Research Article

Models for Patch-Based Image Restoration

Mithun Das Gupta,[1] Shyamsundar Rajaram,[1] Nemanja Petrovic,[2] and Thomas S. Huang[1]

[1] *Beckman Institute, Department of Electrical and Computer Engineering (ECE),*
  University of Illinois at Urbana-Champaign (UIUC), IL 61801, USA
[2] *Google Inc., NY 10011, USA*

Correspondence should be addressed to Mithun Das Gupta, mdgupta@uiuc.edu

We present a supervised learning approach for object-category specific restoration, recognition, and segmentation of images which are blurred using an unknown kernel. The novelty of this work is a multilayer graphical model which unifies the low-level vision task of restoration and the high-level vision task of recognition in a cooperative framework. The graphical model is an interconnected two-layer Markov random field. The restoration layer accounts for the compatibility between sharp and blurred images and models the association between adjacent patches in the sharp image. The recognition layer encodes the entity class and its location in the underlying scene. The potentials are represented using nonparametric kernel densities and are learnt from training data. Inference is performed using nonparametric belief propagation. Experiments demonstrate the effectiveness of our model for the restoration and recognition of blurred license plates as well as face images.

1. Introduction

Restoration is a neat showcase of ill-posedness of computer vision. Given a blurred image, there can be several sharp natural images which, when blurred, will generate the original image. The inherent ambiguities in restoration are usually overcome by using different regularization techniques or the Bayesian remedy. In several important applications like surveillance, tracking, and license plate recognition systems, images are mostly severely blurred. But the quality of the blur is usually application specific and hence the need for restoration systems.

There are numerous methods for inferring the sharp image from the blurred input. A reasonable estimate of the high-resolution image may be obtained if we have a priori knowledge about the blurring kernel. If no additive noise is present, Wiener filter is the optimal filter. In the noisy case, Wiener filter gives the MMSE solution. Restoration can be made easier by incorporating several images as in [1]. Further, image restoration can be thought of as a special case of superresolution and as such image deblurring and superresolution have been treated concurrently by many authors. Superresolution algorithms can be classified into many categories based on different criteria such as frequency/image domain, single/multiple images, and so on. Earlier works in this field utilized the band limitedness of the images to interpolate subpixel values from a series of aliased images. Recently, time domain methods have been the principal research fields. Among the time domain methods, two broad sections are iterative methods and learning-based methods. Iterative methods [2–6] mostly use a Bayesian framework, where an initial guess about the high-resolution frame is refined at each iteration. The image prior is usually assumed to be a smoothness prior.

However, it seems that machine learning and specifically, probabilistic inference techniques are currently the most promising line of research. The principal idea of the machine learning approach is to use a set of high-resolution (sharp) images and their corresponding low-resolution (blurred) images to build a compatibility model. The images are stored as patches or as coefficients of other feature representations. Recently, an impressive amount of work has been reported in this field [7–11], to name a few. In [11], PCA-based techniques were used to capture the relationship between the high-resolution and low-resolution patches while nonparametric modeling was used to estimate the missing details.
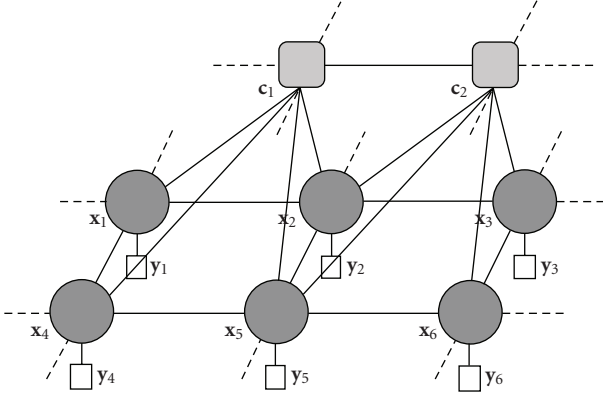
FIGURE 1: Multilayer MRF model. The restoration layer consists of nonoverlapping hidden image patches indicated by dark gray circles and each node has a local observation indicated by empty boxes. The recognition layer consists of class membership nodes and are denoted by light gray boxes.

FIGURE 2: The neighborhood model. The node i is the black node. The dark gray nodes are the neighbors which belong to the top layer defined by E_{ti} and the light gray nodes are the neighbors in the bottom layer defined by E_{bi}.

In [9], an example-based learning method was employed for superresolving natural images up to a zoom factor of 8. Along the same lines, Bishop et al. [10] performed video superresolution by considering additional priors to take into account the temporal coherence between successive frames. Machine learning-based restoration methods can be made more powerful and robust if the images are restricted to be of the same type, as in [7], where face images are hallucinated. The spirit of our work is similar to the work of Freeman et al. [9], with several important differences.

In our previous work [12, 13] on restoration using MRFs over a patch image model, we introduced the ideas of partial messages and the restoration-recognition loop. Our algorithm was built on the notion of *partial message propagation,* where any given node (patch) in an MRF is only partially influenced by its neighbor, depending on the spatial orientation of the two neighbors. Restoration and recognition was performed in an iterative loop which resulted in the localization of the search space.

The limitation of our previous work is the need for segmented images before performing restoration and recognition. Further, the recognition was performed as a separate block. In this work, we suggest a unified framework for performing restoration and recognition without prior segmentation. We present a multilayer Markov random field model having two interconnected layers where restoration is performed in the bottom layer and recognition in the top layer. The messages propagated bottom-up help in improving recognition/segmentation and the messages propagated top-down aid in better restoration.

This paper is organized in the following order. In Section 2, we present the image model and define the problem statement and the notations used in the rest of the paper. In Section 3, we illustrate the features used for learning the potential functions for the multilayer MRF structure. Within this section, we briefly review the belief propagation algorithm and the nonparametric extens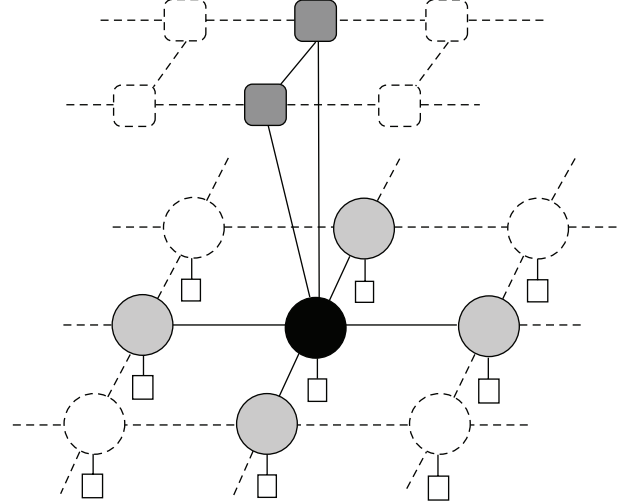ion to it. We illustrate the use of multilayer MRF structure for two different applications and explain the recognition with respect to the two different applications. We propose an analogous model in the frequency domain and show improvement in handling compression artifacts in images and videos in Section 5. Finally, we present experiments and results in Section 6 and conclude in Section 7.

