HOMOTOPY-BASED SEMI-SUPERVISED HIDDEN MARKOV TREE FOR TEXTURE ANALYSIS

Nilanjan Dasgupta, Shihao Ji, Lawrence Carin

Department of Electrical and Computer Engineering, Duke University, NC 27708

ABSTRACT

A semi-supervised hidden Markov tree (HMT) model is developed for texture analysis, incorporating both labeled and unlabeled data for training; the optimal balance between labeled and unlabeled data is estimated via the homotopy method. In traditional EM-based semi-supervised modeling, this balance is dictated by the relative size of labeled and unlabeled data, often leading to poor performance. Semi-supervised modeling may be viewed as a source allocation problem between labeled and unlabeled data, controlled by a parameter $\lambda \in [0, 1]$, where $\lambda = 0$ and 1 correspond to the purely supervised HMT model and purely unsupervised HMT-based clustering, respectively. We consider the homotopy method to track a path of fixed points starting from $\lambda = 0$, with the optimal source allocation identified as a critical transition point where the solution is unsupported by the initial labeled data. Experimental results on real textures demonstrate the superiority of this method compared to the EM-based semi-supervised HMT training.

1. INTRODUCTION

Semi-supervised learning exploits both labeled and unlabeled data to estimate parameters of an underlying model, yielding natural adaptivity to new unlabeled data. Labeled data are generally expensive to acquire and are often sparse, while unlabeled data are relatively inexpensive to acquire and therefore are often abundant. Presuming the existence of an underlying structure in the data, unlabeled samples may provide information about the data manifold, and they may be used to regularize a purely supervised solution. A conventional approach for parameter optimization with a generative model (*e.g.*, HMTs), incorporating both labeled and unlabeled data, is to use the expectation-maximization (EM) algorithm [1], in which the labels of the unlabeled data are treated as hidden variables, and the optimality criterion is the likelihood maximization of both labeled and unlabeled data [2]-[3]. However, the EM approach for semi-supervised learning is unstable, and Nigam *et al.* [3] proposed a heuristic way to alleviate the instability by weighting the contribution from the unlabeled data, while the choice of suitable scaling parameter remains an important issue. Corduneanu *et al.* [4] proposed the homotopy method for stable estimation of a naive Bayes classifier, where the optimal scaling parameter λ is selected at the point at which a critical transition occurs in the homotopy. We propose the homotopy method, a generalization of continuation [5], as an alternative to the EM-based solution for both supervised and semi-supervised HMTs in the context of texture classification, along with estimating the optimal balance λ specific to semi-supervised modeling. We present an overview of the HMT model and associated EM update equations in Sec. 2. The homotopy method is presented in Sec. 3, followed by its application to HMT parameter estimation in Sec. 4. Quantitative performance analyses of the proposed approach and conclusions are presented in Sec. 5 and 6, respectively.

2. WAVELET-BASED HMT MODEL AND EM-BASED PARAMETER ESTIMATION

Consider an image I_{cd}, sampled uniformly in two dimensions. Defining $LL_{cd}^0 = I_{cd}$, a sequential ν-level wavelet decomposition of I_{cd} yields four subsampled images: LL_{cd}^ν, HL_{cd}^ν, LH_{cd}^ν, and HH_{cd}^ν. Each point in HL_{cd}^ν, LH_{cd}^ν, and HH_{cd}^ν corresponds to the root node of a wavelet tree [6], and each node within a tree has four children at the next finer level (hence termed quadtrees). Each quadtree corresponds to a $2^\nu \times 2^\nu$ block in the original image I_{cd} [7], and our objective is to obtain a parametric model that captures the underlying statistics within the wavelet quadtrees.

