2D Model of Hidden Markov Mesh Based on Complete Lattices

Bong-Kee Sin, Sung-Un Kim

2D Model of Hidden Markov Mesh Based on Complete Lattices

Bong-Kee Sin
Computer Multimedia Div.,
Pukyong Nat'l Univ., Busan, Korea
bkshin@pknu.ac.kr

Sung-Un Kim
Electronics and Telecomm. Div.,
Pukyong Nat'l Univ., Busan, Korea
kimsu@pknu.ac.kr

Abstract

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

Keywords: 2D HMM, Image model, Shape distortion.

1. Introduction

Image registration is the task of determining the nature of distortion and transformation of an image with respect to a reference image. The reference image is usually of the same format and dimension in many applications. In general, however, the test image to register can be of any size, smaller or bigger than the reference image. Therefore it is required to deal with size variations between two images for registration purpose.

On the other hand image matching is a task of image analysis involving comparing 2D patterns of different or changing shape. This task is made difficult by the presence of nonlinear local distortion in the test patterns relative to the reference. These are two important problems found in shape analysis in general and handwritten character recognition in particular.

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. 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 very successful, albeit not a total failure. 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. 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 [2]. The MRF model has been studied and used by numerous researchers from diverse fields who are grappling with texture analysis, image restoration and segmentation [3]. It is, therefore, not surprising that there are a large number of variants like Markov mesh and hidden Markov random field. For reasons of computational complexity some 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 [4]. The paper, however, focused mainly on DP-based image match and made just a passing remark on the use of HMM without mathematical and 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. In addition, although it has not been proved yet, we consider that the causality introduced is not a required feature of the model but rather a useful feature enabling an effective sequential computation. With the lattice constraint, the algorithms for the 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 isotropy, of course) is not enforced. The effect is that many more interesting types of image variation including local shape distortions can be modeled more systematically (see Figure 1). 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. One implication of the complete lattice is that the artificial causality is not a factor constraining the modeling power but one that facilitates the computation.

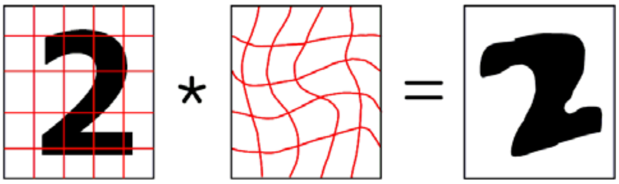


Figure 1. The concept of decoding the distorted 2D lattice; it is assumed that some hidden image on the left is nonlinearly and locally distorted to appear as the rightmost pattern. The middle mesh is the warping function that we want to recover.

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 algorithm is the 2D version of Viterbi algorithm. Therefore the model decoding in Section 3 will be made brief with a few additional remarks. Section 4 presents an MLE-based re-estimation algorithm for training models. The final section discusses the implications of the proposed model, 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 distortion.

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

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

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 each variable is related to only one variable that immediately precedes it when viewed in time dimension. In the case of 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. The model will easily be 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 [6]. 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 $\mathbf{L} = \{(i, j) : 1 \leq i \leq M, 1 \leq j \leq N\}$ 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 together 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.

However, the MRF is still inappropriate 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 define a two-dimensional hidden Markov mesh model (HMMM or simply 2D HMM). The model to be described here is causal and allows 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 one-directional 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 2. 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.

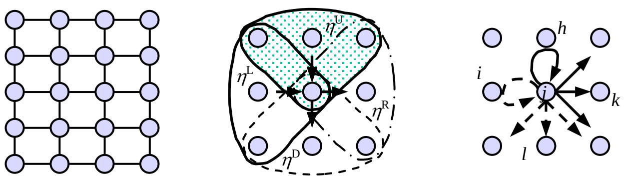


Figure 2. An undistorted mesh lattice, four types of neighborhood to the central node, and four possible arcs(solid arrows) to the right nodes in η^R and another four(broken ones) to the lower nodes in η^D.