2. Model

We propose a multilayer architecture to perform restoration and recognition of blurred images (Figure 1). The lower layer (restoration layer) is an undirected graphical model over image patches with compatibility functions represented as nonparametric kernel densities. The top layer (recognition layer) is an undirected graphical model with each node representing a multinomial distribution over the 10-hypothesized digits present in the neighboring patch nodes in the lower layer. The compatibility functions in the restoration layer are learned using nonparametric kernel density estimation techniques (Section 3). We use an extended version of the nonparametric belief propagation algorithm [14, 15] in the restoration layer, and belief propagation algorithm [16] in the recognition layer.

2.1. Problem Statement and Notation. Consider a training set of triplets given by $\{(\mathbf{X}^1, \mathbf{Y}^1, W^1), (\mathbf{X}^2, \mathbf{Y}^2, W^2), \ldots, (\mathbf{X}^n, \mathbf{Y}^n, W^n)\}$, where $\mathbf{X}^i$ represents a sharp image of one particular digit, $\mathbf{Y}^i$ represents the blurred version, $W^i \in \{0, 1, \ldots, K\}$ indicates the class of the ith training image and K denotes the number of classes. Let there be an unknown kernel $f(\mathbf{X}^i)$ that maps from $\mathbf{X}^i$ to $\mathbf{Y}^i$. The objective is to learn a model which is able to infer the sharp image $\mathbf{X}$ (which has more than one digit) and the classes of all m objects $W_1, \ldots, W_m$ (digits) in it, from the blurred input image $\mathbf{Y}$ consisting of digit fonts not present in the training set.

Interacting nodes	Potential	Notation
$(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$	Patch interaction	$\psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$
$(\mathbf{x}_{i_b}, \mathbf{y}_{i_b})$	Association	$\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b})$
$(\mathbf{x}_{i_b}, c_{j_t})$	Classification	$\varphi(\mathbf{x}_{i_b}, c_{j_t})$
(c_{i_t}, c_{j_t})	Class interaction	$\theta(c_{i_t}, c_{j_t})$

We model the image $\mathbf{X}$ as an undirected graphical model or more specifically a Markov random field (MRF) [9]. This MRF will be referred to as the restoration field and is the bottom layer in Figure 1. The restoration field is defined by the bottom graph $G_b = \{V_b, E_b\}$, where each node represents a random variable $\mathbf{x}_{i_b}$, $i_b \in [1, \ldots, N]$ (where N is the total number of nodes in the restoration layer), corresponding to a patch in the unknown, sharp image, which is associated with an observation node $\mathbf{y}_{i_b}$ which represents the corresponding patch in the observed image. An edge between node $\mathbf{x}_{i_b}$ and node $\mathbf{x}_{j_b}$ indicates that they are spatial neighbors.

We model the class membership of the m objects $W_1, \ldots, W_m$ as another MRF defined by the top graph $G_t = \{V_t, E_t\}$. We refer to this MRF as the recognition field and is shown as the top layer in Figure 1. Each node in this MRF represents a random variable c_{i_t}, $i_t \in [1, \ldots, L]$, where $c_{i_t} \in \{1, \ldots, K\}$ and L is the total number of nodes in the recognition layer. We note that G_t and G_b denote the graph structure corresponding to the top layer and bottom layer, respectively, and they do not include the interlayer connectivity information. Further, for every node i in our two-layer graphical model, we denote the neighborhood in the restoration layer by E_{bi} and similarly we denote E_{ti} as the neighborhood of node i in the recognition layer. The model is illustrated in Figure 2.

The potential functions modeling the various interactions are summarized in Table 1.

The association and the patch interaction potentials can be modeled as continuous parametric distributions that are tractable for the belief propagation algorithm [16] (mixture of Gaussian, etc.). However, as noted in our previous work [17], parametric approaches introduce averaging effects which are against the spirit of the restoration problem. Hence, we use nonparametric kernel density estimation techniques for learning the association, patch interaction, and classification potentials. We elaborate the learning of these potentials in Section 3.

During the learning phase, the potentials ϕ, ψ, θ, φ are learned from the training data. The inference phase involves computing the marginals of posterior distribution $p(\mathbf{x}_{i_b} \mid \mathbf{Y})$, for all the nodes $i_b \in V_b$ and the marginals $p(c_{j_t} \mid \mathbf{Y})$, for all the nodes $j_t \in V_t$. In Section 3.1, we discuss the application of the belief propagation algorithm for approximate inference in the multilayer MRF model.

3. Learning Potentials

We model the association potential $\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b})$ as a function over the vectorized representation of patches $\mathbf{x}_{i_b}$ and $\mathbf{y}_{i_b}$ and

has the form

$$\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b}) = \frac{1}{M} \sum_{m=1}^{M} \mathcal{N}\left([\mathbf{x}_{i_b}, \mathbf{y}_{i_b}]^T; \boldsymbol{\mu}_m, \boldsymbol{\Lambda}_m\right), \qquad (1)$$

where M is the number of components and $\mathcal{N}([\mathbf{x}_{i_b}, \mathbf{y}_{i_b}]^T; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$ over the random vector $[\mathbf{x}_{i_b}, \mathbf{y}_{i_b}]^T$. From the training set, the patch association vectors corresponding to the image and its blurred version are constructed. The patch association vectors are pruned to avoid redundancy. The potential is constructed by considering a Gaussian kernel with the mean chosen as the patch association vector and the covariances are chosen using the leave-one-out cross validation technique [18].

The patch interaction potential $\psi_{\zeta(i_b, j_b)}(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$ is a function over the vectorized two-pixel thick nonoverlapping patch boundary and learnt using the above-mentioned nonparametric estimation technique. The classification potential $\varphi(\mathbf{x}_{i_b}, c_{j_t})$ is split into a conditional term $\varphi(\mathbf{x}_{i_b} \mid c_{j_t})$ and a marginal term $\varsigma(c_{j_t})$. The conditional term is nonparametrically represented as a function over vectorized sharp patches obtained from images belonging to the class c_{j_t} and the marginal term $\varsigma(c_{j_t})$ is given by $1/K$. The class interaction potential $\theta(c_{i_t}, c_{j_t})$ is modeled as a probability table given by

$$[\theta(c_{i_t}, c_{j_t})]_{kl} = p(c_{i_t} = k, \, c_{j_t} = l) = \begin{cases} q, & \text{if } k = l, \\ \dfrac{1 - q}{K - 1}, & \text{if } k \neq l, \end{cases} \qquad (2)$$

where K is the number of classes and we set q to be 0.7 for our experiments.

3.1. Belief Propagation.
For acyclic graphs, the marginal distributions can be calculated efficiently and exactly by a local message passing algorithm known as belief propagation (BP) [16]. In the case of graphs with cycles, the BP algorithm is not exact. The iterative version of BP algorithm produces beliefs which do not converge to true marginals. But, it has been empirically shown that loopy BP produces excellent results for several hard problems. Recently, Yedidia et al. [19] established the link between the fixed points of belief propagation algorithm and stationary points of the "variational free energy" defined on the graphical model. This important result sheds more light on the properties of loopy BP approximation.

