Two-Dimensional Model of Hidden Markov Mesh

신봉기 Bong-Kee Sin
부경대학교 Pukyong National University
bkshin@pknu.ac.kr

Abstract

The new model proposed in this paper is the hidden Markov mesh model or the 2D HMM with the causality of top-down and left-right direction. With the addition of the causality constraint, two algorithms for the evaluation of a model and the maximum likelihood estimation of model parameters have been developed theoretically which are based on the forward-backward algorithm. It is a more natural extension of the 1D HMM than other 2D models. The proposed method will provide a useful way of modeling highly variable image patterns such as offline cursive characters.

Keyword : Hidden Markov model, 2D Markov mesh, Image model, Image analysis

1. Introduction

Hidden Markov model or HMM is a well-known statistical modeling tool for a variety of highly variable time-series or time-series-like signals. Its success, however, is limited to the analysis of 1D signal with an intrinsic order relation. Motivated by the success of the HMM, a number of researchers have tried to apply the model or its extensions to spatial 2D signals like digital images. The efforts thus far, however, have not been successful, although not a total failure either. This paper discusses a theoretical development of a 2D extension of the HMM.

The research on Markov models for 2D patterns is not new. It has been studied in several related areas such as image processing and character recognition. Although historically later to appear, the pseudo 2D HMM or P2DHMM is a simplified model of two-level hierarchy [1]; it is essentially a 1D HMM with vertical frame observations [2]. The P2DHMM is cost-effective for modeling patterns free of global shape deformation. This is also generally true of the truly 2D model of Markov random field or MRF [3, 4]. The MRF model has been studied and used by numerous researchers from diverse fields who are grappling with texture analysis, image restoration and segmentation [5]. It is, therefore, not strange that there are a large number of variants like Markov mesh and hidden Markov random field. For reasons of computational complexity most of the researchers adopted the causal types of MRFs for modeling images. Although there are studies on symmetric local dependence without directional causality, those methods suffer from cost-ineffectiveness problems and thus often resort to heuristic recipes. There is one study referring to 2D planar HMM [6]. The paper, however, focussed mainly on DP-based image match and made just a passing remark on the use of HMM without mathematical and/or practical development.

This paper describes a new development of 2D HMM. Distinct from the previous oversimplified Markov random field, the new model is called a hidden Markov mesh model or HMMM or simply 2D HMM that involves the causality of top-down and left-right direction. This causality, although not explicitly present in images, allows an efficient computation. In addition this paper introduces a lattice constraint under which 2D HMM can be locally scaled up or down just like the time-warping of 1D HMM. With the lattice constraint, the algorithms for evaluation of a model and maximum likelihood estimation of model parameters have been developed.

The proposed model is different from the MRF in that homogeneity (and isotropism, of course) is not enforced. The effect is that many more interesting forms of image variations including local shape distortions can be modeled more systematically. In other words, with the HMMM, many more types of image distortions can now be explained with the most likely Markov mesh lattices constructed by the decoding algorithm.

In the rest of the paper, we will first address the definition of the HMMM and a DP-based evaluation algorithm in Section 2. The description of the algorithm is the 2D version of Viterbi algorithm. Therefore the model decoding in Section 3 will be made short with a few additional remarks. Section 4 presents an MLE-based re-estimation algorithm for training model parameters, which is the most difficult task of the three. The final section discusses implications of the proposed method, and then concludes the paper.

2. Hidden Markov Mesh Model

HMM is a statistical model for analyzing 1D time sequential signal. Many interesting time series data are characterized by a strict order that can be described by the time evolution of the model states. In the HMM theory, the evolution is modeled by an underlying Markov chain with probabilistic transitions between states. In 2D HMM it is the Markov mesh lattice that models the local spatial deformation.