The HMT is a statistical model that assumes a Markovian relationship between any wavelet node with its parent within a quadtree [6]. For simplicity the HH, HL and LH quadtrees are treated as statistically independent, and each node in a quadtree is modeled by a hidden M-state process, with each state represented by a Gaussian distribution parameterized by its mean and variance. Crouse *et al.* [6] developed an EM algorithm obtaining a *most likely* estimate of the model parameters $\boldsymbol{\theta}^y = \{\pi_m^y, \epsilon_{t,k,l}^y, \mu_{tm}^y, \sigma_{tm}^y\}$ for texture y with $m, k, l \in \{1, \cdots, M\}$, and $t, t' \in \{1, \cdots, R\}$ representing indices of the wavelet nodes numbered sequentially from the root to the leaves. The term π_m^y denotes the probability of hidden state m associated with the root node, and $\epsilon_{t,t'}^{y,k,l}$ defines the transition probability to hidden state $s_t (= k)$ from its parent $(= l)$. The terms μ_{tm}^y and σ_{tm}^y denote the mean and variance of the Gaussian distribution representing the mth state of the tth wavelet node. We have three such independent models for each texture, one for each subband, but we suppress band-specific notation here for simplicity.

Assume we have L labeled texture blocks $\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_L, y_L)\}$ and U unlabeled texture blocks $\{\boldsymbol{x}_{L+1}, \cdots, \boldsymbol{x}_{L+U}\}$, where each $\boldsymbol{x}_i$ corresponds to three quadtrees one for each subband. For both labeled and unlabeled texture blocks the associated wavelet coefficients are observable, whereas the underlying Markov states are hidden. In the context of classification among C textures, we employ distinct HMT models, one for each texture, and we wish to estimate the joint set of model parameters $\boldsymbol{\Theta} = \{w^y, \boldsymbol{\theta}^y\}_{y=1}^C$ based on the set of labeled and unlabeled data, where w^y represents the probability of class membership. Assuming that λ represents the balance between labeled and unlabeled data, the semi-supervised EM updates [8] for $\boldsymbol{\Theta}$ can be written as

$$w^y = \frac{(1-\lambda)}{L} \sum_{i=1}^L \delta_{y,y_i} + \frac{\lambda}{U} \sum_{j=L+1}^{L+U} p(y|\boldsymbol{x}_j)$$

$$\tilde{\pi}_m^y = \frac{(1-\lambda)}{L} \sum_{i=1}^L \gamma_{1,m}^{i,y} \delta_{y,y_i} + \frac{\lambda}{U} \sum_{j=L+1}^{L+U} \gamma_{1,m}^{j,y} p(y|\boldsymbol{x}_j) \quad (1)$$

$$\tilde{\epsilon}_{t,k,l}^y = \frac{(1-\lambda)}{L} \sum_{i=1}^L \xi_{t,k,l}^{i,y} \delta_{y,y_i} + \frac{\lambda}{U} \sum_{j=L+1}^{L+U} \xi_{t,k,l}^{j,y} p(y|\boldsymbol{x}_j)$$

where $\gamma_{1,m}^{i,y} = p(s_1 = m|\boldsymbol{x}_i, y)$, $\xi_{t,k,l}^{i,y} = p(s_t = k, s_{parent(t)} = l|\boldsymbol{x}_i, y)$. The update for the Gaussian parameters (mean $\boldsymbol{\mu}^y$ and variance $\boldsymbol{\sigma}^y$) [8] are not shown here for brevity. The above equations only present the M-step of the EM algorithm [1] for a particular iteration, whereas the E-step involves evaluating γ, ξ and $p(y|\boldsymbol{x})$ in terms of model parameters $\boldsymbol{\Theta}$, estimated at the previous iteration. Note that '$\sim$' in the LHS of the first two expressions denote unnormalized model parameters which are subsequently normalized during the E-step [6]. Iterative refinement of the model-parameters based on the E and M step yields a guaranteed convergence (albeit a local optima) on parameters $\boldsymbol{\Theta}$. Note that the above expressions assume λ to be known *a priori*, generally set to $U/(L + U)$ for semi-supervised EM modeling [8]. As an alternative, we propose the homotopy method for optimizing the HMT parameters, along with obtaining an optimal balance λ between the labeled and unlabeled data.