In our multilayer MRF model, BP is first performed in the restoration layer which is followed by BP in the recognition layer. There are four different types of messages, namely, the message passed between neighboring nodes $(i_b, j_b) \in E_b$ in the restoration layer, the messages passed between neighboring nodes $(i_t, j_t) \in E_t$ in the recognition layer, and the bidirectional messages passed between neighboring nodes $(i_b, j_t) \in E_l$ in the recognition and restoration layer interface.
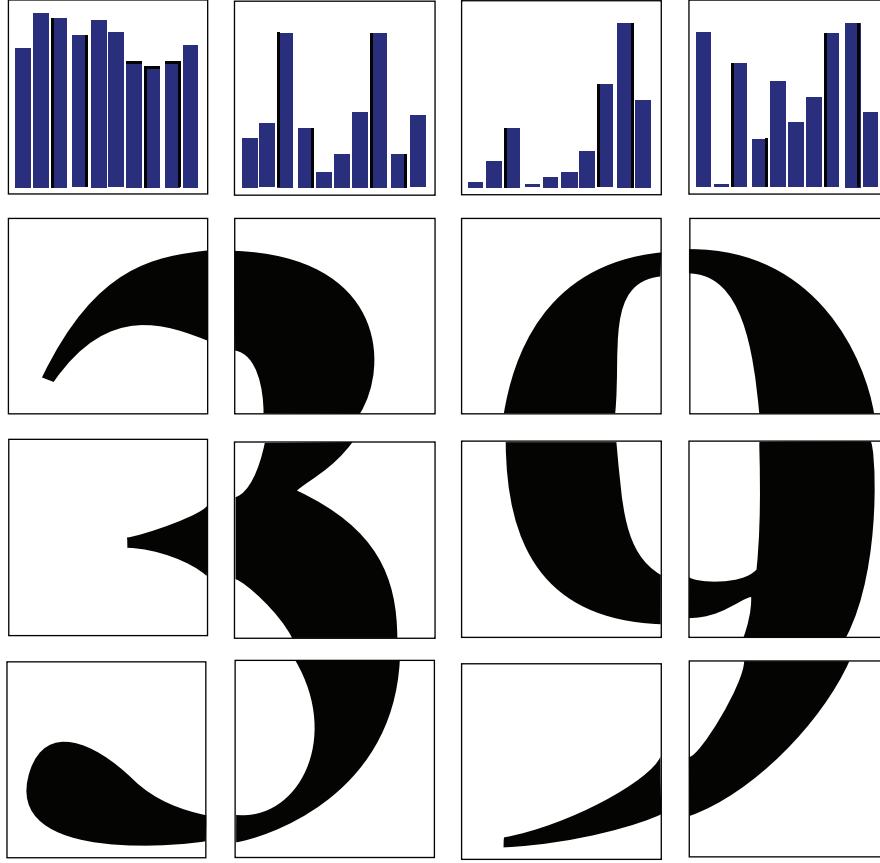
FIGURE 3: The restoration nodes $\mathbf{x}$ converge to the digits. The recognition layer multinomials peak at the digit locations.
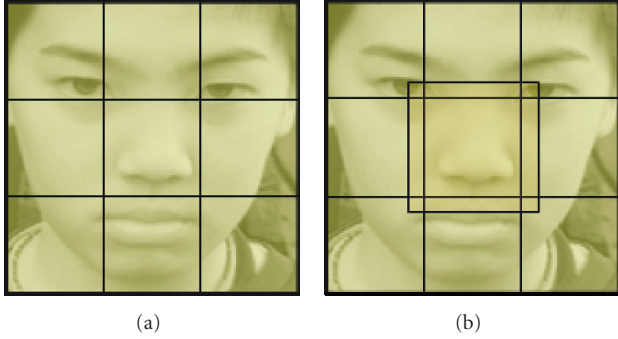

| (a) | (b) |

FIGURE 4: The division of a face into regions for better restoration. Overlapping nodes ensure better interaction for the boundary patches.

Recognition-to-Restoration Layer Messages. The message propagated from a node in the recognition layer c_{i_t} to a node in the restoration layer $\mathbf{x}_{j_b}$ is given by

$$m_{i_t, j_b}(\mathbf{x}_{j_b}) = \alpha \sum_{c_{i_t}} \varphi(\mathbf{x}_{j_b}, c_{i_t}) \prod_{k_t \in E_{ti}} m_{k_t, i_t}(c_{i_t}) \prod_{r_b \in E_{bi}, r_b \neq j_b} m_{r_b, i_t}(c_{i_t}). \tag{3}$$

Intrarestoration Layer Messages. The message propagated from node i_b to node j_b in the restoration layer during the nth iteration represented as $m_{i_b, j_b}^n(\mathbf{x}_{j_b})$ is given by

$$m_{i_b, j_b}^n(\mathbf{x}_{j_b}) = \alpha \int \psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b}) \phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b}) \prod_{k_t \in E_{ti}} m_{k_t, i_b}^{n-1}(\mathbf{x}_{i_b}) \\ \times \prod_{h_b \in E_{bi}, h_b \neq j_b} m_{h_b, i_b}^{n-1}(\mathbf{x}_{i_b}) d\mathbf{x}_{i_b}. \tag{4}$$

Restoration-to-Recognition Layer Messages. The message propagated from node i_b to j_t, $m_{i_b, j_t}(c_{j_t})$ is

$$m_{i_b, j_t}^n(c_{j_t}) = \alpha \int \varphi(\mathbf{x}_{i_b}, c_{j_t}) \phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b}) \prod_{k_b \in E_{bi}} m_{k_b, i_b}^{n-1}(\mathbf{x}_{i_b}) \\ \times \prod_{r_t \in E_{ti}, r_t \neq j_t} m_{r_t, i_b}^{n-1}(\mathbf{x}_{i_b}) d\mathbf{x}_{i_b}. \tag{5}$$

Intrarecognition Layer Messages. The message propagated from node i_t to node j_t, during the nth iteration represented as $m_{i_t, j_t}^n(c_{j_t})$ is given by