2.1 2D Markov Models

A stochastic process $\mathbf{X} = \{X_n,\ n = 1, 2, \ldots\}$ where each variable taking on a finite number of possible values is a Markov chain if there is a fixed probability

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1\} = P_{ij}$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and $n \geq 0$. Usually this type of Markov chain is sufficient for modeling one-dimensional time-series signals where a variable is related to only one or less variable that precedes it when viewed in time dimension. In higher dimensions, however, this type of simple chain structure is not adequate, and one or more additional variables are required. Let us from now on limit our discussion to two-dimensional signals such as a rectangular image consisting of a lattice of pixels. Needless to say, the model will be easily extended to third or higher dimensions.

There are two equivalent ways of defining random configurations of points on a lattice. One is based on the formulation of statistical mechanics according to J. Gibbs. Called as Gibbs ensemble or Gibbs random field, it is generally accepted as the simplest useful mathematical model of discrete or lattice gas. The second class of random fields is Markov random field, whose foundation dates back to the physics literature on ferromagnetism originating in the work of E. Ising in 1925 [8]. This extends in a simple way the notion of Markov process with one dimensional, integer valued, time to the case of higher dimensional, lattice valued, space parameter.

Let $\mathbf{L} = \{(i, j) : 1 \leq i \leq M,\ 1 \leq j \leq M\}$ be a two-dimensional rectangular lattice with $L = MN$ sites arranged as a planar mesh. M and N denote the vertical and horizontal dimension of the lattice respectively. For convenience let us denote the state or site identifiers as $i = 1, \ldots, L$ in row-major order. For each site i in the lattice we find a set of sites which are adjacent to and thus condition the state of the current site. It is called a neighborhood. The neighborhood of a site i in the lattice $\mathbf{L}$ is a set of sites that can influence behavior of the site i. In general the neighborhood system is defined as follows: $\eta = \{\eta_i \subseteq \mathbf{L}: i \in \mathbf{L}\}$. Here η_i is the neighborhood of a site i, and satisfies that $i \notin \eta_i$ and $j \in \eta_i$ if and only if $i \in \eta_j$. Then the definition of the MRF follows: given a lattice $\mathbf{L}$ and a neighborhood η, a random field $\mathbf{X} = \{X_j, j \in \mathbf{L}\}$ is an MRF if and only if

$$P(X_j = x_j \mid X_i = x_i, i \in \mathbf{L} - \{j\})$$
$$= P(X_j = x_j \mid X_i = x_i,\ i \in \eta_j), \quad \forall j \in \mathbf{L}$$

By definition, the MRF is homogeneous and isotropic. This property is highly appropriate for modeling systems of homogeneous gas particles or fluids, and restoring images corrupted by random noise. But the problem of such a noncausal random field model is that there is no known efficient and effective algorithm other than the formulation based on the Gibbs distribution. Several researchers have tried to solve the problem by introducing causality in the lattice. Two recent studies were reported by Park *et al.* [9, 10].

However, the MRF is still insufficient for modeling general image distortions other than random corruption of images. There are many more types of characteristic variations of images arising not from purely random sources but from sources explainable in statistical terms. This is particularly true of hand-written script. Such images involve local distortions characteristic of the target patterns in the images. We believe that they should be modeled with a new type of modeling framework that can represent various local variations. The one proposed in this paper is based on the model of Markov mesh lattice.

Just like an MRF, a general Markov mesh lattice has it that a site is determined by a set of neighbor sites. The difference lies in the definition of anisotropic inhomogeneous click potential which is defined as a probabilistic transition parameter

$$P(X_j = x_j \mid X_i = x_i, i \in \eta_j) = \prod_{i \in \eta_j} P_{ij}, \quad \forall j \in \mathbf{L}$$

with the stochastic constraints $P_{ij} \geq 0$ and $\sum_{j \in \eta_j} P_{ij} = 1$.

It is not necessarily that $P_{ij} = P_{ji}$ and $P_{ij} = P_{i+k,j+k}$, and this property allows the modeling of local spatial distortion.