3. HOMOTOPY METHOD

The theory of the globally-convergent homotopy method involves finding zeros or fixed points of nonlinear system of equations [5],[9]. Rather than solving an original difficult problem $F(\boldsymbol{\Theta}) = \boldsymbol{0}$ directly, we start from an 'easy' problem $G(\boldsymbol{\Theta}) = \boldsymbol{0}$ having a known solution (or roots). We then track the solution while gradually transforming the 'easy' problem into the original one. A simple choice of the transformation function is

$$H(\boldsymbol{\Theta}, \lambda) = (1-\lambda)(\boldsymbol{\Theta} - \boldsymbol{a}) + \lambda F(\boldsymbol{\Theta}) = \boldsymbol{0}, \quad (2)$$

where $\boldsymbol{a} \in R^n$ and $F : R^n \to R^n$ is the original system of equations we want to solve, with $\lambda \in [0, 1]$ being a scalar parameter. Starting from a 'trivial' solution ($\boldsymbol{\Theta} = \boldsymbol{a}, \lambda = 0$), we gradually track the solution of $H(\boldsymbol{\Theta}, \lambda)$, with a final objective of obtaining ($\boldsymbol{\Theta} = \boldsymbol{\Theta}^*, \lambda = 1$) ($F(\boldsymbol{\Theta}^*) = 0$). The solution of $H(\boldsymbol{\Theta}, \lambda) = \boldsymbol{0}$ is a trajectory, found by solving the differential equation

$$\underbrace{\left[\frac{\partial}{\partial \boldsymbol{\Theta}} H(\boldsymbol{\Theta}, \lambda) \quad \frac{\partial}{\partial \lambda} H(\boldsymbol{\Theta}, \lambda)\right]}_{J} \left[\begin{array}{c} \dot{\boldsymbol{\Theta}} \\ \dot{\lambda} \end{array}\right] = 0, \quad \|\dot{\boldsymbol{\Theta}}, \dot{\lambda}\|_2 = 1, \quad (3)$$

with initial conditions $\lambda = 0$ and $\boldsymbol{\Theta} = \boldsymbol{a}$. The tangent vector $[\dot{\boldsymbol{\Theta}} \ \dot{\lambda}]^T$ lies in the one-dimensional null space of J (J is always full rank along the path [9]), and the direction is chosen to maintain an acute angle with the previous tangent vector (initial tangent direction is chosen to be $\dot{\boldsymbol{\Theta}} = \boldsymbol{0}$ and $\dot{\lambda} = 1$).

4. HOMOTOPY METHOD FOR HMT MODELING

4.1. Supervised HMT modeling

The homotopy method may be employed as an alternative to the EM-based parameter estimation for most generative models. In a purely supervised environment, a generative model is trained exclusively on data from one particular class (a particular texture in our problem). This is a special case of the semi-supervised scenario shown in Eq. (1) with $\lambda = 0$. In addition, w^y is irrelevant here since data from only a single class is present. Purely supervised fixed-point EM expressions of Eq. (1) may be written in a concise form as $\boldsymbol{\theta}^y = EM_0(\boldsymbol{\theta}^y)$ (for texture class 'y'), where $\boldsymbol{\theta}^y$ and $EM_0(\boldsymbol{\theta}^y)$ represent the LHS and RHS of Eq. (1) for $\lambda = 0$. Analogous to Eq. (2), the supervised HMT expressions may be written in terms of the transformation function H as

$$H(\boldsymbol{\theta}^y, \lambda) = (1-\lambda)(\boldsymbol{\theta}^y - \boldsymbol{\theta}_0) + \lambda(\boldsymbol{\theta}^y - EM_0(\boldsymbol{\theta}^y)) = 0 \quad (4)$$