$$m_{i_t, j_t}^n(c_{j_t}) = \alpha \sum_{c_{i_t}} \theta(c_{i_t}, c_{j_t}) \prod_{k_b \in E_{bi}} m_{k_b, i_t}^{n-1}(c_{i_t}) \prod_{r_t \in E_{ti}, r_t \neq j_t} m_{r_t, i_t}^{n-1}(c_{i_t}). \tag{6}$$
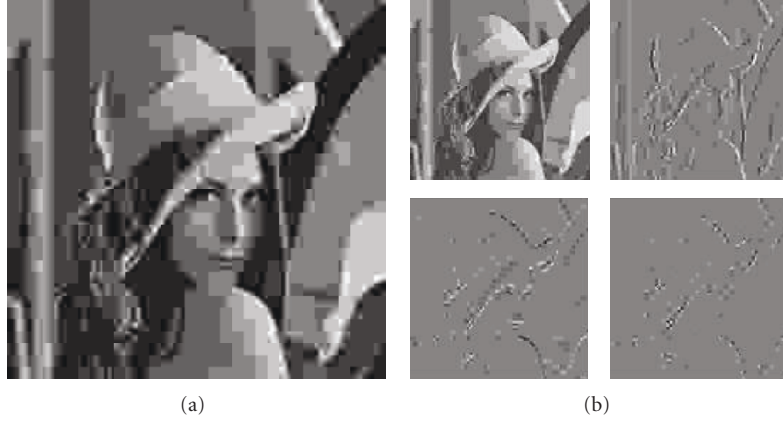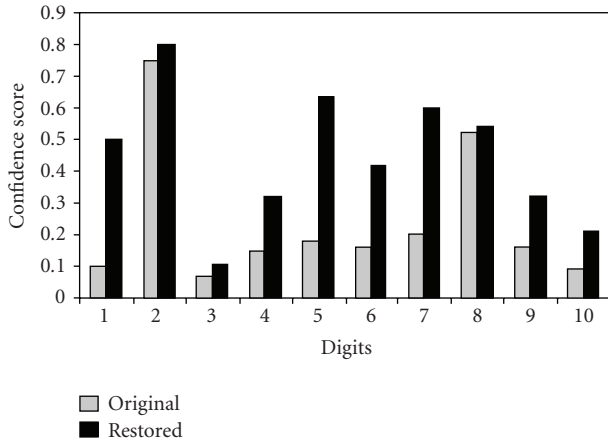
(a)　　　　　　　　　(b)

FIGURE 5: Subband representation of the 2×2 block DCT coefficients.

FIGURE 6: Confidence scores versus true digit class, before and after restoration.

3.2. Nonparametric Belief Propagation (NBP).

We note that the messages computed using (4) are mixtures of Gaussians and computing a message $m_{i_b, j_b}^k(\mathbf{x}_{j_b})$ involves the product of the interaction potential $\psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$, the association potentials $\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b})$, the messages from the neighbors in the recognition layer $m_{k_t, i_b}^{n-1}(\mathbf{x}_{i_b})$ for all $k_t \in E_t i$, and the messages from the neighbors in the restoration layer $m_{h_b, i_b}^{n-1}(\mathbf{x}_{i_b})$ for all $h_b \in E_b i$, $h_b \neq j_b$, where each term is a mixture of Gaussians. In order to evaluate (4), the mixture components in the potentials and the messages have to be pruned drastically so that the number of components in the product is within tractable limits to solve the integral. Such an approximation is unsuitable for the restoration problem and alternatively we use the nonparametric extension proposed by Sudderth et al. [15] and independently invented by Isard [14].

The interaction potential can be decomposed into a marginal influence term given by $\xi(\mathbf{x}_{i_b}) := \int \psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b}) d\mathbf{x}_{j_b}$ and an interaction term $\psi(\mathbf{x}_{j_b} | \mathbf{x}_{i_b})$. The message update equation (4) can be solved in two phases. The first phase involves computing the product $\pi_{i_b, j_b}^n(\mathbf{x}_{i_b})$:

$$\pi_{i_b, j_b}^n(\mathbf{x}_{i_b}) := \xi(\mathbf{x}_{i_b}) \prod_{h_t \in E_{ti}} \phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b}) \prod_{h_b \in E_{bi}, h_b \neq j_b} m_{h_b, i_b}^{n-1}(\mathbf{x}_{i_b})$$
$$\times \prod_{k_t \in E_{ti}} m_{k_t, i_b}^{n-1}(\mathbf{x}_{i_b}). \tag{7}$$

Each term in the product $\boldsymbol{\pi}_{i_b, j_b}^n(\mathbf{x}_{i_b})$ is a mixture of Gaussians and if each term has M mixture components, then the product is a mixture of M^L Gaussians, where L is the number of terms. Exact computation of the product can be performed, however, it is not feasible because of the $\mathcal{O}(M^L)$ computations. Pruning of the mixture components can be performed to restrict the number of computations, but it turns out to be a very coarse approximation for the restoration problem. Sequential Gibbs sampling [20] and importance weighting were used in [14, 15] to generate M asymptotically unbiased samples without explicitly computing the product. In this work, we use *alternating Gibbs sampling (parallel sampling)* [21] to obtain asymptotically unbiased samples $\mathbf{x}_{i_b}^1, \mathbf{x}_{i_b}^2, \ldots, \mathbf{x}_{i_b}^M$ from the product $\boldsymbol{\pi}_{i_b, j_b}^n(\mathbf{x}_{i_b})$. The second phase involves integrating the combination of the above product $\boldsymbol{\pi}_{i_b, j_b}^n(\mathbf{x}_{i_b})$ with the interaction term. Sudderth et al. [15] and Isard [14] proposed Gibbs sampling to solve the first phase and handled the second phase using stochastic integration and further, represented the messages nonparametrically using a kernel density estimate as

$$m_{i_b, j_b}(\mathbf{x}_{j_b}) = \sum_{m=1}^M w_{j_b}^m \mathcal{N}(\mathbf{x}_{j_b}; \boldsymbol{\mu}_{j_b}^m, \boldsymbol{\Lambda}_{j_b}^m), \tag{8}$$