2.2 2D Hidden Markov Mesh Model

Based on the concept of the Markov mesh lattice of the preceding section and the traditional HMM theory, we can define a two-dimensional hidden Markov mesh model (HMMM or simply 2D HMM). The model to be described henceforth is causal and allows an efficient computation. Formally the HMMM is defined as follows:

- **State transition parameters**

In a multi-dimensional space free of temporal arrow it is difficult to justify the introduction of any order, or causality. However, we have assumed an intuitive causality to reduce the computational requirement in the following way: first, there are two types of transitions: the downward transition from the upper neighborhood and the rightward from the left neighborhood. Second, we restrict the site transitions to those to and from the set of 8-neighbors.

The resulting mesh topology of the model is shown in Figure 1. For a given node, say *j*, the set of upper neighbors that can act as an upper source node is denoted by η_j^U. Similarly the left neighborhood is denoted by η_j^L. The right and the lower neighborhoods η_j^R and η_j^D respectively denote the sets of right and the lower destination nodes of rightward and downward transitions respectively from the site *j*. Using the two types of transitions between neighbor sites, we can construct a complete lattice **Y** of 2D HMM states corresponding to an image, a lattice of pixels.

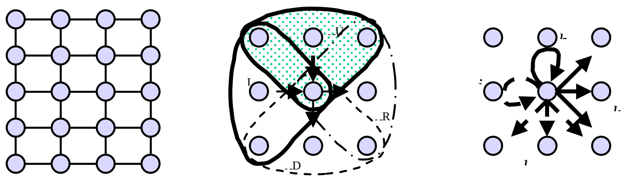

Figure 1. Mesh lattice structure, four types of neighborhood, and (two types of) state transition arcs from a state: broken arrows and solid arrows.

In 2D space there are two types of transitions: vertical downward transitions and horizontal rightward transitions, each parameterized as follows:

$$a_{hi}^{\downarrow} = P(q_u = j \mid q_{u-V} = h),$$
$$h \in \eta_j^{\mathrm{U}} = \{\, j-N-1,\ j-N,\ j-N+1,\ j \,\}$$
$$a_{ij}^{\rightarrow} = P(q_u = j \mid q_{u-1} = i),$$
$$i \in \eta_j^{\mathrm{L}} = \{\, j-N-1,\ j-1,\ j+N+1,\ j \,\}$$

These parameters define the causality, vertical and downward, as assumed before. Also it is noted that the following stochastic constraints are satisfied

$$\sum_{k \in \eta_j^{\mathrm{R}}} a_{jk}^{\rightarrow} = 1 \ \text{ and } \ a_{jk}^{\rightarrow} \ge 0$$
$$\sum_{l \in \eta_j^{\mathrm{D}}} a_{jl}^{\downarrow} = 1 \ \text{ and } \ a_{jl}^{\downarrow} \ge 0 \ . \qquad (1)$$

Here η_j^{R} and η_j^{D} denote the right and the lower neighborhood respectively.

• Observation parameters

In the conventional HMM, the observation is another stochastic process which is a probabilistic function of an underlying Markov chain. The observation of an HMMM is an image $\mathbf{X} = \{x_{uv} \in \Omega : 1 \le u \le U,\ 1 \le v \le V\}$ in the rectangular arrangement of $W = UV$ pixels. $\Omega = \{1, 2, \ldots, K\}$ is a set of K color or gray scale values. Here again let us identify the pixels in the row-major order as $u = 1, \ldots, W$.

The observation of of $\mathbf{X}$ is a function of the above lattice process parameters. The observation symbols x_{uv} are independent of all the others. This type of conditional independence assumption is grossly inaccurate, but allows an efficient computation and usually works well enough. The parameters are:

$$b_j(v) = P(x_u = v \mid q_u = j), \quad j \in \mathbf{L}, v \in \Omega$$

where

$$\sum_v b_j(v) = 1, \quad j \in \mathbf{L}\ .$$

Every site in a Markov mesh is conditioned by its neighboring sites, and the orderly collection of the sites organizes a lattice through an artificial causal chain. Each pixel in an image $\mathbf{X}$ is observed from an HMMM site as a result of a conditionally independent process of the corresponding site. The capability of modeling spatial distortions or spectral variations depends on the organization of neighborhood system (as in the case of MRF) or the transition probability parameters of the 2D HMM.

