# A Study on Character Recognition using HMM and the Mason's Theorem

Sang-kyu Lee* and Jung-youn Hur**

* Department of Computer Engineering, Kyungnam University, Masan, KOREA
Tel : +82-55-249-2651 Fax : +82-55-248-2554 E-mail : 9941017@netian.com
** Department of Computer Engineering, Kyungnam University, Masan, KOREA
Tel : +82-55-249-2651 Fax : +82-55-248-2554 E-mail : hurj@zeus.kyungnam.ac.kr

**Abstract:**

In most of the character recognition systems, the method of template matching or statistical method using hidden Markov model is used to extract and recognize feature shapes. In this paper, we used modified chain-code which has 8-directions but 4-codes, and made the chain-code of hand-written character, after that, converted it into transition chain-code by applying to HMM(Hidden Markov Model). The transition chain code by HMM is analyzed as signal flow graph by Mason's theory which is generally used to calculate forward gain at automatic control system. If the specific forward gain and feedback gain is properly set, the forward gain of transition chain-code using Mason's theory can be distinguished depending on each object for recognition. This data of the gain is reorganized as tree structure, hence making it possible to distinguish different hand-written characters. With this method, 91% recognition rate was acquired.

**Key words :** Character recognition, HMM(Hidden Markov Model), Mason's theory

## 1. INTRODUCTION

As Industry develops, communication skills between men, machines, and between men and machines are frequently required. Especially, for data communication between men and machines, many improvements are made in Voice recognition, Voice synthesis, image recognition, and character recognition.

Recently, with improvements in computer signal-processing technique, mathematical theory like template matching, statistical modeling, parsing method is used. In the method using statistical modeling, likelihood is generally determined using Hidden Markov Model (HMM).

In this paper, HMM using chain-code in character recognition is applied to the transfer function of the signal flow graph, which is generally used in automatic control systems. Hence the transition state graph of the character's chain-code is drawn, and transfer function is calculated.

For every state transition gain, the gain of forward path and the feedback is properly set so that the input and output gain of the character can be different. The resulting input and output gain is reorganized as a tree structure, and by tree structure searching character recognition is done. This method is conducted on 10 numbers ranging from 0 to 9.

## 2. HMM AND MASON'S THEOREM

### 2.1 Hidden Markov Model

Hidden Markov Model was researched by Baum etc. in the late 1960s, and efficient algorithm began to be found. Baker first found out that this model could be applied to voice-recognition. The HMM can be defined as Markov chain, probability Markov chain that can be defined as probability distribution of observation symbol at each state, or statistic function of the Markov chain. Markov Model (or Markov chain) is defined as finite number of state (or node) and the set of transition having direction between states (nodes). At each state, one observation symbol is allocated, transition is probably defined, and node connected by transition forms one network. Probability variable of the state $q_t$ at each time $t=0, 1\cdots$ can be defined as a conditional statistics of the present state and the previous state. With first-order Markov assumption, It can be rewritten as a following [3,5,7]

$$P(q_{t+1} = S_j | q_0 = S_a, q_1 = S_b, \ldots, q_t = S_i)$$
$$= P(q_{t+1} = q_j | q_t = S_i)$$

In other words, the state transition has no relation with the past sate, and only indicates the function just previous the transition. This value is called transition probability, assuming that it is independent of time is represented as $a_{ij}$. Probability distribution of the transition in the entire network is represented as matrix $A=\{a_{ij}\}$. Probability of Markov Model being the state $S_j$ at a certain time $t$ can be represented recursively.

$$P_t(j) = \sum_i P_{t-1}(i) a_{ij} \qquad 1 \le j \le N$$

Extending this, by probably presenting observation symbol at each state, modeling ability can be improved. This probability variable is represented as probability function $b_j(k)$ at the state, then probability $P_t(j)$ at the state $S_j$ is redefined as following.

$$P_t(j) = \sum_i P_{t-1}(i) a_{ij} b_k(k) \quad 1 \le j \le M \,, \, 1 \le j \le N$$

This model has a complex structure determined by two probability variables. This is called Hidden Markov Model. General characteristics of HMM follow.