where $w_{j_b}^m$, $\boldsymbol{\mu}_{j_b}^m$, $\boldsymbol{\Lambda}_{j_b}^m$ correspond to the weight, mean, and covariance associated with the mth kernel. The message update is performed using stochastic integration, where every sample $\mathbf{x}_{i_b}^m$ is propagated to node j_b by sampling $\mathbf{x}_{j_b}^m$ from the interaction potential $\psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$. Now, nonparametric density estimation is used to obtain a kernel density estimate (8) for the message $m_{i_b, j_b}^n(\mathbf{x}_{j_b})$, where the means of the kernels are the propagated samples. Covariances are chosen to be diagonal and identical and are obtained using leave one outcross validation [18].
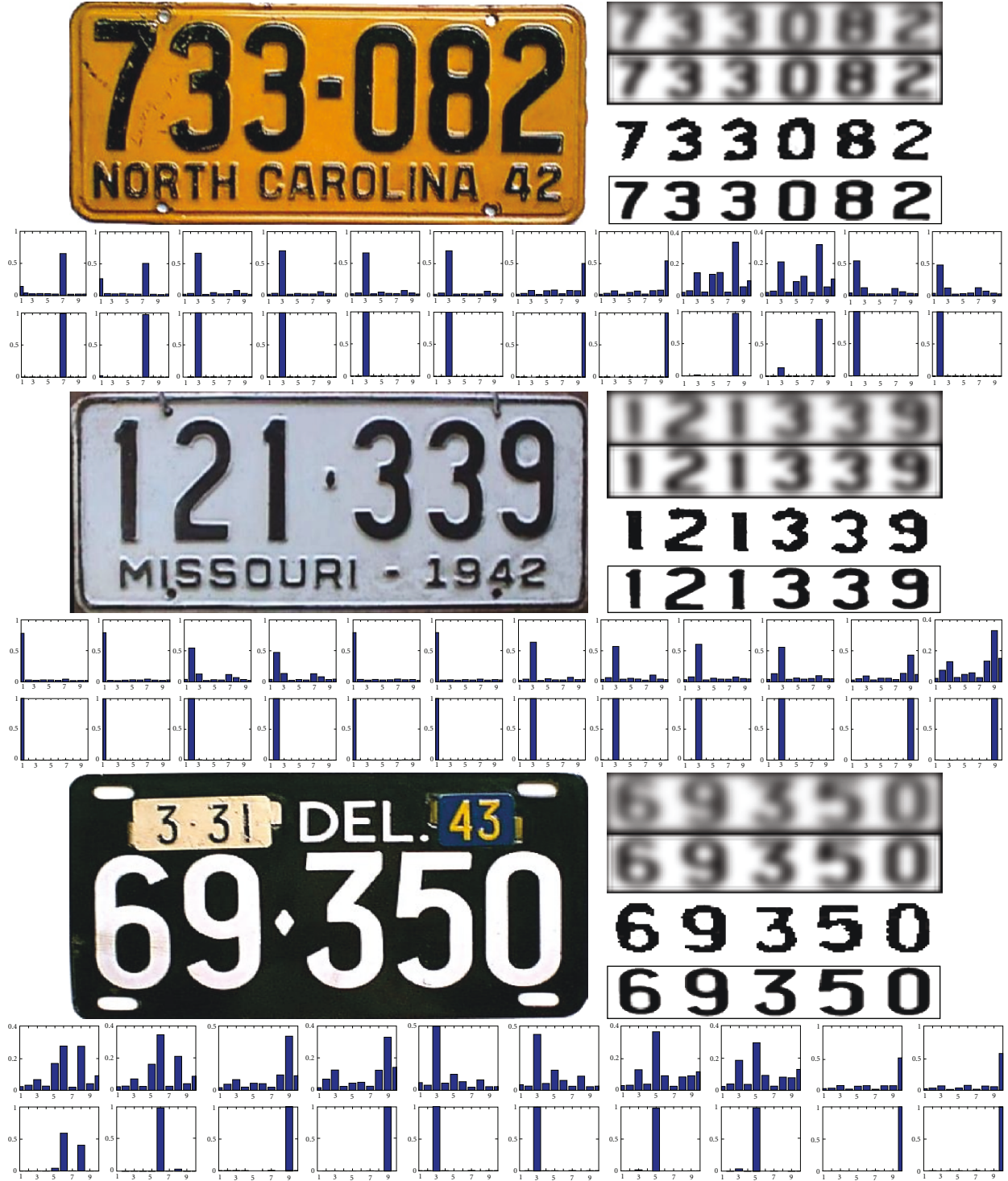
FIGURE 7: Recognition and restoration results: (top Left) original license plate; (top right top to bottom) blurred input, deblurred using deconvolution methods, result of our algorithm, actual sharp image. (Bottom left to right) multinomial distributions (X-axis: digit class, Y-axis: confidence score) over the nodes in the recognition layer. The top row is the result after 3 iterations and the bottom row is the result after 6 iterations.

3.3. Data-Driven Local Retraining.

Nonparametric belief propagation can be used to approximately evaluate (4), but, however, we note that the first two terms in (4), $\psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$ and $\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b})$, are nonparametric densities each having more than 10^6 components. It is computationally infeasible to compute $m_{i_b,j_b}^n(\mathbf{x}_{j_b})$ and, hence, we need a technique for pruning the nonparametric densities to be computationally reasonable and meaningful. We adopt
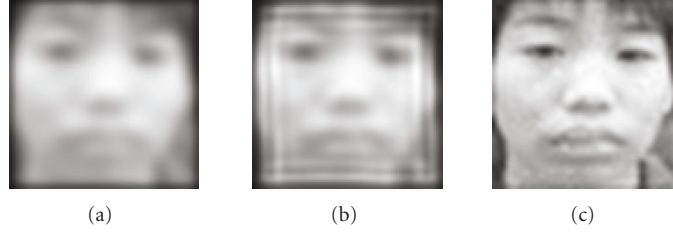
(a) (b) (c)

FIGURE 8: Multilayer MRF-based face superresolution. (Left) Blurred input; (center) deblurred using deconvolution methods; (right) our method.

FIGURE 9: Comparative results. (Top row from left to right) LENA, a part of JPEG-coded LENA at 0.188 bpp, result of the proposed method, and Zou and Yan result [22]. (Bottom row) Yang and Galatsanos result [23], Paek et al. result [24], Park and Kim result [25], Jeong et al. result [26], and MPEG-4 result [27].

a data-driven retraining step in which the potentials $\psi(\mathbf{x}_{i_b}, \mathbf{x}_{j_b})$ and $\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b})$ are learned from a subset sampled from the training set based on the posterior marginal $p(c_{k_t}|\mathbf{Y})$ at the recognition layer node c_{k_t}, where k_t indicates the common neighbor of i_b and j_b in the recognition layer.

3.4. Multilayer Message Passing.
Belief propagation for our two-layer model is given by the following schedule.

(1) Intrarestoration layer messages given by (4) are computed using nonparametric belief propagation and the data-driven local retraining algorithm.

(2) Restoration-to-recognition layer message is approximately evaluated at the maximum posterior marginal estimate $\mathrm{argmax}_{\mathbf{x}_{i_b}} p(\mathbf{x}_{i_b}|\mathbf{Y})$ owing to computational difficulties involved in evaluating the exact integral.

(3) Intrarecognition layer messages are evaluated using (6).

(4) Recognition-to-restoration layer messages are computed using (3) and the resulting nonparametric message is pruned based on the weight of the components.

The above steps are performed iteratively till convergence.

3.5. Restoration and Recognition.
After convergence of the message passing algorithm, the marginal distributions $p(\mathbf{x}_{i_b}|\mathbf{Y})$ and $p(c_{i_t}|\mathbf{Y})$ can be evaluated using

$$
\begin{aligned}
p(\mathbf{x}_{i_b}|\mathbf{Y}) &\approx \alpha\phi(\mathbf{x}_{i_b}, \mathbf{y}_{i_b}) \prod_{k_b \in E_{bi}} m^{n-1}_{k_b, i_b}(\mathbf{x}_{i_b}) \prod_{r_t \in E_{ti}} m^{n-1}_{r_t, i_b}(\mathbf{x}_{i_b}), \\
p(c_{i_t}|\mathbf{Y}) &\approx \alpha \prod_{k_t \in E_{ti}} m_{k_t, i_t}(c_{i_t}) \prod_{r_b \in E_{bi}} m_{r_b, i_t}(c_{i_t}).
\end{aligned}
\tag{9}
$$

