

NGHIÊN CỨU MÔ HÌNH MARKOV ẨN VÀ ỨNG DỤNG TRONG NHẬN DẠNG CHỮ VIẾT TAY

Nguyễn Thị Hà Phương

Trường Đại học Quảng Bình

Tóm tắt. Nhận dạng chữ viết tay là một đề tài rất quan trọng trong những ứng dụng khác nhau như tình báo, kỹ thuật robot... Các nghiên cứu về nhận dạng chữ viết tay đã được phát triển từ hơn nửa thập kỷ qua và đạt được nhiều thành quả thiết thực tại nhiều nước trên thế giới. Tuy nhiên, ở Việt Nam vẫn còn ít người nghiên cứu. Bài báo này trình bày kết quả nghiên cứu về mô hình Markov ẩn và ứng dụng mô hình đó trong nhận dạng chữ viết tay. Để xây dựng được mô hình Markov ẩn ta phải xác định hệ thống có bao nhiêu trạng thái, xác suất chuyển đổi giữa các trạng thái, từ đó tạo được hệ thống huấn luyện (học chữ viết tay). Chữ viết tay sẽ được trích rút các đặc trưng theo vec tơ định hướng trước khi đưa vào huấn luyện và nhận dạng. Dựa vào những giải pháp đó, bài báo xây dựng được chương trình thực nghiệm nhận dạng chữ viết tay với hiệu suất cao.

Từ khóa: nhận dạng; mô hình Markov ẩn; nhận dạng chữ viết tay; trích rút đặc trưng.

1. ĐẶT VẤN ĐỀ

Hiện nay, những văn bản được soạn thảo bằng chữ viết tay còn tồn tại rất nhiều và điều này làm cho việc nhập liệu bằng tay sẽ tốn nhiều công sức và thời gian, do đó việc nhận dạng để đọc dữ liệu tự động vào máy tính rất cần thiết. Nhận dạng chữ viết tay là vấn đề đang được quan tâm đối với các nhà nghiên cứu. Đã có một số phương pháp nhận dạng chữ viết tay như phương pháp noron [1], phương pháp tiếp cận cấu trúc, phương pháp mô hình Markov ẩn... trong đó, nhận dạng dựa vào mô hình Markov ẩn (Hidden Markov Model-HMM) đã và đang được nhiều nhà nghiên cứu lựa chọn vì hiệu suất nhận dạng cao [3] nhưng chủ yếu ứng dụng trong nhận dạng cử chỉ tay bằng video [2,3,4] (nhận dạng chữ viết tay bằng ảnh động). Vì vậy, bài báo đã nghiên cứu HMM và ứng dụng trong nhận dạng chữ viết tay bằng ảnh tĩnh.

2. MÔ HÌNH MARKOV ẨN

2.1. Mô hình Markov ẩn là gì?

Xét một hệ thống có N trạng thái khác nhau được ký hiệu $\{1, 2, \dots, N\}$. Trong khoảng không gian rời rạc, hệ thống sẽ thay đổi trạng thái (có thể quay lại các trạng thái giống nhau) tùy theo các tập xác suất liên kết với trạng thái. Chúng ta ký hiệu thời gian tức thời tương ứng với trạng thái thay đổi là $t = 1, 2, 3, \dots$ và ký hiệu trạng thái q hiện hành tại thời điểm t là q_t . Trong trường hợp chuỗi Markov bậc nhất, rời rạc theo thời gian, ta có công thức:

$$P[q_t = j | q_{t-1} = i, q_{t-2} = k, \dots] = P[q_t = j | q_{t-1} = i] \quad (1)$$

Như vậy, quá trình Markov là một quá trình ngẫu nhiên $q(t)$ mà trạng thái hiện tại tại thời điểm t chỉ phụ thuộc vào trạng thái tại thời điểm $(t-1)$ mà không phụ thuộc vào các

trạng thái trước thời điểm $(t-1)$. Hơn thế, chúng ta chỉ xét với về bên phải của quá trình là phụ thuộc vào thời gian, bằng cách ấy chỉ ra tập xác suất chuyển đổi trạng thái a_{ij} :