· the number of states $N$ : represents the number of every state of the model, every state is represented as

$$S = \{S_1, S_2, \cdots, S_N\}.$$

· the number of observation symbols $M$ : represents total number of observation symbol at each state. Observation symbols are represented as $V = \{v_1, v_2, \cdots, v_M\}$. · probability distribution of the state transition $A=\{a_{ij}\}$ : represents vector parameter of the every Markov chain's state transition probability. It's value is defined as

$$a_{ij} = P(q_{t+1} = S_j \mid q+t = S_i) \qquad 1 \le i, j \le N$$

$a_{ij}$ indicates the probability of transition from the state $S_i$ to the state $S_i$ at a certain time. This probability is assumed to be independent of time, and satisfies following probability condition.

$$\sum_j a_{ij} = 1$$

· probability distribution of the observation symbol $B = \{b_i(k)\}$ : represents the probability of observing a certain target symbol, and it's values are defined as following.

$$b_i(k) = P(v_k \mid S_i) \quad 1 \le k \le M \qquad 1 \le i \le N$$

$b_i(k)$ is the probability of observing the target symbol $v_k$ at the state $S_i$, and satisfies following probability condition.

$$\sum_k b_i(k) = 1 \qquad 1 \le i \le N$$

Also, it is assumed to be independent of time, and the probability of observation symbol at a certain time is only dependent upon the state located at the time(output-independence assumption).
The probability distribution of the initial state $\prod = \{\pi_i\}$ : represents the probability distribution of the initial state*(t=0)*,

$$\pi_i = P(q_0 = S_i), \quad 1 \le i \le N$$

and defined as above. These parameters satisfy the following probability condition.

$$\sum_i \pi_i = 1$$

HMM is generally described by parameter $\lambda = (A, B, \prod)$ which describes dynamic characteristics of the model [5,7,8].

### 2.2 Mason's theorem

Mason's theorem is the characteristics of the complex system represented as a flow graph using transfer function at the flow of signal. By using signal flow graph, complex system becomes easy to analyze. Especially, noise interfering in the course of system becomes easy to distinguish, and also by representing flow graph, this can be used in circuits of analog calculator.


Suppose there is a simple equation $X_i=A_{ij}*X_j$. $X_i$ and $X_j$ can be any time conjugate value or any other value, and $A_{ij}$ represents a transfer function. The flow graph shows

With block diagram, flow graph can be easily acquired. However for the complex system, to acquire the gain of the system or the transfer function, Mason's theorem like equation (1) is generally used[1,2].

$$M = \frac{x_{out}}{x_{in}} = \frac{\sum_k M_k \Delta_k}{\Delta} \quad - - - (1)$$

$M_k$ : $k^{th}$ forward path gain of $k^{th}$ path

$\Delta : 1 - \sum_m P_{m1} + \sum_m P_{m2} - \sum_m P_{m3} + \ldots\ldots$ *=1- (sum of every each gain) + (sum of gain point of 2 non-touching loop's every possible combination) - (sum of gain point of 3 non touching loop's every possible combination) + ……*

$\Delta_k$ : $k^{th}\Delta$ *which is non touching with* $k^{th}$ *forward path*

In this equation, non-touching loop indicates the loop not sharing node, $\Delta$is called the determinant of signal flow graph or characteristic determinant[1].

## 3. EXPRESSION OF HMM BY TRANSFER FUNCTION

Fig. 1 shows chain code and example of application upon number '5'. Depending upon the start point and direction of motion, most 8-direction chain code can create upside down code. In this paper, to eliminate this obscurity 4-direction modified chain code is used shown in the Fig.1.

The application upon number '5' on the right side of Fig. 1 starts at the top and have 000022200112233000 directions code. To apply to the HMM, transfer chain code is made to be 0201230.

Table 1 shows transfer chain codes from the number '0' to '9', Fig. 2 shows transfer flow graph of various hand-written numbers. At Fig.2 the value of transfer function from one state to another is set as 'a' and for feedback 'b'