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

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

where u is the index of a pixel and V is the width of the input image. The above parameters define the causality, vertical and downward, as assumed before. Also it is noted that the following stochastic constraints are satisfied

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

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 $\mathbf{X} = \{x_{uv} \in \Omega : 1 \leq u \leq U,\ 1 \leq v \leq V\}$ 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 **X** is a function of the above lattice process parameters. The observation symbols x_u are independent of all the others. This type of conditional independence assumption is not accurate, but enables 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 **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 (just like 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:

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

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

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 $\mathbf{Y}_{1,u} = y_1 y_2 \ldots y_u$ where $y_u = j$. 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 $\mathbf{Y}_{u+1,W} = y_{u+1} y_{u+2} \ldots y_W$ where $y_W = L$. The boundary conditions are:

$$\alpha_1(1) = b_1(x_1)$$
$$\alpha_u(j) = \max_{i \in \eta_j^L} a_{ij}^{\rightarrow} b_j(x_u) \alpha_{u-1}(i), \ \ j = 1, \cdots, N, \ \ u = 2, \cdots, V$$
$$\beta_W(L) = 1 \qquad u = W, \cdots, W-V+1$$
$$\beta_u(j) = \max_{k \in \eta_j^R} a_{jk}^{\rightarrow} b_k(x_{u+1}) \beta_{u+1}(k), \ \ j = L, \cdots, L-N+1,$$

The forward DP continues while keeping the lattice-related information in

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

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

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 Eqs. (2) and (3):

$Left(h^*) = Up(i^*)$,

$Down(k^*) = Right(l^*)$. (5)

And

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

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

Here $R_h = Right^{N-1}(h)$ and $L_l = Left^{N-1}(l)$, and each indicates the rightmost boundary node of h and the leftmost boundary node of l respectively. 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 Eqs. (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

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

Here again any complete mesh lattice requires that

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

$$j_k = y_{u+k} = Right(j_{k-1}) \quad \text{and} \quad j_0 = j$$

the condition for ‘stitching’ 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 in that this does not impose the lattice constraint which leads to the construction of a regular mesh of states. There are two different types of transitions. And the lattice constraint is not temporal, but spatial, a desired characteristic for modeling spatial distortion.

3. Lattice Decoding

The decoding of 2D HMM Λ is the problem of finding the optimal Markov mesh lattice **Y*** of maximum likelihood given an image **X**. Mathematically speaking it is defined as the task of maximizing $P(\mathbf{X},\mathbf{Y}\,|\,\Lambda)$ 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 Eq. (2) in terms of the best realization of initial partial lattices. The final probability is none other than the result of decoding. After the forward pass the optimal Markov mesh lattice can be obtained by backtracking the result of forward computation using the information of Eq. (4).

In the 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

$$P(\mathbf{X}\,|\,\Lambda) = \sum_{\mathbf{L}} P(\mathbf{X}\,|\,\mathbf{L},\Lambda)\; P(\mathbf{L}\,|\,\Lambda) \tag{8}$$

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

$$P(\mathbf{X}\,|\,\Lambda) = \max_j \alpha_u(j)[\prod_{k=1}^{V-1} a_{h_k j_k}^{\downarrow}]\beta_u(j) \tag{9}$$

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 given a sample image **X**. The likelihood of observing **X** from the model Λ is

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

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

$$P(\mathbf{X},\mathbf{Y}\,|\,\Lambda) = \prod_{u=1}^{W} [a_{y_{u-1},y_u}^{\rightarrow} a_{y_{u-V},y_u}^{\downarrow} b_{y_u}(x_u)] \;. \tag{10}$$

By taking the logarithm of it, we have

$$\log P(\mathbf{X},\mathbf{Y}\,|\,\Lambda) = \sum_{u=1}^{W} (\log a_{y_{u-1},y_u}^{\rightarrow} + \log a_{y_{u-V},y_u}^{\downarrow} + \log b_{y_u}(x_u)) \tag{11}$$