where $\boldsymbol{\theta}_0$ represents initialized HMT parameters. Note that λ used in Eq. (1) and (4) have different meanings since they are used for semi-supervised and supervised EM updates respectively. We obtain the partial derivatives of the fixed-point expressions ($\boldsymbol{\theta}^y = EM_0(\boldsymbol{\theta}^y)$) with respect to model parameters $\boldsymbol{\theta}^y$ and develop the Jacobian matrix J (see Eq. (3)), from which we obtain the direction and next set of parameter updates. Starting with ($\lambda = 0, \boldsymbol{\theta}^y = \boldsymbol{\theta}_0$), the homotopy function tracks the EM solution as it reaches $\lambda = 1$, corresponding to the purely supervised HMT model. A basic idea of the homotopy method for HMT parameter optimization is presented here without deliberating on details regarding partial derivatives of the fixed-point EM updates, which we present subsequently for the semi-supervised HMT modeling.

4.2. Semi-supervised HMT modeling

The iterative EM approach to semi-supervised learning of HMT (as shown in Eq. (1), with λ fixed at $U/(L + U)$) is often unreliable [4],[8]. As an alternative, we apply the homotopy method to obtain an optimal balance λ, along with the corresponding HMT model-parameters representing the textures. The fixed-point equations of a semi-supervised HMT (for any arbitrary λ), listed in Eq. (1), may be written in a concise form as

$$H(\tilde{\boldsymbol{\Theta}}, \lambda) = (1-\lambda)(\tilde{\boldsymbol{\Theta}} - EM_0(\tilde{\boldsymbol{\Theta}})) + \lambda(\tilde{\boldsymbol{\Theta}} - EM_1(\tilde{\boldsymbol{\Theta}})) = 0, \quad (5)$$

where $\tilde{\boldsymbol{\Theta}} = \{w^y, \tilde{\boldsymbol{\theta}}^y\}_{y=1}^C$ is an unnormalized version of $\boldsymbol{\Theta}$ (LHS of Eq. (1)). The expressions $EM_0(\tilde{\boldsymbol{\Theta}})$ and $EM_1(\tilde{\boldsymbol{\Theta}})$ represent the RHS of Eq. (1) for $\lambda = 0$ and 1, respectively (purely supervised and unsupervised EM updates). Note that the only difference between the above expression and the generic homotopy form (see Eq. (2)) is that the homotopy starts with a 'trivial' solution, whereas the above transformation function is itself a fixed-point EM (supervised EM) for $\lambda = 0$. We approximate the above expression as $EM_0(\tilde{\boldsymbol{\Theta}}) \approx EM_0$, where EM_0 is the supervised EM solution obtained using only labeled texture blocks. One may use either traditional EM or the homotopy-based solution proposed in Sec. 4.1 to obtain EM_0. The homotopy path is obtained by solving the differential equation (differentiating Eq. (5) to form the Jacobian J)

$$\underbrace{\left[\mathbf{I} - \lambda \nabla_{\tilde{\boldsymbol{\Theta}}} EM_1(\tilde{\boldsymbol{\Theta}}) \quad EM_0 - EM_1(\tilde{\boldsymbol{\Theta}})\right]}_{J} \left[\begin{array}{c} d\tilde{\boldsymbol{\Theta}} \\ d\lambda \end{array}\right] = 0 \quad (6)$$

Hence the direction $[d\tilde{\boldsymbol{\Theta}} \ d\lambda]^T$ can be obtained, under first-order approximation, as the one-dimensional null-space of the Jacobian matrix J. Note that evaluation of $\nabla_{\tilde{\boldsymbol{\Theta}}} EM_1(\tilde{\boldsymbol{\Theta}})$ involves partial derivatives of γ, ξ and $p(y|\boldsymbol{x})$ with respect to model parameters $\{w^y, \tilde{\pi}^y, \tilde{\epsilon}^y, \boldsymbol{\mu}^y, \boldsymbol{\sigma}^y\}$. Analytical derivations of these derivatives are elaborate; we only present the final expressions here for brevity:

$$\frac{\partial \gamma_{t,k}^j}{\partial \pi_m} = \frac{\gamma_{t,k}^j}{\pi_m}\left[\phi_{1,t,m,k}^j - \gamma_{1,m}^j\right]$$