2.3 Lattice Process

Now, given the causality for a 2D lattice, we can proceed to define the following two recurrence relations based on the Markovian property and the Bellman's optimality principle of dynamic programming. They are the forward probability and the backward probability:

$$\alpha_u(j) = \max_{h \in \eta_j^U, i \in \eta_j^L} a_{ij}^{\rightarrow} a_{hj}^{\downarrow} b_j(x_u) \alpha_{u-1}(i), \quad j = 1, \ldots, L,\ u = 1, \ldots, W \qquad (2)$$

$$\beta_u(j) = \max_{k \in \eta_j^R, l \in \eta_j^D} a_{jk}^{\rightarrow} a_{jl}^{\downarrow} b_k(x_{u+1}) \beta_{u+1}(k), \quad j = L, \ldots, 1,\ \ u = W, \ldots, 1 \ . \qquad (3)$$

The forward probability $\alpha_u(j)$ is the maximum probability of observing the partial region of the image $\mathbf{X}_{1,u} = x_1\, x_2 \ldots x_u$ from the partial mesh of states $\mathbf{Y}_{1,j} = y_1\, y_2 \ldots y_j$. The backward probability $\beta_u(j)$ denotes the probability of observing the remaining image region $\mathbf{X}_{u+1,W} = x_{u+1}\, x_{u+2} \ldots x_W$ after x_u from the remaining partial mesh of states $\mathbf{Y}_{j+1,L} = y_{j+1}\, y_{j+2} \ldots y_L$. The boundary conditions are:

$$\alpha_1(1) = b_1(x_1)$$
$$\alpha_u(j) = \max_{i \in \eta_j^L} a_{ij}^{\rightarrow} b_j(x_u) \alpha_{u-1}(i), \ \ j = 1, \ldots, N,\ \ u = 2, \ldots, V$$

$$\beta_W(L) = 1$$
$$\beta_u(j) = \max_{k \in \eta_j^R} a_{jk}^{\rightarrow} b_k(x_{u+1}) \beta_{u+1}(k), \ \ j = L, \ldots, L-N+1, \ .$$
$$u = W, \ldots, W-V+1$$

The forward DP continues while keeping the mesh lattice-related information as

$$(h^*, i^*)_u(j) = \arg\max_{h \in \eta_j^U, i \in \eta_j^L} a_{ij}^{\rightarrow} a_{hj}^{\downarrow} b_j(x_u) \alpha_{u-1}(i), \quad j = i, \ldots, L,\ \ u = 1, \ldots, W \qquad (4a)$$

$$(k^*, l^*)_u(j) = \arg\max_{k \in \eta_j^R, l \in \eta_j^D} a_{jk}^{\rightarrow} a_{jl}^{\downarrow} b_k(x_{u+1}) \beta_{u+1}(k), \quad j = L, \ldots, 1,\ \ u = W, \ldots, 1 \qquad (4b)$$

Here, let us write

$i^* = Left(j)$

$h^* = Up(j)$

$k^* = Right(j)$

$l^* = Down(j)$

Then we have another requirement for building a complete mesh lattice of states as the result of computation in (2) and (3).

$$Left(h^*) = Up(i^*),$$
$$Down(k^*) = Right(l^*) \quad (5)$$

And

$$\alpha_u(j) = \max_{h \in \eta_j^U} a_{hj}^{\downarrow} b_j(x_u) \alpha_{u-1}(R_h),$$
$$j = 1, N+1, \ldots, (M-1)N+1, \quad (6a)$$
$$u = V+1, 2V+1, \ldots, (U-1)V+1$$

$$\beta_u(j) = \max_{l \in \eta_j^D} a_{jl}^{\downarrow} b_k(x_{u+1}) \beta_{u+1}(L_l), \quad (6b)$$
$$j = L, L-N, \ldots, N, \quad u = W-V, W-2V, \ldots, V$$

Here $R_h = Right^{N-1}(h)$, $L_l = Left^{N-1}(l)$, and each indicates the rightmost boundary node of h and the leftmost boundary node of l. The power notation is defined by the recursion for all n as a composite function:

$$Right^n(x) = Right(Right^{n-1}(x))$$
$$Left^n(x) = Left(Left^{n-1}(x))$$

The equations (5) and (6) constitute the lattice constraints for a complete mesh lattice.