Following the line of Baum’s reasoning with the Q-function [5], we can define a similar auxiliary for the 2D HMM as follows:

$$Q(\Lambda,\hat{\Lambda}) = 1/P(\mathbf{X}\,|\,\Lambda)\sum_Y P(\mathbf{X},\mathbf{Y}\,|\,\Lambda)\log P(\mathbf{X},\mathbf{Y}\,|\,\hat{\Lambda})$$

$$= 1/P \sum_{Y=y_1y_2\cdots y_W} P(\mathbf{X},\mathbf{Y}\,|\,\Lambda)\times\sum_{u=1}^{W}(\log \hat{a}^{\rightarrow} + \log \hat{a}^{\downarrow} + \log \hat{b})$$

$$
\begin{aligned}
&= 1/P\sum_i\sum_j\sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda)\log \hat{a}^{\rightarrow} \\
&\quad +1/P\sum_i\sum_j\sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda)\log \hat{a}^{\downarrow} \\
&\quad +1/P\sum_i\sum_j\sum_{u:x_u=k} P(\mathbf{X}, y_u = j \mid \Lambda)\log \hat{b}
\end{aligned} \tag{12}
$$

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

$$
\begin{aligned}
Q(\Lambda, \hat{\Lambda}) &= \sum_h\sum_j c_{hj} \log \hat{a}^{\downarrow}_{hj} + \sum_i\sum_j d_{ij} \log \hat{a}^{\rightarrow}_{ij} \\
&\quad + \sum_j\sum_k e_{jk} \log \hat{b}_j(x_k)
\end{aligned} \tag{13}
$$

where

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

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

$$
\hat{a}^{\rightarrow}_{ij} = \frac{\sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda)}{\sum_j\sum_u P(\mathbf{X}, y_{u-1} = i, y_u = j \mid \Lambda)} \tag{14}
$$

$$
\hat{a}^{\downarrow}_{hj} = \frac{\sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda)}{\sum_j\sum_u P(\mathbf{X}, y_{u-V} = h, y_u = j \mid \Lambda)} \tag{15}
$$

$$
\begin{aligned}
\hat{b}_{jk} &= \sum_{u:x_u=k} P(\mathbf{X}, y_u = j \mid \Lambda) \Big/ \sum_k \sum_{u:x_u=k} P(\mathbf{X}, y_u = j \mid \Lambda) \\
&= \sum_{u:x_u=k} P(\mathbf{X}, y_u = j \mid \Lambda) \Big/ \sum_u P(\mathbf{X}, y_u = j \mid \Lambda)
\end{aligned} \tag{16}
$$

The above formulae can be computed using the Eq. (9) or (10). Using the maximization in the forward pass it considers the single best lattice, and eliminates other less likely possibilities. This problem can be solved by the use of multiple samples and multiple lattice candidates.

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 we write

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

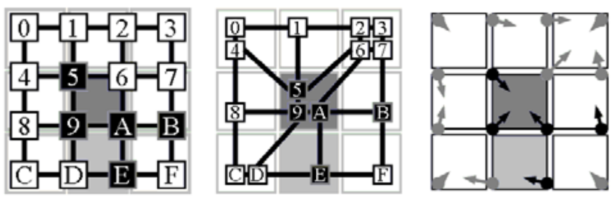
where the inequality is due to the well-known Jensen's inequality. The above inequality says that Λ is a critical point of $P(\mathbf{X} \mid \Lambda)$ if and only if it is a critical point of *Q* as a function of Λ. **QED**.

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 $P(\mathbf{X} \mid \Lambda)$. The improvement then results in $\hat{\Lambda}$ that maximizes the *Q*-function unless a critical point is reached [5].