$$\frac{\partial \gamma_{t,k}^j}{\partial \epsilon_{t',m,n}} = \frac{\gamma_{t,k}^j}{\epsilon_{t',m,n}}\left[\Phi_{t',t,m,n,k}^j - \xi_{t',m,n}^j\right]$$

$$\frac{\partial \gamma_{t,k}^j}{\partial \mu_{t'm}} = \gamma_{t,k}^j\left[\phi_{t',t,m,k}^j - \gamma_{t',m}^j\right](x_{t'j} - \mu_{t'm})/\sigma_{t'm}$$

$$\frac{\partial \gamma_{t,k}^j}{\partial \sigma_{t'm}} = \gamma_{t,k}^j\left[\phi_{t',t,m,k}^j - \gamma_{t',m}^j\right]\frac{(x_{t'j} - \mu_{t'm})^2 - \sigma_{t'm}}{2(\sigma_{t'm})^2},$$

where $\phi_{t,t',m,k} = p(s_t = m|s_{t'} = k, \boldsymbol{x}_j)$ and $\Phi_{t,t',m,n,k} = p(s_t = m, s_{parent(t)} = n|s_{t'} = k, \boldsymbol{x}_j)$. Note that the above set

of expressions is specific to a particular subband (HH,HL, or LH) of a particular texture class y, but these symbols are suppressed to avoid cluttering notation. One can derive analytical expressions for ϕ and Φ in terms of training data and HMT parameters, but these are not shown here due to space constraints. The terms γ and ξ can be evaluated directly using the upward-downward step of the HMT-EM algorithm [6]. Similarly one can show that

$$
\frac{\partial \xi_{t,k,l}^j}{\partial \pi_m} = \frac{\xi_{t,k,l}^j}{\pi_m} \left[\phi_{1,p(t),m,l} - \gamma_{1,m}^j \right]
$$

$$
\frac{\partial \xi_{t,k,l}^j}{\partial \epsilon_{t',m,n}} = \frac{\xi_{t,k,l}^j}{\epsilon_{t',m,n}} \left[p_{t,m,n,k,l}^{t',j} - \xi_{t',m,n}^j \right]
$$

$$
\frac{\partial \xi_{t,k,l}^j}{\partial \mu_{t'm}} = \xi_{t,k,l}^j \left[p_{t,m,k,l}^{t',j} - \gamma_{t',m}^j \right] \frac{(x_{t'j} - \mu_{t'm})}{\sigma_{t'm}}
$$

$$
\frac{\partial \xi_{t,k,l}^j}{\partial \sigma_{t'm}} = \xi_{t,k,l}^j \left[p_{t,m,k,l}^{t',j} - \gamma_{t',m}^j \right] \frac{(x_{t'j} - \mu_{t'm})^2 - \sigma_{t'm}}{2(\sigma_{t'm})^2},
$$

where $p_{t,m,n,k,l}^{t',j} = p(s_{t'} = m, s_{p(t')} = n | s_t = k, s_{p(t)} = l, \boldsymbol{x}_j)$ and $p_{t,m,k,l}^{t',j} = p(s_{t'} = m | s_t = k, s_{p(t)} = l, \boldsymbol{x}_j)$. The computations of $p_{t,m,n,k,l}^{t',j}$ and $p_{t,m,k,l}^{t',j}$ are straightforward but not shown here due to length constraints. Note that computation of the Jacobian matrix J (see Eq. (6)) requires partial derivatives with respect to unnormalized parameters $\tilde{\boldsymbol{\Theta}}$, which are related to their normalized counterpart as

$$
\pi_m^y = \frac{\tilde{\pi}_m^y}{\sum_{n=1}^M \tilde{\pi}_n^y} \text{ and } \epsilon_{t,p(t)}^y(m,n) = \frac{\tilde{\epsilon}_{t,p(t)}^y(m,n)}{\sum_{k=1}^M \tilde{\epsilon}_{t,p(t)}^y(k,n)},
$$

and the derivatives of an arbitrary function f with respect to unnormalized parameters may be written as

$$
\frac{\partial f}{\partial \tilde{\psi}_m} = \left(\frac{\partial f}{\partial \psi_m} - \sum_n \psi_n \frac{\partial f}{\partial \psi_n} \right) / \sum_{n=1}^M \tilde{\psi}_n, \quad \psi \in \{\pi, \epsilon\}.
$$