Using the forward and backward probabilities of (2) and (3), we can complete the calculation as

$$P(\mathbf{X} \mid \Lambda) = \max_j \alpha_u(j) [\prod_{k=1}^{V-1} a_{h_k j_k}^{\downarrow}] \beta_u(j) \quad (7)$$

Here again any complete mesh lattice requires that

$$h_k = y_{u-V+k} = Right(h_{k-1}) \text{ and } h_0 = h,$$
$$j_k = y_{u+k} = Right(j_{k-1}) \text{ and } j_0 = j$$

the condition for 'sewing' together the forward and the backward lattice patches to obtain a complete rectangular lattice. The resulting mesh of states will be a planar lattice locally warped to model the locally deformed 2D patterns. And it is noted that this type of two-dimensional lattice model is different from the second order Markov chain [11] in that this does not impose the lattice constraint which leads to the construction of a regular mesh of states.

3. Lattice Decoding

In this paper an image is defined as a realization of a stochastic process that, in turn, is defined over an underlying Markov mesh. Each observation is a function of the corresponding site of the Markov mesh but is assumed to be independent of other observations. The individual states of a Markov mesh lattice are determined based on their neighborhood. But, since the states in the lattice depend on their location, the model is also distinguished from the homogeneous MRF that is not aware of model topology.

The decoding of 2D HMM Λ is the problem of finding the optimal Markov mesh lattice **Y*** of maximum likelihood given an observation image **X**. Mathematically it is defined as the task of maximizing $P(\mathbf{X}, \mathbf{Y} \mid \Lambda)$ over all possible chains **Y**. **Y** is a complete mesh lattice of sites. In the preceding section, we have already defined the forward probability in (2) in terms of the best realization of initial partial lattices. The final probability is the very result of decoding. Finally the optimal Markov mesh lattice can be obtained by backtracking the result of forward pass using the information of (4) after the forward pass is over.

In standard theory of hidden Markov modeling, the model evaluation is based on the concept of total probability of observing an input signal given a model, which is given by

$$P(\mathbf{X} \mid \Lambda) = \sum_{\mathbf{L}} P(\mathbf{X}, \mathbf{L} \mid \Lambda) \, P(\mathbf{L} \mid \Lambda)$$

Although correct in statistical context, there is a difficulty in interpreting the result of computation. Namely, given a model of planar topology, one is asked whether it is possible to generate the rectangular image without regard to the topology. This problem leads us to define the optimization criterion as the joint probability of the lattice of states as well as the input image. Formally it is given by the following formula

$$P(\mathbf{X} \mid \Lambda) = \max_j \alpha_u(j)[\prod_{k=1}^{V-1} a^{\downarrow}_{h_k j_k}]\beta_u(j)$$

as is given in the preceding section. In effect, this is the equation for decoding the model Λ given an image **X**.

4. Parameter Estimation

The parameter estimation problem is concerned with finding the optimal set of model parameters given a set of typical samples. Let us write **Y** be a Markov mesh chain given a sample image **X**. The likelihood of observing **X** from the Λ of the model is

$$P(\mathbf{X} \mid \Lambda) = \sum_{\mathbf{Y}} P(\mathbf{X}, \mathbf{Y} \mid \Lambda)$$

Each term in the right hand side is the joint probability written as

$$P(\mathbf{X}, \mathbf{Y} \mid \Lambda) = \prod_{u=1}^{W} [a^{\rightarrow}_{y_{u-1}, y_u} a^{\downarrow}_{y_{u-V}, y_u} b_{y_u}(x_u)] \quad . \tag{8}$$