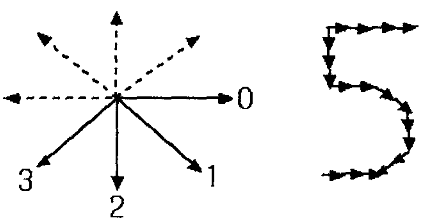


**Fig. 1.** Chain code and example of application upon number '5'

**Table 1.** Chain code of the hand-written characters and transfer function

| Number | Chain Code | Transfer Function |
|---|---|---|
| 0 | 01230123 | $\frac{8a^3}{1-24a^3b}$ |
| 1 | 2 | $\frac{1}{1-b}$ |
| 2 | 0123203 | $\frac{3a+a^3}{1-2a^2b-ab}$ |
| 2 | 012303 | $\frac{3a+a^3}{1-3a^3b}$ |
| 2 | 00123 | $\frac{a^3}{1-b}$ |
| 2 | 001232 | $\frac{a^3}{1-b-ab+ab^2}$ |
| 3 | 012312303 | $\frac{3a+4a^2}{1-8a^2b-12a^3b}$ |
| 4 | 032 | $\frac{1+3a}{1-b}$ |
| 4 | 233 | $\frac{a}{1-b}$ |

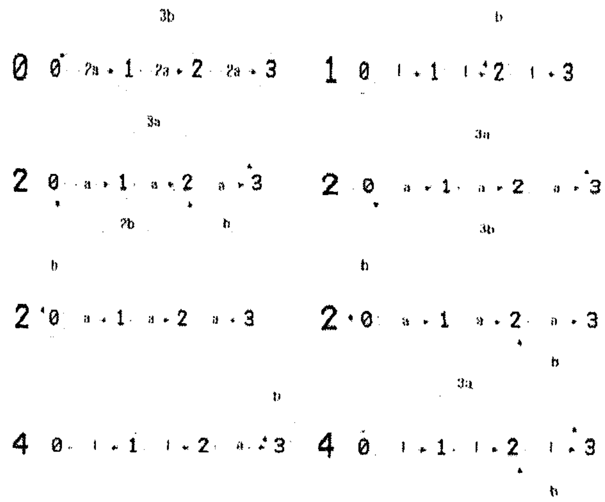

**Fig. 2.** Signal flow graph of written number by HMM

At Fig. 2, because the transferred gain 'a' or 'b' are used only to distinguish each hand-written characters, 'a' and 'b' should be properly set so that the gains between inputs and outputs are distinguishable.

## 4. SIMULATION AND RESULTS

At table 2, to find most distinguishable case for various 14 hand-written characters, the forward path gain 'a' is replaced from -5 to +5 and 'b' from 1.3 +a/2.4. The gain and covariance (*cov*) and minimum difference (*diff*) of each row is calculated. In the most ideal case, as *diff* and *cov* gets larger, distinctiveness of the characters improves.
At table 2, when a=-3, *cov* and *diff* become largest. If *diff* equals zero, the gain of the characters for recognition is the same, hence there are more than 2 characters unable to distinguish.

**Table 2.** Gain of 14 character using constant variable a and b=1.3+a/2.4

Fig. 3 shows decision tree using the gains of 14 characters when a=-3 at table 2. According to the gain of a certain character, after binary searching from root to leaf, the index matching the final character pattern can be acquired.