Now the only partial derivative remains to be specified is $p(y|\boldsymbol{x}^j)$ with respect to $\tilde{\boldsymbol{\Theta}}$, which can be shown as

$$
\frac{\partial p(y|\boldsymbol{x}_j)}{\partial w^{y'}} = p(y'|\boldsymbol{x}_j) \left[\delta_{y,y'} - p(y|\boldsymbol{x}_j) \right] / w^{y'} = W/w^{y'}
$$

$$
\frac{\partial p(y|\boldsymbol{x}_j)}{\partial \tilde{\pi}_m^{y'}} = W \left[\gamma_{1,m}^{j,y'} - \pi_m^{y'} \right] / \tilde{\pi}_m^{y'} \quad y, y' \in \{1, \cdots, C\}
$$

$$
\frac{\partial p(y|\boldsymbol{x}_j)}{\partial \tilde{\epsilon}_{t,m,n}^{y'}} = W \left[\xi_{t,m,n}^{j,y'} - \gamma_{p(t),n}^{j,y'} \epsilon_{t,m,n}^{y'} \right] / \tilde{\epsilon}_{t,m,n}^{y'} \tag{7}
$$

$$
\frac{\partial p(y|\boldsymbol{x}_j)}{\partial \mu_{tm}^{y'}} = W \left(x_{tj} - \mu_{tm}^{y'} \right) \gamma_{t,m}^{j,y'} / \sigma_{t,m}^{y'} \quad \text{and}
$$

$$
\frac{\partial p(y|\boldsymbol{x}_j)}{\partial \sigma_{tm}^{y'}} = W[(x_{tj} - \mu_{tm}^{y'})^2 - \sigma_{tm}^{y'}] \gamma_{t,m}^{j,y'} / 2(\sigma_{tm}^{y'})^2
$$

Given the partial derivative of $EM_1(\tilde{\boldsymbol{\Theta}})$ with respect to unnormalized parameters $\tilde{\boldsymbol{\Theta}}$, we obtain the Jacobian J, from which we obtain the directional vector $[\dot{\tilde{\boldsymbol{\Theta}}} \; \dot{\lambda}]^T$ and update the model parameters $\tilde{\boldsymbol{\Theta}}$ and balance parameter λ. Starting with $\dot{\tilde{\boldsymbol{\Theta}}} = 0$ and $\dot{\lambda} = 1$, we iterate the above procedure until we reach the first local optima in the path of λ. The corresponding $\boldsymbol{\Theta}$ (normalized from $\tilde{\boldsymbol{\Theta}}$) are treated subsequently as *trained* semi-supervised HMT parameters.

5. RESULTS

We present here a quantitative performance analysis of the proposed homotopy method for parameter estimation of both supervised and semi-supervised HMT models in the context of texture classification. Three different textures, *e.g.*, 'sand', 'grass' and 'wool' from publicly available "Brodatz" textural images [10] are chosen for subsequent analyses. Each texture of size 512×512 pixels is subjected to a two-level wavelet decomposition with Haar wavelets yielding 16384 wavelet quadtrees. Each two-level quadtree consisting of a root node and four children nodes (hence $R = 5$) corresponds to a 4×4 textural image block. For supervised HMT modeling, we randomly choose L samples from each texture and train the HMT model using both the EM and homotopy method. Note that the same random parameter initialization is used for both methods, and we observe that both algorithms converge to the same estimated HMT parameters. Figure 1 shows the convergence of expected log-likelihood for both
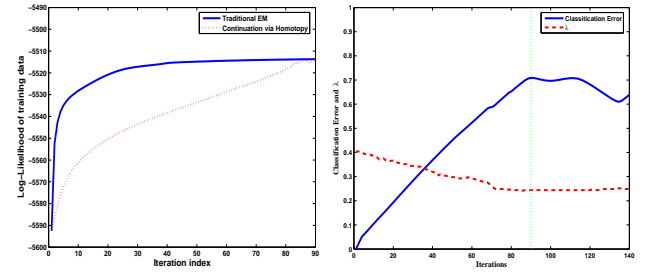

Fig. 1. Comparison of the fixed-point found by EM and homotopy method.

