Chapter 5

CHAPTER 5

HANDWRITTEN CHARACTER MODELING USING STATE-SPACE PARAMETERS BASED ON GRAY-SCALE IMAGES

5.1 Introduction

This chapter presents a novel method for modeling Malayalam handwritten characters directly from gray-scale images without the usual step of binarization. For this purpose the reconstructed state-space for each character pattern is generated based on the information obtained from the pixel intensity distribution of both the foreground and the background of the gray-scale image. Then the trajectory matrix of embedding dimension $d=2$ is formed from this reconstructed state-space of each character image. After that the scatter plot of the row vector of the trajectory matrix named State-Space Map (SSM) is constructed for each character image with one, four and eight directional space variations (space delay). The State-Space Point Distribution (SSPD) parameters are extracted for each character pattern from the state-space map. The extracted parameters are found to be promising and can be effectively used for the recognition purpose.

5.2 Gray-scale Based Features for Modeling Handwritten Characters

In the present study we propose a promising approach to perform feature extraction from gray-scale images of the handwritten characters. Direct feature extraction from gray-scale image is not a new idea [Wang.L
and Pavlidis.T, 1993]. As early as in 1975, Peuker and Douglar [Peuker.T.K and Donglar.D.H, 1975] had proposed methods for the detection of topographic structure in a gray-scale image. Figure 5.1 shows the block diagram of the direct feature extraction method from the gray-scale image. A major challenge in gray-scale image based feature extraction is to identify the candidate character locations in the image [Kahan.S et al., 1987]. Here we introduced a new method of feature extraction based on the state-space representation by trying all possible locations in a gray-scale image.

![Block Diagram of Direct Feature Extraction](image)

**Figure 5.1 Direct feature extraction from gray-scale image**

5.3 The Reconstructed State-space for Handwritten Character Images

In the case of purely deterministic systems, once it’s state is fixed, then the states at all future times can be determined as well. Thus by all accounts it is significant to establish a vector space called State-Space or Phase-Space for the system such that, specifying a point in this space specifies the state of the system. This will definitely help us to study the dynamics of the system by studying the dynamics of the corresponding state-space points.
The concept of the state of a system is powerful for non-deterministic systems also. *Takens theorem* states that under certain assumptions, state-space of a dynamical system can be reconstructed through the use of time delayed (space varying) versions of the original scalar measurements [Takens.F, 1981]. This new state-space is commonly referred in the literature as a reconstructed state-space. It is also proven that the reconstructed state-spaces are topologically equivalent to the original phase-space of the dynamical system, as if all the state variables of that system would have been measured simultaneously [Ott.E, 1993], [Baker.G.L and Gollub.J, 1996]. So a reconstructed state-space can be treated as a powerful signal-processing domain, especially when the dynamical system of interest is nonlinear or even chaotic [Kantz.H and Schreiber.T, 1998], [Broomhead.D.S and King.G, 1986].

A reconstructed state-space for a dynamical system can be produced from a measured state variable, $I_n$, $n = 1, 2, 3 \ldots N$, via the method of delays by creating vectors given by

$$I_n = [i_n \ i_{n+\tau} \ i_{n+2\tau} \ \ldots \ i_{n+(d-1)\tau}]$$

where $d$ is the embedding dimension and $\tau$ is the chosen time or space delay value.

The row vector, $I_n$, defines the position of a single point in the reconstructed state-space. The row vectors then can be compiled into a matrix
(called a trajectory matrix) to completely define the dynamics of the system and create a reconstructed state-space.

\[
I = \begin{bmatrix}
i_1 & i_{1+\tau} & i_{1+2\tau} & \ldots & i_{1+(d-1)\tau} \\
i_2 & i_{2+\tau} & i_{2+2\tau} & \ldots & i_{2+(d-1)\tau} \\
i_3 & i_{3+\tau} & i_{3+2\tau} & \ldots & i_{3+(d-1)\tau} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
i_N & i_{N+\tau} & i_{N+2\tau} & \ldots & i_{N+(d-1)\tau}
\end{bmatrix}
\]

A handwritten character image, denoted by a two-dimensional function \( f(x, y) \), can be treated as a dynamical system with the original scalar measurements including the pixel intensity values or amplitude of \( f \) and the spatial co-ordinates \((x, y)\). Based on the above theory, this study investigates a method to model a reconstructed state-space for handwritten character images, through the use of space varying versions of the original scalar measurements. Here trajectory matrices \( I_1 \) with embedding dimension \( d = 2 \) and space delay or space variation \( \tau = 1 \) and \( I_2 \) with embedding dimension \( d = 3 \) and space delay \( \tau = 1 \) are constructed by considering the pixel intensity values \( i \) taken from all the spatial co-ordinate points of the image. The matrices \( I_1 \) and \( I_2 \) thus obtained are given below.

\[
I_1 = \begin{bmatrix}
i_1 & i_{1+\tau} \\
i_2 & i_{2+\tau} \\
i_3 & i_{3+\tau} \\
\vdots & \vdots \\
i_N & i_{N+\tau}
\end{bmatrix}
\]
A visual representation of the system dynamics is evident from figure 5.2(a) and Figure 5.2(b) obtained by plotting the row vectors of the above given trajectory matrix $I_1$ and $I_2$, constructed for the Malayalam handwritten character ə [-ah].

