Wang, C. Otto, and A. K. Jain (2016), "Face search at scale," *IEEE Trans. on PAMI*
7. G. Nizar ,C. Michel , and B. Nazha (2004), *Unsupervised and Semi –Supervised Clustering : A Brief survey*, INRIA Rocquencourt, B.P. 105
8. <http://cswwww.essex.ac.uk/mv/allfaces/>
9. N. Dalal and B. Triggs (2005), "Histograms of oriented gradients for human detection", *In CVPR*.
10. P. Antonopoulos, N. Nikolaidis and I. Pitas (2007), "Hierarchical face clustering using SIFT image features", in *Proc. IEEE Symposium on Computational Intelligence in Image and Signal Processing*, USA, pp. 325-329
11. S. Eickeler, F. Wallhoff, U. Iurgel, G. Rigoll (2001), "Contentbased Indexing of Images and Videos using Face Detection and Recognition Methods", *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*
- 12.S. Hussien Al-janabi (2005), *The use of soft computing to classify objects for Air photos and satellite image*, M.SC. thesis , university of Babylon.
13. Sanghamitra and M. Ujjwal (2002), "Genetic Clustering For Automatic Evolution of clusters and application of image classification", *journal of Recognition Society*, vol.35 pp.1197- 1208.
14. S. Milborrow (2007), *Locating Facial Features with Active Shape Models*. Master's thesis. University of Cape Town.

ABSTRACT

AN EFFECTIVE FACECLUSTERING METHOD FOR THE SOCIAL NETWORK

Nguyen Huu Quynh*

Electric Power University

In recent years, the amount of information on the social network is growing, as Facebook alone has hundreds of billions of images. Therefore, processing these data to assist users in knowledge discovery and data mining will be essential. In this paper we present a method for clustering faces in an existing set of facial images based features extraction by the PCA algorithm. We then used a hierarchical clustering algorithm (HAC) to faces clustering into distinct clusters. We also provided empirical results on a set of 100 images. to show the accuracy and speed of the method.

Keywords: *hierarchical clustering; principal component analysis; data mining; faces; clustering*

Ngày nhận bài: 12/02/2018; **Ngày phản biện:** 02/3/2018; **Ngày duyệt đăng:** 31/5/2018

* Email: quynhnh@epu.edu.vn