Fig. 2. Evolution of classification error and λ as a function of iteration index.

models as a function of iteration number (the explicit HMT parameters are also very close) for the 'sand' image with $L = 30$. For both algorithms, convergence is guaranteed (albeit to a local optimum), although the homotopy method is approximately 20% slower than its EM counterpart for the chosen example (although both algorithms are very fast). In addition, the rate of convergence for the homotopy method depends on the step size used for piecewise-linear approximation of the homotopy path, to be chosen specific to the nonlinear system of equations at hand. In our Gaussian mixture model-based HMT algorithm, we typically observe smooth convergence using a step size of 0.01 along the direction of λ. This comparison to conventional EM-based HMT training can also be viewed as a verification of accuracy for the homotopy method, and we therefore proceed to our true objective, *i.e.*, semi-supervised learning of HMTs.

We present the semi-supervised HMT performance analysis with $C = 2$ classes, although the method is applicable to an arbitrary number of texture classes (the computational cost increases exponentially with the number of training classes used). We randomly choose L labeled wavelet quadtrees (with known underlying texture) from two textures and U unlabeled ones (each unlabeled texture block or wavelet quadtree is chosen randomly from one of the two textures). We first obtain the HMT parameters trained only on the labeled samples and treat them as a starting point (EM_0 at $\lambda = 0$) for our semi-supervised HMT modeling via homotopy. In Fig. 2 we plot the evolution of λ starting at 0 and observe a sharp discontinuity around $\lambda = 0.7$ while training with 'sand' and ''grass' image with $L = 20$ (10 from each texture) and $U = 400$. According to [4], this kink represents a good operating point for λ within the semi-supervised classifier. In Fig. 2 we also plot the probabil-
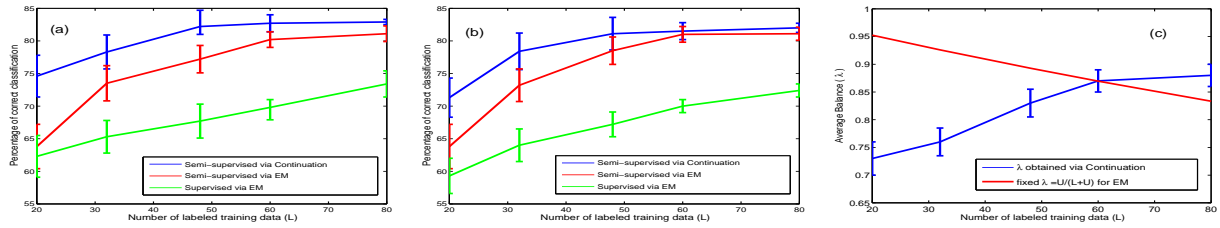
Fig. 3. Texture classification performance using homotopy, supervised-EM and semi-supervised EM method for a texture pair as a function of number of labeled data (L), with fixed number ($U = 400$) of unlabeled data.(a) performance for 'sand' vs 'grass' texture, (b) performance for 'sand' vs 'wood', (c) Variation of optimal balance (λ) averaged over 10 runs as a function of number of labeled data for 'sand' vs. 'grass'.