The state-space reconstruction techniques are not specific to any particular production model of the underlying system with a finite dimension. Theoretically reconstructed state-spaces (in the case of an image it can be treated as space varying vectors of a two dimensional signal) are capable of capturing the full dynamics of the system including the non-linear information.
Figure 5.2 (a) Plot of the two dimensional trajectory matrix $I_1$ of Malayalam handwritten character മ[-ah]

Figure 5.2 (b) Plot of the three dimensional trajectory matrix $I_2$ of Malayalam handwritten character മ[-ah]

88
In the present study the reconstructed state-space maps for segmented Malayalam handwritten characters are generated for one, four and eight directional space variations by considering one neighbour, 4-neighbours and 8-neighbours of each pixel with space delay $\tau=1$ as discussed in the following sections.

5.3.1 State-Space Map (SSM) of Handwritten Characters with One Directional Space Variation

In this section state-space map of handwritten characters are constructed with one directional space variation (space delay). At the outset, the total 2772 pixel intensity values obtained from the size normalized handwritten character image of size 66x42 pixels, ranging from 0 to 255 gray level values, are treated as a state vector denoted by $[i_1, i_2, \ldots, i_{2772}]$. For each image sample, a trajectory matrix is formed by taking intensity values of each pixel $p(x, y)$ and its neighbor $p(x+1, y)$ denoted by $i_n$ and $i_{n+1}$ respectively. In all the cases the embedding dimension 'd' is two and the space delay $\tau$ is one. Now a scatter plot named State-Space Map (SSM) is generated by plotting the row vectors of the above constructed trajectory matrix (ie by plotting $i_n$ versus $i_{n+1}$, where $i^1$ indicates the one directional space variation). Figure 5.3(a-f) shows the state-space maps of the selected Malayalam characters [ə], [ga], [ta], [thha], [pa] and [rha] constructed for one directional space variation.
Fig. 5.3(a) SSM of Malayalam handwritten character ൕ േ [-ah]

Fig. 5.3(b) SSM of Malayalam handwritten character ൔ[ga]
Fig. 5.3(c) SSM of Malayalam handwritten character s[ta]

Fig. 5.3(d) SSM of Malayalam handwritten character th[tha]
Figure 5.3(a-f): State-space maps for Malayalam handwritten characters ൽ[-ah], ൽ[ga], ൽ[ta], ൽ[thha], ൽ[pa] and ൽ[rha] with one directional space variation.
Manually inspecting the SSMs of the forty four Malayalam characters, it is observed that more than fifty percent of the characters are distinguishable from each other. Here only one directional space variation is incorporated to construct state-space maps by considering the information regarding only one neighbour of each pixel in the image. In the case of gray-scale images, additional information of its pixel intensity distributions can be achieved by incorporating higher directional space variations. So it is good enough to construct SSMs by incorporating 4-directional and subsequently 8-directional space variations by considering 4-neighbours and 8-neighbours of each pixel in the character image. Following sections describe the incorporation of four and eight directional space variation by considering information about four and eight neighbours of each pixel in the gray-scale images for the construction of SSMs for the handwritten characters.