The restored image is obtained by estimating the maximum posterior marginal (MPM) estimate $\mathrm{argmax}_{\mathbf{x}_{i_b}} p(\mathbf{x}_{i_b}|\mathbf{Y})$ at all nodes i_b in the restoration layer and similarly recognition
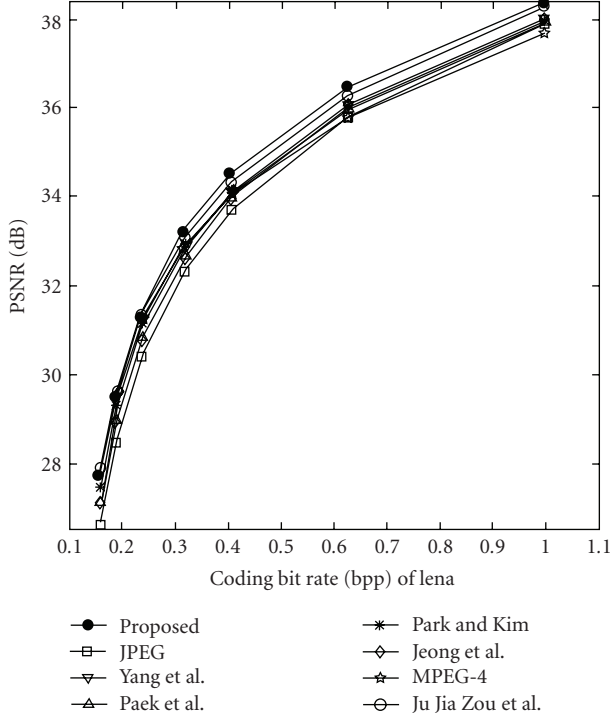
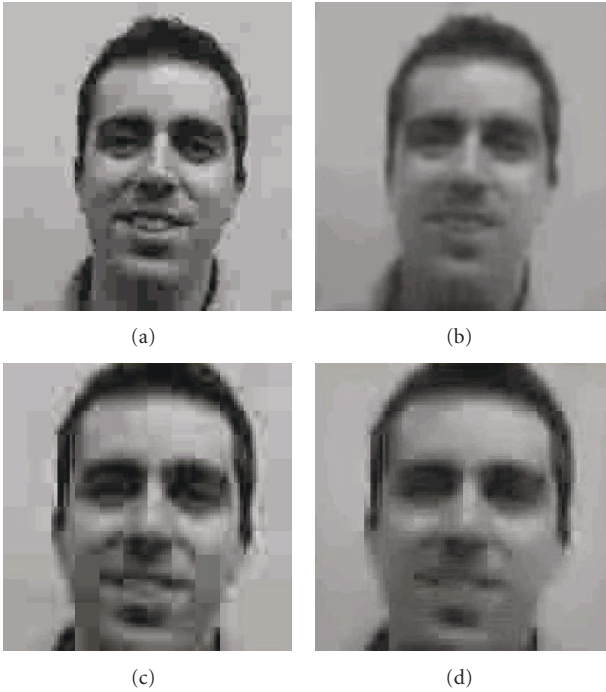FIGURE 10: PSNR values for the methods mentioned in Figure 9.

FIGURE 12: (Top) Compressed image; (bottom) reconstruction using our method (2×2 BDCT).

4. Recognition Layer

The recognition layer is incorporated into the model to aid the restoration layer. The basic need for this layer arises from the fact that the training dataset for any practical application is extremely big and we need an efficient way of subdividing this set for searching through this space of training examples.

One important feature of this kind of model is that it does not expect the nodes as well as their interactions to be modeled in any particular form. We investigate two different applications and two different representations of the recognition layer as well as the interactions in this layer.

4.1. License Plate Recognition. For this class of applications, the recognition layer nodes can be represented by a multinomial distribution over the 10 digits. In essence, this layer captures the digit class to which the restoration nodes ($\mathbf{x}'_\mathbf{i}$s) belong to. For example, let us consider a case where we are trying to restore an image of 39. Let us assume we have 3 rows of $\mathbf{x}$ nodes under each $\mathbf{z}$ node. After a few iterations of our NBP algorithm for the multilayered MRF, the nodes in the two layers converge to the states as shown in Figure 3.

4.2. Face Image Restoration. We note that face images have a spatial coherence which can be exploited for recognition. If the face is divided into 4 quadrants with the nose tip as the center, then we can expect the eyes to be in 1st and 2nd quadrants and the cheeks to be in the other two quadrants. This information can be incorporated in our restoration algorithm and we can search for patches only in the first two quadrants while restoring the eye regions (Figure 4).

FIGURE 11: (Top row, left) Compressed image; (right) reconstruction using Li and Delp [28] method (8×8 BDCT); (bottom row, left) compressed image; (right) reconstruction using Li-Delp method (2×2 BDCT).

results are obtained by evaluating the MPM estimate $\arg\max_{c_{i_t}} p(c_{i_t} | \mathbf{Y})$ at all nodes i_t in the recognition layer.
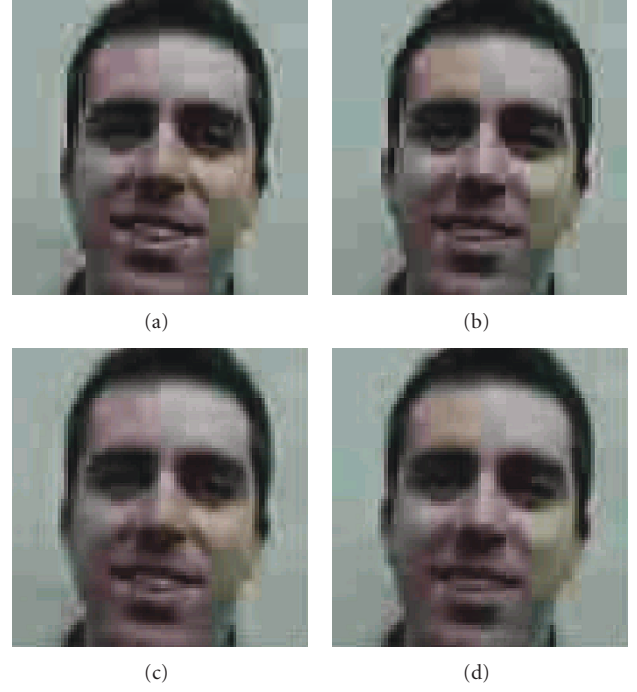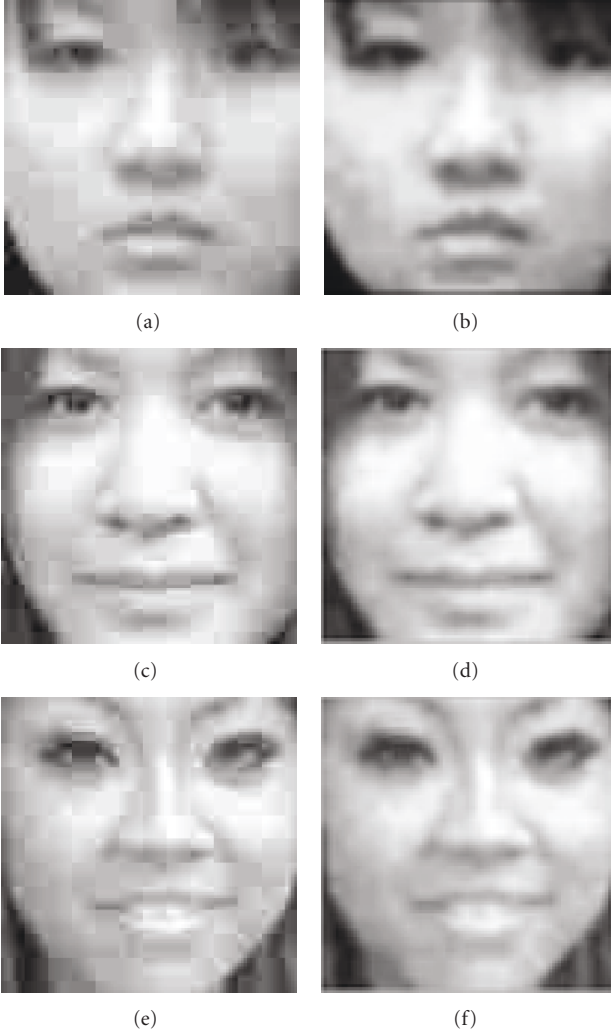
FIGURE 13: (Left) Compressed image; (right) reconstruction using our method (2×2 BDCT).