5. Lattice Construction Tests

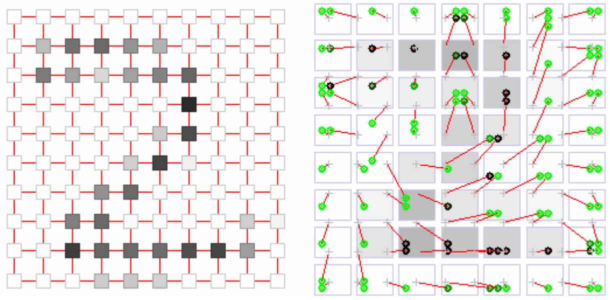
The immediate goal of the model evaluation is the construction of the most likely lattices corresponding to an input image. Unlike MRF modeling, the proposed model can be smaller than input images because the self-transitions in each state model the local shape variations, vertical and horizontal. Figure 3 shows a very simple example with a 3×3 state HMMM and an overlaid 4×4 image. The pixels are indexed in hexadecimal digits. The middle figure shows the best lattice in which three dark inner pixels are correctly mapped (or labeled) to the (dark) center state, while the remaining two dark pixels (B and E) are positioned appropriately because they cannot leave the boundary states. Simple visual analysis shows that the lattice is complete and highly convincing. The rightmost figure illustrates the warping function in the form of flow vector field for reference.

The second test was carried to check whether the model works in realistic setting. A sample result is shown in Figure 4, where 2D HMM for digit two is trained with a number of clean samples. The output distribution is shown in the form of state's (or block's) color. The pixel-to-state mapping function shows a nonlinear local mapping field. It is one 100 candidate lattices computed.



(a) (b) (c)

Figure 3. A simple lattice decoding example; (a) a 4-by-4 image overlaid on a simple 3-by-3 state 2D HMM (bottom or background), and (b) the best candidate lattice obtained. Note that the black pixels in the input image are well mapped to dark shaded states while maintaining a deformed mesh lattice topology. (c) Image warping function in the form of 2D vector flow field.



(a) (b)

Figure 4. (a) A 10×10 test image scaled down from the original 32×30. (b) The model for digit two is shown at the background. The flow field shows highly nonlinear local distortion. (Overall distortion seems to show a global rotation counterclockwise.)

6. Discussion

The proposed 2D HMM has been defined based on the concept of the probabilistic function of Markov mesh lattice. The mesh lattice is considered as the most natural 2D-extension to the sequential chain. And we believe that it is the most intuitive extension to 1D HMM, and the most interesting 2D HMM. We propose this view supported by the observation that the complete lattice constraint makes the introduced causality natural and required rather than awkward and power-limiting.

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 large 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 very well with the latter points.

The 2D HMM is capable of decoding strictly local nonlinear shape distortions, 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.

As is illustrated in Figure 5, however, large scale deformations (translations or rotations) may pose some difficulty. In the current implementation, this problem has occurred due to the boundary constraint requiring that boundary pixels be mapped to boundary nodes. One solution to this problem would be the use of global affine transformation prior to the lattice decoding for (**y**, **v**) as:

$$\mathbf{x} = A\mathbf{y} + \mathbf{v} = FA\mathbf{y}$$

where $\mathbf{x} \in \Re^W$ is an input image vector, A and F are affine and nonlinear transformation matrix respectively, and **v** and **y** are the decoded images after removing the distortions from A and the nonlinear distortion respectively.

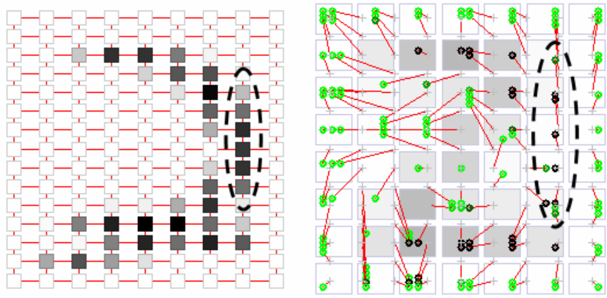


Figure 5. A global distortion example beyond small scale or local distortions. The boundary constraint of the method blocked modeling a large scale distortion marked in a broken circle.

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.

7. Conclusion

A 2D HMM for image distortion analysis has been proposed; the key features include the 'convenient' order or causality to utilize the sequential processing of modern computer, and the complete lattice constraint. Two algorithms for model evaluation and training were given, and the convergence was proved for the latter.

The 2D HMM is believed to be 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] J. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Info. Theory*, vol. IT-18, no. 2, pp. 232-240, March 1972.

[3] 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.

[4] 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.

[5] 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.

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