ity of classification error on the unlabeled data, as a function of the same evolution index (iteration number). It is important to emphasize that one would not have access to this curve in practice, since classifier scoring obviously requires access to the labels. However, it is interesting to note that the probability of error for the semi-supervised HMM is indeed minimized about the point for which the kink in λ is manifested. In Fig. 3(a) and 3(b) we summarize the performance of semi-supervised algorithms using both the EM and homotopy method using two sets of texture pairs, 'sand' vs. 'grass' and 'sand' vs. 'wood', respectively. The traditional EM algorithm uses a fixed balance ($\lambda = U/(L + U)$), whereas we estimate the optimal λ in the homotopy method that corresponds to the first sharp kink encountered in the homotopy path of λ. In both figures the percentage classification (averaged over 10 different runs) on a set of $U = 400$ unlabeled data is presented as a function of the number of labeled training samples (L, with $L/2$ samples from each texture). Note that we perform the classification task only on the U unlabeled image blocks. The variation in classification performance over ten different runs with fixed number of labeled and unlabeled data are shown using errorbars in both Figs. 3(a) and 3(b). Figure 3(c) shows the variation of λ, obtained via the homotopy method and averaged over different runs, along with its variation presented in errorbars. We observe superior classification performance using homotopy-based semi-supervised HMT modeling in comparison to its EM counterpart, while both of the semi-supervised algorithms are significantly superior to the supervised algorithm.

6. CONCLUSIONS

We apply the homotopy method to HMT model-parameter estimation in the context of wavelet-based texture analysis and classification, utilizing both labeled and unlabeled image blocks. The algorithm is focused on determining the proper balance between these two data sets, accounting for the fact that unlabeled data are typically far more plentiful than labeled data. The homotopy method starts from a purely supervised solution and then tracks in the parameter space until a phase transition is manifested. We observe in our experiments that the initial such transition is indicative of a good balance between the labeled and unlabeled data. We also used the homotopy method as an alternative to the widely used EM algorithm for supervised HMT modeling. Results are presented for real textural images from a publicly available image database [10].

The major contributions of this paper are: 1) development of the homotopy method for purely supervised training of the HMT model as an alternative to the EM algorithm, and 2) development of a semi-supervised HMT model, with automatic estimation of the proper balance between labeled and unlabeled data. There are sev-

eral areas of interest for future research. Computation cost of the homotopy based semi-supervised training is much higher than its EM counterpart, albeit significant performance gain. In addition, no incremental learning procedure is available for new unlabeled data. Using 20 labeled and 400 unlabeled data, the supervised-EM, semi-supervised EM, and the homotopy method take approximately 10 seconds, 2 minutes and 100 minutes of CPU time, respectively using Matlab on a 2.8 GHz Pentium machine. The complexity of homotopy is $O(n^3)$ (n being the the size of training data), which is too slow for many applications.

7. REFERENCES

[1] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[2] M. Seeger, "Learning with labeled and unlabeled data," Tech. Rep., Institute for Adaptive and Neural Computation, University of Edinburgh, Edinburgh, UK, 2000.

[3] K. Nigam, A. Mccallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using em," *Machine Learning*, vol. 39, pp. 135–167, 2000.

[4] A. Corduneanu and T. Jaakkola, "Continuation methods for mixing heterogeneous sources," in *Proceedings of the Eighteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 2002.

[5] S. N. Chow, J. Mallet-Paret, and J. A. Yorke, "Finding zeros of maps: Homotopy methods that are constructive with probability one," *Math. Comput.*, vol. 32, pp. 887–899, 1978.

[6] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. on Sig. Proc.*, vol. 46, no. 4, pp. 886–902, 1998.

[7] N. Dasgupta and L. Carin, "Context-based graphical modeling for wavelet domain signal processing," in *ICASSP*, 2003, vol. 3, pp. 485–488.

[8] M. Inoue and N. Ueda, "Exploitation of unlabeled sequences in hidden markov models," *IEEE Trans. on PAMI*, vol. 25, no. 12, pp. 1570–1581, 2003.

[9] L. T. Watson, "Theory of globally convergent probability-one homotopies for nonlinear programming," *Siam J. Optim.*, vol. 11, no. 3, pp. 761–780, 2000.

[10] "University of southern california image database," http://sipi.usc.edu.