The recognition layer in this model is just the index of the grid structure imposed over the face image. The interactions for the nodes in this layer are captured in the overlapped structure of the grids.

5. Compression Artifacts Removal by Frequency Domain MRF Model

We propose a new progressive and fast approach based on the transform domain MRF model. The $L \times L$ DCT transform is interpreted as an L^2-subband transform. The transform coefficients at the same subband are modeled as an image. This decomposition is similar to the wavelet transform techniques proposed in [29]. We pose the problem of finding the optimal coefficients as an inference problem in Markov random fields (MRFs). Example-based inference techniques have been used by [9, 30] for image superresolution. In a similar tone, we try to learn the mapping between uncompressed and compressed images with many training

patches. But the novelty is that now the patches are no longer image patches, but contiguous blocks of DCT coefficients which form the subband image for a particular frequency. We learn the potential functions for each subband image separately and inference is carried out for each subband image independently of the others. An attractive feature of this technique (originally introduced by Li and Delp [28]) is the fact that not all the subband coefficient images need to be inferred.

5.1. Block DCT Basics. Classical BDCT coding techniques divide the input image of size $M \times N$ into blocks of size $L \times L$, and then the 2-D DCT is obtained for each block. The 2-D DCT coefficients of the (i, j)th block $B_{(i,j)}$ can be described by

$$F_{(i,j)}^D(u, v) = \frac{2C(u)C(v)}{L} \sum_{m=0}^{L-1} \sum_{n=0}^{L-1} f_{(i,j)}(m, n)$$
$$\times \cos\left(\frac{(2m+1)}{2L}u\pi\right) \cos\left(\frac{(2n+1)}{2L}v\pi\right), \tag{10}$$

where $f_{(i,j)}(m, n)$ is the pixel value in $B_{(i,j)}$, $u, v = 0, \ldots, L-1$, $i = 0, \ldots, M/L$, $j = 0, \ldots, N/L$, and

$$C(u) = \begin{cases} \dfrac{1}{\sqrt{(2)}}, & u = 0 \\ 1, & u \neq 0. \end{cases} \tag{11}$$

After the DCT transform, the transform coefficients are quantized independently in each block (for the sake of simplicity in analysis, we assume that there is no intraprediction in the DCT coefficients across blocks) and the quantized coefficients can be determined by

$$F_{(i,j)}^Q(u, v) = \text{round}\left(\frac{F_{(i,j)}^D(u, v)}{\gamma Q(u, v)}\right), \tag{12}$$

where $Q(u, v)$ is a quantization table and γ is a quantization parameter which controls the overall quality level.

"Dequantization" and the inverse DCT (IDCT) are determined by

$$\hat{F}_{(i,j)}^D = \gamma F_{(i,j)}^Q(u, v) Q(u, v), \tag{13}$$

$$\hat{f}_{(i,j)}(m, n) = \frac{2}{L} \sum_{u=0}^{L-1} \sum_{v=0}^{L-1} C(u)C(v)\hat{F}_{(i,j)}^D(u, v)$$
$$\times \cos\left(\frac{(2m+1)}{2L}u\pi\right) \cos\left(\frac{(2n+1)}{2L}v\pi\right). \tag{14}$$

Since the dequantization process is not a lossless operation, it can be shown that the reconstructed DCT coefficients can actually have any value inside the QCS instead of the one defined in (13). The values lying outside the QCS are usually clipped. The independent reconstruction of each block DCT coefficients leads to discontinuities along the block boundaries which subsequently causes the blocky
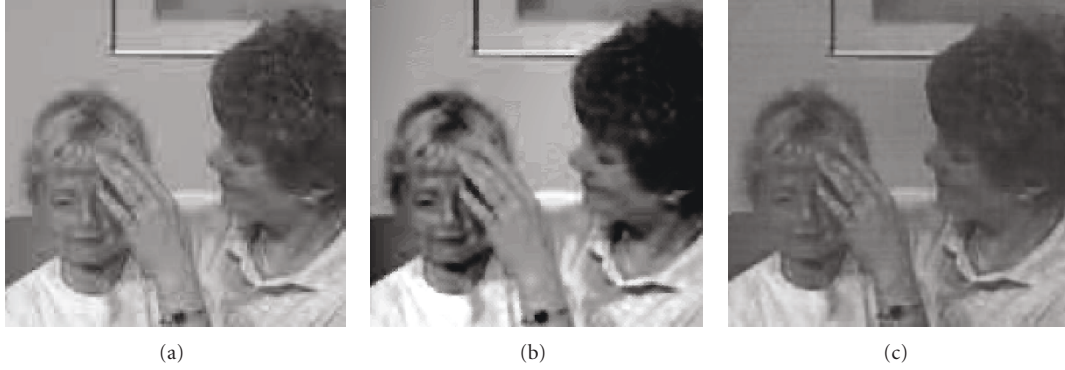
(a)   (b)   (c)

FIGURE 14: (Left) Compressed image; (center) reconstruction using our method (2×2 BDCT); (right) reconstruction using Yang and Galatsanos [23] method.

artifacts. It has been recognized that the DCT coefficients at the same frequency are highly correlated from block to block. On the other hand, the DCT coefficients at different frequencies are nearly uncorrelated which can be attributed to the suboptimality in energy compaction and decorrelation of the DCT. As a result, the DCT coefficients at the same frequency in different blocks can be regrouped as one coherent group. We define each group as $F(u, v)$, where (u, v) is the frequency or order of the coefficients. $F(u, v)$ is then considered as a subband image of size $(M/L) \times (N/L)$. A 2×2 BDCT-coded image and its subband images are shown in Figure 5. Clipping is done to ensure a proper display.