5.3.2 State-Space Map (SSM) of Handwritten Characters with Four Directional Space Variations

This section deals with the construction of state-space maps for size normalized (66x42 pixel size) handwritten character images with four directional space variations. Here for each pixel \( p(x, y) \) in the image the set of its two horizontal and two vertical neighbours, \( N4p \), whose co-ordinates are given by \( (x-1, y), (x+1, y) \) and \( (x, y-1), (x, y+1) \) respectively, are taken into account. Since each of these neighboring pixels are of unit distance from \( p(x, y) \), the space variation or space delay \( \tau \) is one. The intensity value of
pixel \( p(x, y) \) is taken as \( i_n \) and that of its four neighbours are taken as \( i^1_{n+1}, i^2_{n+1}, i^3_{n+1}, i^4_{n+1} \). Here \( i^1, i^2, i^3 \) and \( i^4 \) indicate four directional space variations. Finally the state-space map with four directional space variation is generated by plotting \( i_n \) versus \([i^1_{n+1}, i^2_{n+1}, i^3_{n+1}, i^4_{n+1}]\) i.e., plotting \( i_n \) versus \( i^k_{n+1} \) where \( k = 1,2,3,4 \). This plot is a scatter plot of the trajectory of the system with embedding dimension \( d = 2 \) and the space delay \( \tau = 1 \), varies in four directions. Figure 5.4(a-f) shows the state-space maps of the selected Malayalam handwritten characters \( \text{m[ah]}, \text{n[ga]}, \text{s[ta]}, \text{D[tha]}, \text{n[pa]} \) and \( \text{n[rha]} \) constructed for four directional space variations. From these figures it is evident that the state-space map for different characters are dissimilar in nature. It is also noted that the above obtained SSMs with four directional space variations are more informative than the state-space map with one directional space variation that are obtained in section 5.3.1. Manually inspecting the SSMs with four directional space variations of the forty four Malayalam characters it is observed that more than two third of the characters are distinguishable from each other. The modified version of the state-space map with eight directional space variations by comprising the information about eight neighbours of each pixel in the image is constructed as follows.
Fig. 5.4(a) SSM of Malayalam handwritten character ऊँ [-ah]

Fig. 5.4(b) SSM of Malayalam handwritten character ऊँ[ga]
Fig. 5.4(c) SSM of Malayalam handwritten character sipa

Fig. 5.4(d) SSM of Malayalam handwritten character thha
Figure 5.4(a-f): State-space maps for Malayalam handwritten characters $\text{ae} [-\text{ah}]$, $\omega [\text{ga}]$, $\delta [\text{ta}]$, $\omega [\text{thha}]$, $\alpha [\text{pa}]$ and $\alpha [\text{rha}]$ with four directional space variations
5.3.3 State-Space Map (SSM) of Handwritten Characters with Eight Directional Space Variations

This section describes the proposed method to generate the state-space map of the handwritten character images with eight directional space variations by considering the eight neighbours, $N8p$, of each pixel $p(x, y)$ in the character image. In addition to four horizontal and vertical neighbours, $N4p$, mentioned in section 5.3.2, the four diagonal neighbours, $N4D$, of $p(x, y)$ with the co-ordinates $(x-1, y-1)$, $(x-1, y+1)$, $(x+1, y-1)$ and $(x+1, y+1)$ are also considered here for constructing state-space maps. Since each of these neighbouring pixels are of unit distance from $p(x, y)$, the space variation or space delay $\tau$ is one. The intensity value of pixel $p(x, y)$ is taken as $i_n$ and that of its eight neighbours are taken as $i_{n+1}^k$ where $k = 1, 2, \ldots, 8$. Finally the state-space map with eight directional space variation is generated by plotting $i_n$ versus $i_{n+1}^k$ where $k = 1, 2, \ldots, 8$. This plot is a scatter plot of the trajectory of the system with embedding dimension $d = 2$ and the space delay $\tau = 1$, varies in eight directions. Figure 5.5 (a-f) shows the state-space maps of the selected Malayalam characters മ[-ah], ഗ[ga], സ[ta], മ[thha], പ[pa] and ര[rha] constructed for eight directional space variations. From this figure it is evident that the SSM for different characters are dissimilar in nature.
Fig. 5.5(a) SSM of Malayalam handwritten character Ω [-ah]

Fig. 5.5(b) SSM of Malayalam handwritten character ω [ga]
Fig. 5.5(c) SSM of Malayalam handwritten characters $s[ta]$

Fig. 5.5(d) SSM of Malayalam handwritten character $\theta[th\text{ha}]$
Figure 5.5(a-f): State-space maps for Malayalam handwritten characters .Delay [-ah],  [ga],  [ta],  [thha],  [pa] and  [rha] with eight directional space variations.
Manually inspecting this state-space maps with eight directional space variations of the forty four Malayalam characters, it is observed that more than three fourth of the characters are distinguishable from each other. The state-space maps with eight directional space variations generated for different samples of the Malayalam handwritten character "$\text{m-ah}$" are shown in figure 5.6. From these figures it is evident that the state-space maps generated for different samples of the same character are similar in nature.