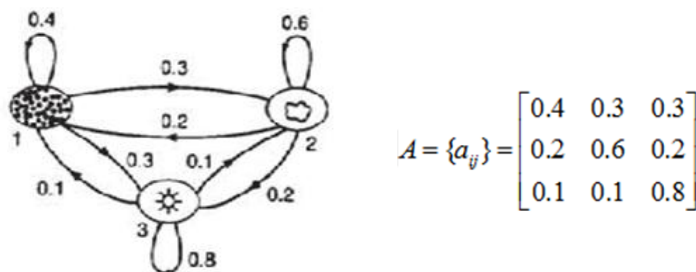
$$a_{ij} = P[q_t = j | q_{t-1} = i], 1 \leq i, j \leq N \quad (2)$$

với điều kiện: $a_{ij} \geq 0 \quad \forall i, j$ thì $\sum_{j=1}^N a_{ij} = 1 \quad \forall i$ (3)

từ đó chúng tuân theo các ràng buộc về xác suất.

Quá trình xác suất trên có thể được gọi là mô hình Markov quan sát vì đầu ra của quá trình là tập các trạng thái rời rạc theo thời gian, trong đó mỗi trạng thái tương ứng là một sự kiện quan sát. Xét một mô hình Markov đơn giản với 3 trạng thái dùng cho dự báo thời tiết như Hình 1[2]. Giả sử rằng 1 ngày (*buổi trưa*) thời tiết được quan sát như sau: Trạng thái 1: Mưa hoặc tuyết. Trạng thái 2: Mây. Trạng thái 3: Nắng.

Giả sử thời tiết của ngày t được biểu thị một trong ba trạng thái trên và ma trận A của xác suất chuyển đổi trạng thái là:



Hình 1. Mô hình Markov dự báo thời tiết.

Xem mô hình Markov là mô hình trong đó mỗi trạng thái tương ứng với một sự kiện quan sát xác định. Mở rộng khái niệm mô hình Markov bằng cách đưa thêm vào hàm xác suất quan sát ứng với trạng thái, đó là hàm mô hình kết quả (mô hình Markov ẩn) thêm hai quá trình xác suất mà nó không trực tiếp (ẩn) quan sát nhưng có thể quan sát tập các trạng thái sản xuất ra chuỗi quan sát. Để minh họa cho khái niệm HMM, chúng ta thực hiện thí nghiệm tung đồng xu.

2.2. Mô hình tung đồng xu

Giả sử có một diễn biến như sau: bạn ở trong một phòng với một tấm chắn, trong suốt quá trình tung đồng xu xảy ra bạn không thể nhìn thấy gì. Mặt bên kia của vật chắn là một người khác, người đó sẽ làm thí nghiệm tung đồng xu (sử dụng một hoặc nhiều đồng xu). Người đó sẽ không nói cho bạn biết đồng xu nào được tung mà chỉ nói cho bạn biết kết quả của mỗi lần tung đồng xu. Sau đó một chuỗi thí nghiệm tung đồng xu ẩn được tiến hành, với chuỗi quan sát bao gồm dãy mặt H và T . Chuỗi quan sát là :

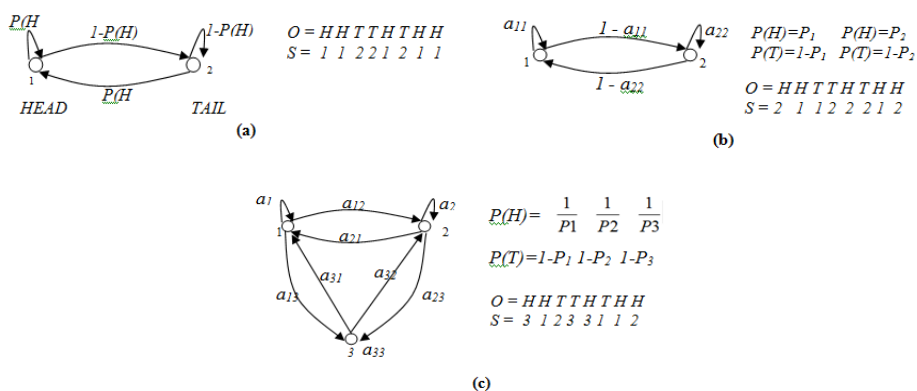
$$O = (O_1 O_2 \dots O_T) = (HTTT \dots H)$$

Chúng ta quyết định trạng thái trong mô hình tương ứng với cái gì và quyết định có bao nhiêu trạng thái sẽ có. Giả sử rằng chỉ có duy nhất một đồng xu được tung, ta có thể xây dựng vị trí với mô hình hai trạng thái mà mỗi trạng thái tương ứng với kết quả của lần tung

trước (Hình 2a), trong đó trạng thái tương ứng với đồng xu không đối xứng và tham số chưa biết là độ dịch chuyển của đồng xu.