By taking the logarithm of it, we have

$$\begin{aligned} &\log P(\mathbf{X}, \mathbf{Y} \mid \Lambda) \\ &= \sum_{u=1}^{W} (\log a^{\rightarrow}_{y_{u-1}, y_u} + \log a^{\downarrow}_{y_{u-V}, y_u} + \log b_{y_u}(x_u)) \end{aligned} \tag{9}$$

Following the line of Baum's reasoning with the *Q*-function [7], we can now define a similar auxiliary for the 2D HMM as follows:

$$\begin{aligned} Q(\Lambda, \hat{\Lambda}) &= 1/P(\mathbf{X} \mid \Lambda) \sum_{Y} P(\mathbf{X}, \mathbf{Y} \mid \Lambda) \log P(\mathbf{X}, \mathbf{Y} \mid \hat{\Lambda}) \\ &= 1/P \sum_{Y = y_1 y_2 \cdots y_W} P(\mathbf{X}, \mathbf{Y} \mid \Lambda) \times \sum_{u=1}^{W} (\log \hat{a}^{\rightarrow} + \log \hat{a}^{\downarrow} + \log \hat{b}) \\ &= 1/P \sum_i \sum_j \sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda) \log \hat{a}^{\rightarrow} \\ &\quad + 1/P \sum_i \sum_j \sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda) \log \hat{a}^{\downarrow} \\ &\quad + 1/P \sum_i \sum_j \sum_{u: x_u = k} P(\mathbf{X}, y_u = j \mid \Lambda) \log \hat{b} \end{aligned} \tag{10}$$

where $P = P(X \mid \Lambda)$. The last expression can be reorganized as

$$\begin{aligned} Q(\Lambda, \hat{\Lambda}) &= \sum_h \sum_j c_{hj} \log \hat{a}^{\downarrow}_{hj} + \sum_i \sum_j d_{ij} \log \hat{a}^{\rightarrow}_{ij} \\ &\quad + \sum_j \sum_k e_{jk} \log \hat{b}_j(x_k) \end{aligned} \tag{11}$$

where

$$\begin{aligned} c_{hj} &= \sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda) / P(\mathbf{X} \mid \Lambda) \\ d_{ij} &= \sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda) / P(\mathbf{X} \mid \Lambda) \\ e_{ik} &= \sum_{u: x_u = k} P(\mathbf{X}, y_u = j \mid \Lambda) / \sum_u P(X, y_u = j \mid \Lambda) \end{aligned}$$

Then the resulting formulae for re-estimating the parameters are as follows:

$$\hat{a}^{\rightarrow}_{ij} = \frac{\sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda)}{\sum_j \sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda)} \tag{12}$$

$$\hat{a}^{\downarrow}_{hj} = \frac{\sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda)}{\sum_j \sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda)} \tag{13}$$

$$\begin{aligned} \hat{b}_{jk} &= \sum_{u: x_u = k} P(\mathbf{X}, y_u = j \mid \Lambda) \Big/ \sum_k \sum_{u: x_u = k} P(\mathbf{X}, y_u = j \mid \Lambda) \\ &= \sum_{u: x_u = k} P(\mathbf{X}, y_u = j \mid \Lambda) \Big/ \sum_u P(\mathbf{X}, y_u = j \mid \Lambda) \end{aligned} \tag{14}$$

Let us consider the *Q*-function of EM algorithm as a function of $\hat{\Lambda}$. Although the above function has more parameters than the corresponding function of 1D HMM, they are essentially of the same form. Therefore we can say the above re-estimation algorithm converges. It is stated in the following theorem.

[Theorem 1] *If* $Q(\Lambda, \hat{\Lambda}) \geq Q(\Lambda, \Lambda)$, *then* $P(\mathbf{X} \mid \hat{\Lambda}) \geq P(\mathbf{X} \mid \Lambda)$. *The equality holds when* $P(\mathbf{X} \mid \hat{\Lambda}) = P(\mathbf{X} \mid \Lambda)$.