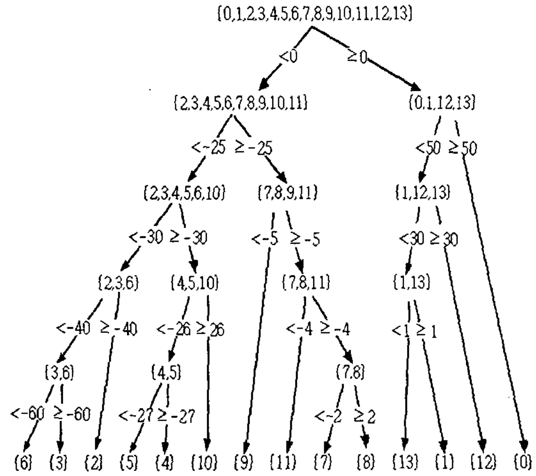

**Fig. 3.** Binary decision tree from a = -3 column of table 2

| a<br>Index | -5 | -4 | -3 | -2 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|
| 0 | 0.38 | 0.76 | **118.85** | -0.77 | -0.15 | -0.13 | -0.11 | -0.10 |
| 1 | 0.53 | 0.69 | **1.00** | 1.76 | -0.86 | -0.62 | -0.49 | -0.40 |
| 2 | -3.47 | -5.72 | **-33.80** | 8.85 | -0.68 | -0.67 | -0.70 | -0.74 |
| 3 | 0.43 | 0.91 | **-55.57** | -1.24 | -0.27 | -0.17 | -0.13 | -0.11 |
| 4 | -66.71 | -44.47 | **-26.88** | -14.05 | -6.84 | -16.83 | -31.39 | -50.53 |
| 5 | 19.80 | 58.78 | **-27.24** | -7.55 | 2.05 | 2.47 | 2.81 | 3.09 |
| 6 | -0.07 | -0.19 | **-282.27** | 0.35 | -0.08 | -0.04 | -0.03 | -0.02 |
| 7 | -2.13 | -2.08 | **-1.99** | -1.76 | -2.57 | -2.49 | -2.45 | -2.43 |
| 8 | -2.67 | -2.78 | **-2.99** | -3.51 | -1.71 | -1.87 | -1.96 | -2.02 |
| 9 | -1.68 | -2.13 | **-8.35** | 0.00 | -0.98 | -0.98 | -1.00 | -1.01 |
| 10 | -3.98 | -6.62 | **-25.12** | 9.99 | 7.69 | 3.84 | 2.71 | 2.14 |
| 11 | -4.80 | -4.86 | **-4.98** | -5.27 | -4.28 | -4.36 | -4.41 | -4.45 |
| 12 | 0.21 | 0.41 | **48.87** | -0.40 | -0.13 | -0.09 | -0.08 | -0.06 |
| 13 | -0.49 | -0.59 | **0.00** | -0.49 | -1.44 | -1.82 | -2.62 | -4.92 |
| cov | 67.9 | 74.5 | **310.1** | 21.6 | 11.6 | 17.3 | 30.4 | 48.5 |
| diff | 0.05 | 0.04 | **0.36** | 0.09 | 0.02 | 0.03 | 0.02 | 0.01 |

## 5. CONCLUSION

In this paper, Using modified chain code with 8-directions but 4-codes, chain code is made. To apply to HMM model, it is converted into transfer chain code, and then transfer chain code made by HMM is analyzed as a signal flow graph using Mason's theory that is generally used to calculate forward gain at the automatic control.

The input and output gain of HMM using Mason's theorem is organized as binary tree structure, and by tree searching pattern matching is done.

After applying to 100 hand-written numbers, 91% recognition rate was acquired, and those with too distorted or having too many jutted strokes was mismatched.

According to the analysis, the forward gain of transfer chain code using Mason's theorem has very high distinctiveness, the recognition rate will improve, if the kinds of reference test are diversified, and these methods can be applied to other character recognitions.


### Acknowledgements

This work was supported by Kyungnam University Research Fund.


### References


[1] Richard C. Dorf, "modern Control Systems", Addison Wesley, (1967)

[2] Distno III, "Feedback and Control Systems", Schaum Pub. Co.,(1971)

[3] B. K. Sin and J. H. Kim, "A statistical approach with HMMs for on-line cursive Hangul(Korean script)recognition", in *proceedings of the 2nd International Conference on Document Analysis and Recognition,*(Tsukuba Science City, Japan), pp.147-150, Oct. (1993)

[4] T. J. Sung and S. Y. Pang, "On-line korean character recognition using learning rules for character composition,", *Journal of Korean Information Science Society,* vol. 20, pp.305-316, Mar. (1993) in Korean.

[5] N. Merhav and Y. Ephraim. Maximum likelihood hidden Markov modeling using a dominant sequence of states. *IEEE Trans. on Signal Processing,* 39(9):2111-2115. Sep. (1991)

[6] J.-Y. Ha, S.-C. Oh, and Jin H. Kim. Unconstrained handwritten word recognition with interconnected hidden Markov models. In Third Int. *Workshop on Frontiers in Handwriting Recognition,* pp.455-460, Buffalo, New York, USA, May (1993)

[7] Michel Gilloux, Manuel Leroux, and Jean-Michel Bertille, "Strategies for Handwritten Words Recognition Using Hidden Markov Models", *Proc. of 2nd Int. Cong. on Document Analysis and Recognition, Tsukuba Science City,* Japan, 299-304, (1993)

[8] H. Bunke, M. Roth, and E. G. Schukat-Talamazzini, "Off-Line Cursive Handwriting Recognition Using Hidden Markov Model", *Pattern Recognition,* Vol.28, No.9, pp.1399-1413, (1995).