HMM thứ hai giải thích chuỗi quan sát kết quả tung đồng xu (Hình 2b). Trong trường hợp này có hai trạng thái trong mô hình và mỗi trạng thái tương ứng với mỗi đồng xu khác nhau. Mỗi trạng thái được mô tả bởi một phân phối xác suất mặt H và mặt T , việc chuyển đổi trạng thái được mô tả bằng một ma trận chuyển đổi trạng thái.

HMM thứ ba giải thích chuỗi quan sát lần tung đồng xu (Hình 2c). Mô hình sử dụng ba đồng xu và lựa chọn giữa ba đồng xu dựa vào một số sự kiện xác suất. Mỗi mô hình đều có tham số chưa biết. Vì vậy, HMM được xem như mô hình thí nghiệm tung đồng xu.



Hình 2. Các mô hình Markov tung đồng xu.

Mô hình HMM được định nghĩa gồm :

○ N , là số trạng thái của mô hình. Trong chữ viết tay mỗi trạng thái tương ứng với một nét chữ. Các trạng thái riêng lẻ là $\{1, 2, \dots, N\}$ và trạng thái tại thời điểm t là q_t .

○ M , là số ký hiệu quan sát riêng lẻ trên một trạng thái. Trong chữ viết tay ký hiệu quan sát riêng lẻ là mã codeword xuất hiện tại mỗi vị trí. Ký hiệu $V = \{v_1, v_2, \dots, v_M\}$.

○ $A = \{a_{ij}\}$, là xác suất chuyển đổi giữa các trạng thái.

$$a_{ij} = P[q_{t+1} = j | q_t = i], \quad 1 \leq i, j \leq N \quad (4)$$

○ $B = \{b_j(k)\}$, là xác suất quan sát ký hiệu k của trạng thái j .

$$b_j(k) = P[o_t = v_k | q_t = j], \quad 1 \leq k \leq M \quad (5)$$

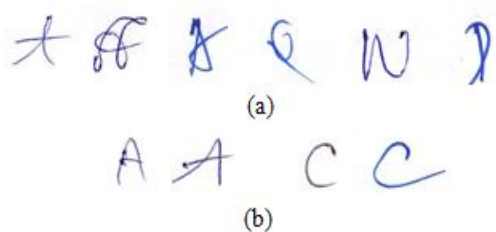
$$\pi = \{\pi_i\}, \text{ xác suất trạng thái ban đầu, } \pi_i = P[q_1 = i] \quad \forall 1 \leq i \leq N \quad (6)$$

Vậy cấu trúc chúng ta sử dụng là ký hiệu $\lambda = (A, B, \pi)$.

3. MỘT SỐ ĐIỀU KIỆN RÀNG BUỘC VỀ MẪU CHỮ

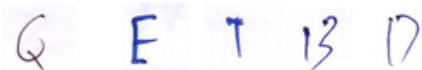
Vì chữ viết tay khá đa dạng, có thể gãy nét hay có thể là các tiếng nước ngoài như tiếng Lào, Trung Quốc... Nếu xét như vậy thì quá phức tạp không thể xét trong một thuật toán, do đó đề tài chỉ giới hạn ở các chữ cái hoa tiếng Anh.

- Chữ viết gồm các chữ cái từ A đến Z và viết trong một số kiểu font hạn chế. Không quá bay bướm, không quá nghiêng.



Hình 3. Mẫu chữ in hoa: a) Mẫu sai, b) Mẫu đúng.

- Chữ viết phải rõ ràng, không được quá cầu thả, mất nét. Đặc biệt cần có sự khác nhau giữa ký tự ‘C’ và ký tự ‘O’, ký tự ‘O’ và ký tự ‘Q’, ký tự ‘E’ và ‘F’...