Proof: From the concavity of the log function it follows that

$$\begin{aligned} \log \frac{P(\mathbf{X} \mid \hat{\Lambda})}{P(\mathbf{X} \mid \Lambda)} &= \log \left[\frac{\sum_{\mathbf{Y}} P(\mathbf{X}, \mathbf{Y} \mid \hat{\Lambda})}{P(\mathbf{X} \mid \Lambda)} \right] \\ &= \log \sum_{\mathbf{Y}} \frac{P(\mathbf{X}, \mathbf{Y} \mid \Lambda)}{P(\mathbf{X} \mid \Lambda)} \times \frac{P(\mathbf{X}, \mathbf{Y} \mid \hat{\Lambda})}{P(\mathbf{X}, \mathbf{Y} \mid \Lambda)} \\ &\geq \sum_{\mathbf{Y}} \frac{P(\mathbf{X}, \mathbf{Y} \mid \Lambda)}{P(\mathbf{X} \mid \Lambda)} \log \frac{P(\mathbf{X}, \mathbf{Y} \mid \hat{\Lambda})}{P(\mathbf{X}, \mathbf{Y} \mid \Lambda)} \\ &= Q(\Lambda, \hat{\Lambda}) - Q(\Lambda, \Lambda) \end{aligned}$$

where the inequality is due to the well-known Jensen's inequality. The above inequality says that Λ is a critical point of $P(\mathbf{X} \mid \Lambda)$ if and only if it is a critical point of *Q* as a function of Λ. QED.

According to the above result, if the newly estimated model $\hat{\Lambda}$ makes the right-hand side positive, the algorithm is guaranteed to improve the model likelihood $P(\mathbf{X}\,|\,\Lambda)$. The improvement then results in $\hat{\Lambda}$ that maximizes the Q-function unless a critical point is reached [7].

5. Discussion and Conclusion

Before a detailed discussion of the proposed model, let us look at the sample results from the current prototype system. Figure 2 illustrates a test result with a very simple model and an image, which lends to visual analysis.

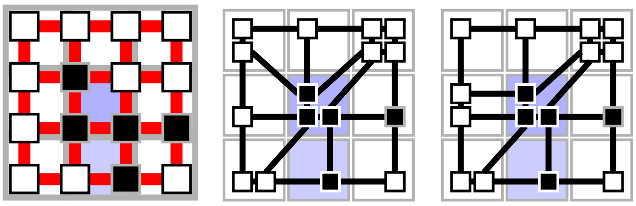

Figure 2. A simple image overlaid on a simple 3-by-3 state 2D HMM, and the two best candidates lattices obtained. Note that the black pixels in the input image are well mapped to dark states (shown in shaded blocks) while maintaining a deformed mesh lattice topology.

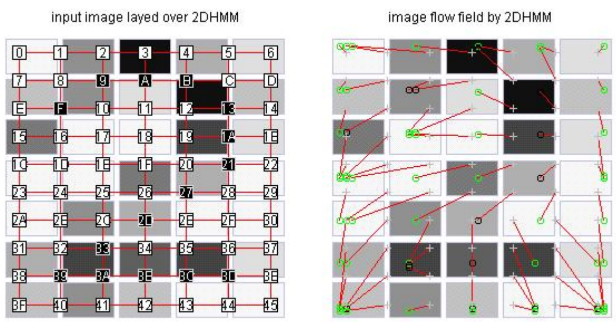

Figure 3. A test model for digit '2' is shown at the background. It has been trained from noisy images. The mesh small blocks in the foreground are then warped with the resulting warping field is shown in the right. The black pixels are appropriately mapped to the dark states. The line segments denote the vector of displacement with the small circles marking their head.

The most essential difference between 1D time series data and 2D image is the existence or absence of intrinsic order between components in the signals. Thus we have raised the issue of introducing a putative order in order to utilize the sequential processing of modern computers.