6. Experiments and Results

In this section, we will provide results for the different methods we have illustrated throughout this paper. The images are always represented as patches of size 4×4. Our training set consists of face images as well as license plate images. The low-resolution training dataset consists of gray scale images obtained by convolving Laplacian kernels of different scales with the high-resolution images. For generating the training data for the digit images, we collected blurred as well as sharp images from 20 different font families. For the face images, we use a proprietary face database which has 5 different poses for the same person. The database consists of aligned faces and hence any test face which is not in the training set needs to be aligned with respect to the training images.

First set of experiments are related to the multilayer MRF structure as elaborated in Section 2. We present results for both the applications mentioned above, namely, license plate image restoration and face image restoration. For the digit case, the results are shown in Figure 7. We test the recognition accuracy and confidence of the proposed alternating restoration and recognition algorithm. As mentioned earlier, the potentials during the first iteration are learned using random images from the training set. These potentials are then refined after the completion of the recognition algorithm, by making use of the confidence scores to obtain representative samples. For this experiment, the training

set consists of 200 synthetic images and the test set is composed of 200 real images of the blurred license plates. We present results of recognition accuracy and the improvement in confidence scores, before and after restoration, after 5 runs of the restoration and recognition loop. There was a significant improvement in recognition rate 92% after restoration compared to the 40% recognition rate before restoration. In Figure 6, we present the average confidence scores corresponding to the true digit class before and after restoration. We observe that there is a clear improvement in the confidence score for most of the digits ("0", "3", "4", "5", "6", "8"). In some cases ("1", "7"), we observe that the gain is not significant, as the confidence scores are already high. For the experiments with face images, the results are shown in Figure 8.

We demonstrate the superiority of the BP technique for inferring the DCT coefficients over other existing methods that have been proposed over the recent years. We have tried different block sizes resulting in different number of subband images. The overall performance is much more dependent on the compression bit rate rather than the block size. Also for most blocks around 60% of the subband images are modified thereby, saving on runtime. The training data for the experiments have been generated from Feret face recognition dataset [31] and the CMU PIE dataset [32]. The experiments are mainly done to handle face images, but the technique can be easily extended to other areas, as we show in the final set of experiments. We used 200 images for training. For color images, we process the different channels separately. Each training image is compressed with 5 different compression rates resulting in total 1000 training pairs of images.

The experiments are mainly divided into two categories. First, we try our method on the Lena image because of the wealth of published results for Lena. We report improvement over the method proposed by Zou and Yan [22], which reported improvements over most other existing methods. The comparative results are shown in Figure 9. Note that the results have been taken from published images and have not been regenerated. The eye portion in the result shows that it is sharper than the result obtained by Zou and Yan [22]. The PSNR values obtained for the Lena image are shown

in Figure 10. Though PSNR may not be the best metric for evaluation, since it has been used by many researchers to claim improvement, we show the results in Figure 10.

In the second set of experiments, we test on frames obtained from a real video for which no ground truth was available. The closest method to the one proposed in this work is the one proposed by Li and Delp [28]. The results shown in Figure 11 have been produced by Zhen Li, on his copyright software. Result for a frame of the same video, but using the method proposed in this paper, is shown in Figure 12. Clearly, the face looks much sharper in our method. The reconstructed images show much better performance near the eye and the teeth areas. Figure 13 shows face results generated from images compressed at higher bit rates.

We also compare our method to the method proposed by Yang and Galatsanos [23] for nonface-based experiments, as shown in Figure 14. The training data used for these experiments were still based on face images. Yang and Galatsanos [23] used explicit horizontal, vertical, and diagonal smoothness constraints which are expensive to compute. We try to capture all the information in the potential functions for the MRF.

7. Conclusion and Future Directions

In this paper, we have tried to address the problem of image restoration through multiple methods. We model the images as MRF's over a patch-based representation. The two spatial domain methods mentioned move toward the idea that high-level concepts like recognition can be used to aid low-level operations like restoration. We introduce a transformed domain method analogous to the spatial domain patch-based MRF and implement the system for removing compression artifacts from images and videos.

References

[1] B. Bascle, A. Blake, and A. Zisserman, "Motion deblurring and super-resolution from an image sequence," in *Proceedings of the 4th European Conference on Computer Vision (ECCV '96)*, pp. 573–581, Cambridge, UK, April 1996.

[2] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.

[3] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy under-sampled multiframes," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.

[4] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[5] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[6] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[7] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[8] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 2, pp. 627–634, Kauai, Hawaii, USA, December 2001.

[9] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[10] C. M. Bishop, A. Blake, and B. Marthi, "Super-resolution enhancement of video," in *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics (AI&Stats '03)*, Key West, Fla, USA, January 2003.

[11] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 192–198, Kauai, Hawaii, USA, December 2001.

[12] M. Das Gupta, S. Rajaram, N. Petrovic, and T. S. Huang, "Restoration and recognition in a loop," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 638–644, San Diego, Calif, USA, June 2005.

[13] S. Rajaram, M. Das Gupta, N. Petrovic, and T. S. Huang, "Learning-based nonparametric image super-resolution," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 51306, 11 pages, 2006.

[14] M. Isard, "PAMPAS: real-valued graphical models for computer vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 1, pp. 613–620, Madison, Wis, USA, June 2003.

[15] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 1, pp. 605–612, Madison, Wis, USA, June 2003.

[16] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, Calif, USA, 1988.

[17] "The frobnicatable foo filter," ECCV '06 submission ID 324. Supplied as additional material eccv06.pdf, 2006.

[18] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, CRC Press, Boca Raton, Fla, USA, 1986.

[19] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," in *Exploring Artificial Intelligence in the New Millennium*, pp. 239–269, Morgan Kaufmann, San Francisco, Calif, USA, 2003.

[20] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[21] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[22] J. J. Zou and H. Yan, "A deblocking method for BDCT Compressed images based on adaptive projections," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 3, pp. 430–435, 2005.

[23] Y. Yang and N. P. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process

modeling," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1345–1357, 1997.

[24] H. Paek, R.-C. Kim, and S.-U. Lee, "On the POCS-based postprocessing technique to reduce the blocking artifacts in transform coded images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 3, pp. 358–367, 1998.

[25] S. H. Park and D. S. Kim, "Theory of projection onto the narrow quantization constraint set and its application," *IEEE Transactions on Image Processing*, vol. 8, no. 10, pp. 1361–1373, 1999.

[26] Y. Jeong, I. Kim, and H. Kang, "A practical projection-based postprocessing of block-coded images with fast convergence rate," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 4, pp. 617–623, 2000.

[27] MPEG-4 video verification model version 18.0, iso/iecjtc1/ sc29/wg11, 2001.

[28] Z. Li and E. J. Delp, "Block artifact reduction using a transform-domain Markov random field model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1583–1593, 2005.

[29] Z. Xiong, M. T. Orchard, and Y.-Q. Zhang, "A deblocking algorithm for JPEG compressed images using overcomplete wavelet representations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 433–437, 1997.

[30] T. A. Stephenson and T. Chen, "Adaptive Markov random fields for example-based super-resolution of faces," *EURASIP Journal on Applied Signal Processing*, vol. 2006, Article ID 31062, 11 pages, 2006.

[31] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[32] T. Sim and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.