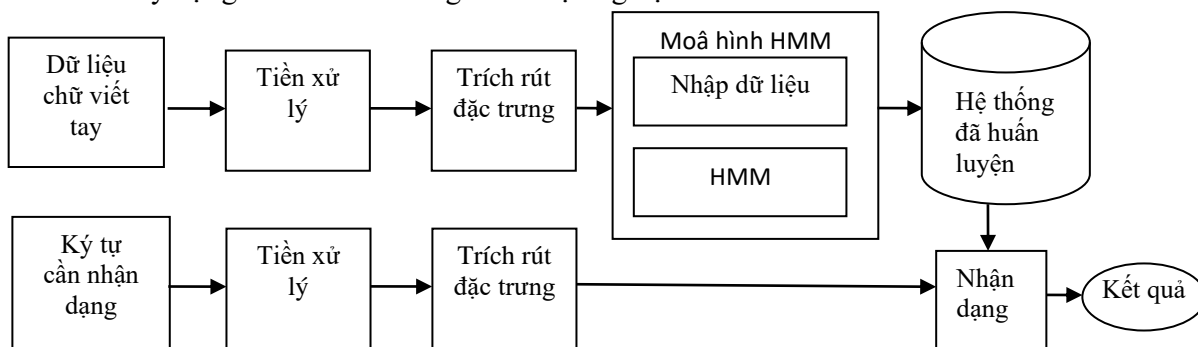


Hình 4. Chữ viết mất nét.

4. GIẢI PHÁP ĐỀ XUẤT

4.1. Sơ đồ tổng quát

Ta xây dựng sơ đồ của chương trình thực nghiệm như Hình 5.



Hình 5. Sơ đồ chương trình thực nghiệm.

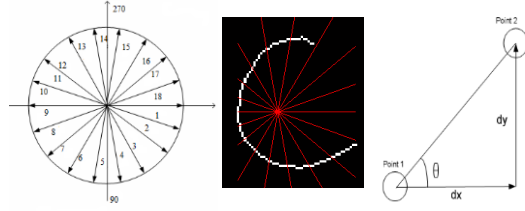
Bước nhập dữ liệu được lưu vào CSDL là chuỗi codeword của các mẫu ký tự và xây dựng HMM (xây dựng HMM của các ký tự).

4.2. Trích lọc đặc trưng

Quá trình trích lọc đặc trưng rất quan trọng. Có hai đặc trưng cơ bản: vị trí, hướng, trong đó hướng là đặc trưng cơ bản nhất. Vì vậy hướng dựa vào x_c , y_c (điểm trọng tâm) của ảnh ký tự và được xác định theo [5] bằng cách:

$$x_c = \frac{1}{n} \sum_{t=1}^n x_t \quad y_c = \frac{1}{n} \sum_{t=1}^n y_t \quad (7)$$

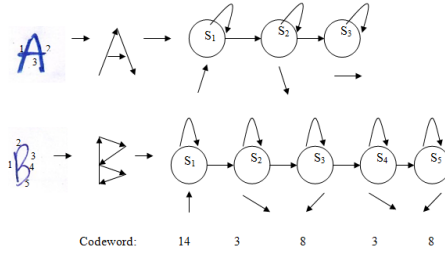
Trong đó n là độ dài chữ viết tay (số điểm tọa độ của chữ viết). Từ trọng tâm xác định các vectơ có hướng cách nhau 20° để tạo ra “codeword” từ 1 đến 18 (Hình 6).

**Hình 6. Vector định hướng.**

Các vector có hướng sẽ cắt ảnh ký tự tại các giao điểm. Khi đó ta xác định các đặc trưng của ký tự chính là xác định hướng của điểm $t+1$ so với điểm t bằng góc θ_t (Hình 6).

$$\theta_t = \arctan\left(\frac{y_{t+1} - y_t}{x_{t+1} - x_t}\right) \quad \backslash t = 1, 2 \dots T-1 \quad (8)$$

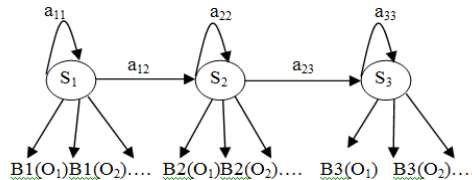
Sau khi trích lọc đặc trưng chúng ta xác định các trạng thái của ký tự, số trạng thái là một tham số quan trọng vì số trạng thái dư có thể phát sinh ra vấn đề tràn bộ nhớ nếu số mẫu huấn luyện không đủ để so sánh với tham số mô hình. Khi đó không có đủ số trạng thái, năng suất phân biệt của mô hình HMM bị giảm và sẽ có nhiều hơn một mẫu được làm mẫu trên một trạng thái.