The above results indicate that, state-space map can be effectively utilized for representing each Malayalam handwritten character since it is similar for different samples of the same character and differ from character to character. Moreover the state-space maps with eight directional space variations are found to be well-informed compared to the state-space map with one and four directional space variations and hence it is used further in the feature extraction stage. The following section describes the proposed method of parameter extraction from the state-space maps with eight directional space variations for Malayalam handwritten characters.
Figure 5.6: State-space maps for different samples of Malayalam handwritten characters ோ [ ah ] with eight directional space variations
5.4 State-Space Point Distribution (SSPD) Features from the State-space Map

In handwritten character recognition problem, selection of feature extraction method is certainly the most important factor in achieving high recognition performance. This section describes a scheme for extracting features from the gray-scale images of the handwritten characters based on their state-space map with eight directional space variations. For this purpose the state-space map of the character image is divided into grids with 16x16 locations (boxes) as shown in figure 5.7. The box defined by the co-ordinates (0, 0), (1, 1) is taken as box one and box just right side to it is taken as box two and it is extended towards X direction with the last box being (15, 0), (16, 1) is taken as box sixteen. Then the next row is scanned from the beginning and boxes are numbered consecutively. The same process is repeated for all rows and finally 256 boxes are spotted. The State-Space Point Distribution (SSPD) for each pattern is calculated by estimating the number of points in each of these 256 boxes. Using the above information the State-Space Point Distribution graph is plotted by taking the box number in X-axis and the number of points in each box in Y-axis.
Figure 5.7 State-Space Map with 16 x16 locations

Figure 5.8 (a-f) shows the state-space point distribution graph for the selected Malayalam handwritten characters अ [ah], ओ [ga], ए [ta], व [th], र [pa] and ओ [r] obtained from the SSMs with eight directional space variations respectively. The state-space point distribution graph generated for different samples of the above characters are shown in figure 5.9 (a-f). The state-space map and its corresponding state-space point distribution graph obtained for different characters show the identity of that character as regard to the pattern. From the above analysis it is found that an efficient feature vector based on the SSPD can be effectively used to represent Malayalam handwritten characters and hence it can be set as an input to the recognition systems.
Fig. 5.8 (a) SSPD graph of Malayalam handwritten character ആ [-ah]

Fig. 5.8 (b) SSPD graph of Malayalam handwritten character ഗ [ga]
Fig. 5.8 (c) SSPD graph of Malayalam handwritten character s[ta]

Fig. 5.8 (d) SSPD graph of Malayalam handwritten character ω[thha]
Fig. 5.8 (e) SSPD graph of Malayalam handwritten character α [pa]

Fig. 5.8 (f) SSPD graph of Malayalam handwritten character ρ[rha]

Figure 5.8 (a-f): SSPD graph of Malayalam handwritten characters ṃ [-ah], ω [ga], ṣ [ta], ॐ thha], α [pa] and ρ[rha] obtained from state-space maps with eight directional space variations
Figure 5.9 (a-f): SSPD graphs of different samples of Malayalam handwritten characters ʍ[-ah], ʍ[ga], ʍ[ta], ʍ[thha], ʍ[pa] and ʍ[rha] obtained from the SSMs with eight directional space variations.
The feature vector is estimated by computing the average distribution of points in the sixteen consecutive boxes in the SSPD graph. Since there are 256 boxes in the SSPD graph, the feature vector dimension is reduced to sixteen. Figure 5.10 (a-c) shows the feature vector graphs obtained from Malayalam handwritten characters മ[-ah], ഓ [ga] and സ [ta] plotted for fifteen samples of each character along with the mean curve. The graphs obtained for different samples of same character are similar whereas graphs for different characters are highly distinguishable. Detailed studies on the recognition experiments based on the SSPD features are performed in succeeding chapters.

![SSPD Feature Graph](image)

**Figure 5.10(a) SSPD feature graph for Malayalam handwritten character മ[-ah]**
Figure 5.10 (b) SSPD feature graph for Malayalam handwritten character ө [ga]

Figure 5.10(c) SSPD feature graph for Malayalam handwritten character ș [ta]
5.5 Conclusions

A novel method for modeling Malayalam handwritten characters based on the gray-scale images is introduced. Reconstructed state-spaces in the form of State-Space Maps (SSMs) for character images are generated for one, four and eight directional space variations. State-space maps with eight directional space variations are found to be well-informed and hence it is used further in the feature extraction process. Sate-Space Point Distribution (SSPD) features are extracted from each of the character samples for recognition purpose. The SSPD feature vector graph plotted for each character shows that the extracted features are similar for different image samples of a given character and at the same time they considerably vary from character to character.

The proposed method is very easy to implement since it does not require binarization. The difficulties in threshold selection and the subsequent loss of information caused by the process of binarization (that may results in broken strokes in binarized character image) are also eliminated in this method. Added advantage is that the feature extraction procedure is computationally less complex and the dimension of the feature vector (sixteen) is comparatively small. So the SSPD parameters derived from the gray-scale based SSMs of handwritten character samples can be effectively utilized for high speed HCR applications.