The MRF as a model for 2D image is described by a small number of parameters defining clique potential (instead of transition probability) subject to the global homogeneity condition. Unlike the MRF or other Markov mesh models, the proposed 2D HMM has been defined based on the concept of the probabilistic function of Markov mesh lattice. The mesh lattice is a natural extension of the sequential chain. Therefore the HMMM is a natural extension 1D HMM, and a truer 2D HMM.

The causality referred to in the paper is not new; it has already been used in the previous studies on mesh models such as mesh random field. One distinguishing feature of the current method is the lattice constraint that constrains the search for only complete lattices. Naturally this has led to the use of Viterbi-type of decoding algorithm. Another noteworthy feature is that, with the introduction of site-to-site transition parameters, local spatial distortions can be modeled. We believe that this type of capability should be considered in modeling patterns of high variability, which is highly unpredictable globally, and thus less likely to be parameterized, but can be anticipated and modeled locally, remotely based on the study the psychomotor of handwriting. The author believes that the Markovian assumption fits appropriately with the latter points.

In Section 2 we have assumed a model topology with 2nd order neighborhood in addition to directional causality. This will enable us to reduce the computational load drastically from $O(L^2W) = O(M^2N^2UV)$ of general ergodic models down to $O(LW) = O(MNUV)$ without decreasing the modeling power in general.

The 2D HML proposed in this paper is different from HMMRFs or mesh MRFs, and it is true even in the basic assumptions. The previous field models are too simple to model a wide range of shape variations occurring in images. But the HMMM is capable of decoding strictly local shape deformation, a task that may be called as dynamic space warping, in contrast to the dynamic time warping. The structure of the model is better suited for local nonlinear variation of a reference image, be it scaling or distortion either globally or locally.

The final remark is that the 2D HMM is not limited to two-dimensional space modeling, therefore the model can be called simply as HMMM instead of 2D HML. For 2D case, however, it is certain that the model will make a useful tool in such tasks as off-line handwritten character recognition, nonlinear motion field analysis.

References

[1] O. Agazzi and S. Kuo, “Pseudo two-dimensional hidden Markov models for document recognition,” *AT&T Technical Journal*, vol. 72, no.5, pp. 60-72, Sep.-Oct. 1993.

[2] W. Cho, S.-W. Lee and J. H. Kim, “Modeling and recognition of cursive words with hidden Markov models,” *Pattern Recognition*, vol. 28, no. 12, pp. 1941-1953, 1995.

[3] J. Woods, “Two-dimensional discrete Markovian fields,” *IEEE Trans. Info. Theory*, vol. IT-18, no. 2, pp. 232-240, March 1972.

[4] A. Jain, “Advances in mathematical models for image processing,” *Proc. IEEE*, vol. 69, pp. 502-528, 1981.

[5] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. PAMI*, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984.

[6] E. Levin and R. Pieraccini, Dynamic planar warping for optical character recognition,” in *Proc. Int’l Conf. On Accoustics, Speech and Signal Processing*, San Francisco, California, pp. 23-26, March 1992.

[7] L. Baum and J. Eagon, “An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a models for ecology,” *Bull. Ann. Math. Stat.*, vol. 73, pp. 360-363, 1967.

[8] E. Ising, *Zeitschrift Physik*, vol. 31, p. 253, 1925.

[9] H.-S. Park and S.-W. Lee, “A truly 2-D hidden Markov model for off-line handwritten character recognition,” *Pattern Recognition*, vol. 31, no. 12, pp. 1849-1864, 1998.

[10] G. Saon, “Cursive word recognition using a random field based hidden Markov model,” *Int’l J. Doc. Aanal. and Recog.*, vol. 1, no. 4, pp. 199-208, 1999.

[11] A. Kundu, Y. He, and P. Bahl, “Recognition of handwritten word: first and second hidden Markov model based approach,” *Pattern Recognition*, vol. 22, no. 3, pp. 283-297, 1989.