**Hình 7. Mẫu ký tự với các trạng thái.**

4.3. Xây dựng mô hình HMM (Huấn luyện hệ thống)

○Nhập dữ liệu: lưu kết quả của quá trình trích rút đặc trưng vào CSDL. Kết quả của quá trình trích rút đặc trưng là chuỗi codeword sau khi đã chuyển từ chuỗi các góc θ .

○Xây dựng mô hình HMM mà cụ thể là xác định $\lambda_A, \lambda_B, \dots, \lambda_Z$.

**Hình 8. Mô hình Markov ẩn.**

Chuỗi quan sát là dãy codeword sau khi đã trích lọc đặc trưng và xác suất chuyển đổi trạng thái a_{ij} là ẩn, $B_i(O_j)$ là xác suất trạng thái i thấy ký hiệu O_j (mã codeword) hay là xác suất để codeword tiếp theo của O_{j-1} là O_j .

4.4. Nhận dạng ký tự

Để nhận dạng một ký tự ta tìm xác suất chuyển đổi điểm đặc trưng trên ký tự cần nhận dạng (xác suất chuyển đổi từ một điểm sang điểm kế tiếp lệch một góc θ_t là bao nhiêu %, cụ thể trong chương trình là xác suất chuyển đổi từ codeword này sang codeword khác, xác suất này đã được tính trong quá trình huấn luyện hệ thống). Như vậy việc tính xác suất cho đến điểm đặc trưng cuối cùng ta sẽ được **xác suất nhận dạng** của ký tự cần nhận dạng với mẫu trong hệ thống. Sau khi so sánh với tất cả các mẫu, mẫu có xác suất nhận dạng cao nhất chính là ký tự nhận dạng được.

Quá trình nhận dạng đó được thực hiện bằng thuật toán Viterbi [2]. Để tìm chuỗi trạng thái tốt nhất q cho ra chuỗi quan sát O chúng ta cần định nghĩa đại lượng:

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1 q_2 \dots q_{t-1}, q_t = i, o_1 o_2 \dots o_t / \lambda] \quad (9)$$

$\delta_t(i)$ là kết quả tốt nhất (xác suất cao nhất) dọc theo một quỹ đạo tại thời điểm t , tính toán cho quan sát t đầu tiên và kết thúc ở trạng thái i . Theo tính chất quy nạp ta có:

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] b_j(o_{t+1}) \quad (10)$$

Để thực sự lấy ra chuỗi trạng thái, thì cần phải giữ vết của đối số lớn nhất với mỗi t và j . Chúng ta thực hiện thông qua mảng $\psi_t(j)$. Hoàn thành thủ tục tìm chuỗi trạng thái tốt nhất:

- Khởi tạo: $\delta_1(i) = \pi_i b_i(o_1) \quad 1 \leq i \leq N \quad (11a)$

$$\psi_1(i) = 0 \quad (11b)$$

- Vòng lặp:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t) \quad 2 \leq t \leq T, 1 \leq j \leq N \quad (12a)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] \quad 2 \leq t \leq T, 1 \leq j \leq N \quad (12b)$$

- Kết thúc: $P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (13a)$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (13b)$$

- Đường dẫn (chuỗi trạng thái) tìm ngược:

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad \backslash \quad t = T-1, T-2 \dots 1 \quad (14)$$

Vậy chuỗi trạng thái là $q_1^* q_2^* \dots q_T^*$, $\delta_t(j)$, $\psi_t(j)$ lần lượt là giá trị lớn nhất, chỉ mục của S tại thời điểm t và P^* là tình trạng của một hàm tối ưu hóa gần nhất.

5. KẾT QUẢ NGHIÊN CỨU

Nghiên cứu thành công mô hình Markov ẩn và ứng dụng nhận dạng chữ viết tay trên ảnh ký tự đạt hiệu suất cao.

Kết quả thử nghiệm nhận dạng trên các mô hình Markov ẩn được xây dựng từ tập CSDL chữ viết tay, tập CSDL này được thu thập mẫu chữ từ nhiều người với các kiểu chữ khác nhau. Cụ thể các mẫu chữ được thu thập từ 50 người trên các đối tượng là văn phòng, công nhân (hoặc nông dân), giáo viên và học sinh. Mỗi người viết ít nhất 4 mẫu chữ/một chữ cái và thực hiện đối với 26 chữ cái. Như vậy đối với mỗi chữ sẽ có trên 200 mẫu chữ viết. Mẫu chữ để thực nghiệm nhận dạng được thu thập từ 20 người khác nhau với 5 mẫu chữ/một chữ cái/một người.

Bảng 1. Bảng kết quả của quá trình nhận dạng

TT	Ký tự	Mẫu Huấn luyện	Mẫu nhận dạng	Kết quả nhận dạng đúng	Hiệu suất
1	A	210	101	97	96%
2	C	257	95	93	97.8%
3	L	220	100	95	95%
Trung bình					96.3%

Chương trình thực nghiệm đạt hiệu suất 96,3% với số lượng mẫu chữ thực nghiệm không lớn. Như vậy, nếu dữ liệu huấn luyện và nhận dạng càng nhiều thì độ chính xác nhận dạng càng cao. Ngoài ra, trong quá trình thực nghiệm có sử dụng phương pháp để nâng cao chất lượng ảnh bằng cách lọc và khử nhiễu, do đó chương trình cũng đã nhận dạng được các ảnh bị mờ hoặc bị lệch tọa độ ít (trong quá trình scan).

Nguyên nhân dẫn đến kết quả nhận dạng sai:

- Chữ viết cẩu thả dẫn đến quá trình trích lọc đặc trưng không chính xác và cho ra kết quả nhận dạng sai.
- Một số chữ cái có đặc trưng gần giống nhau nên dễ nhận dạng nhầm sang ký tự khác.

6. KẾT LUẬN

Kết quả nghiên cứu các đặc trưng của ảnh, thuật toán nhận dạng, mô hình Markov ẩn, phương pháp nhận dạng bằng mô hình Markov ẩn và xây dựng chương trình thực nghiệm nhận dạng trên một số mẫu ký tự đạt hiệu suất cao. Trong các nghiên cứu sau này nếu tiếp tục phát triển với hệ thống mẫu chữ đa dạng hơn (thu thập nhiều mẫu chữ với nhiều đối tượng), nhận dạng những ký tự còn lại trong bảng chữ cái và thực hiện thêm một số chức năng tách chữ, nhận dạng đoạn văn bản viết tay thì sẽ có tính ứng dụng rất cao trong các bài toán thực tế về nhận dạng.

TÀI LIỆU THAM KHẢO

- [1] Huỳnh Hữu Hưng, Nguyễn Trọng Nguyên và các tác giả (2012), *Nhận dạng ngôn ngữ ký hiệu tiếng Việt sử dụng mạng Neuron nhân tạo*, No 5, Hội nghị khoa học ĐH Đà Nẵng lần V.
- [2] Lawrence Rabiner, Biing-Hwang Juang, *Fundamental of speech recognition*, Prentice-Hall International, Inc.

- [3] Mahmoud Elmezain, Ayoub Al-Hamadi, và các tác giả (2007), Gesture Recognition for Alphabets from Hand Motion Trajectory Using Hidden Markov Models, *Otto – von-Guericke-University Magdeburg, Germany*, IEEE Inte Symposium on signal Processing and Information Technology, p. 1192 - 1197.
- [4] Nianjun Liu, Brian C. Lovell, Peter J. Kootsookos, and Richard I.A. Davis Intelligent Real-Time Imaging and Sensing (IRIS) Group School of Information Technology and Electrical Engineering (2004), Model Structure Selection & Training Algorithms for an HMM Gesture Recognition System, *The University of Queensland, Brisbane, Australia 4072*, IEEE, p. 100 – 106.
- [5] Nianjun Liu, Brian C. Lovell, Peter J. Kootsookos (2003), Evaluation of HMM Training Algorithms for Letter Hand Gesture Recognition, *The University of Queensland, Brisbane, Australia 4072*, IEEE, p. 648 – 651.

RESEARCH HIDDEN MARKOV MODEL AND APPLICATION IN HANDWRITING RECOGNITION

Abstract. *Handwriting recognition plays an very important roles in various applications such as intelligence, robotics ... A great number of studies have been conducted over the half of decade and reached many practical achievements in many countries worldwide. However, the study on handwrriting recognition has remained open. This paper examinies the model of hidden Markov and its application to recognize the handwriting. In order to build a hidden Markov model we have to determine how many status the system have, the transition probabilities between states, thereby creating training system (learn handwriting). Handwriting must be extracted features from the orientation vector before being put into training and recognition. Based on this solution, the paper built experimental program handwriting recognition have been built with high performance.*