

Learning-Based Nonparametric Image Super-Resolution

Shyamsundar Rajaram,¹ Mithun Das Gupta,² Nemanja Petrovic,³ and Thomas S. Huang²

¹Beckman Institute, University of Illinois at Urbana-Champaign, IL 61801, USA

²Beckman Institute, at Urbana-Champaign, University of Illinois Urbana, IL 61801, USA

³Siemens Corporate Research, Princeton, NJ 08540-6632, USA

Received 1 December 2004; Revised 19 April 2005; Accepted 25 April 2005

We present a novel learning-based framework for zooming and recognizing images of digits obtained from vehicle registration plates, which have been blurred using an unknown kernel. We model the image as an undirected graphical model over image patches in which the compatibility functions are represented as nonparametric kernel densities. The crucial feature of this work is an iterative loop that alternates between super-resolution and restoration stages. A machine-learning-based framework has been used for restoration which also models spatial zooming. Image segmentation is done by a column-variance estimation-based “dissection” algorithm. Initially, the compatibility functions are learned by nonparametric kernel density estimation, using random samples from the training data. Next, we solve the inference problem by using an extended version of the nonparametric belief propagation algorithm, in which we introduce the notion of partial messages. Finally, we recognize the super-resolved and restored images. The resulting confidence scores are used to sample from the training set to better learn the compatibility functions.

Copyright © 2006 Hindawi Publishing Corporation. All rights reserved.

1. INTRODUCTION

Restoration plays a major role in most vision-based systems as the inputs in most cases are blurred/noisy. Blurred images are a nightmare for any recognition system. Many segmentation algorithms fail when the image is badly blurred. Restoration is a neat showcase of the ill-posedness of computer vision problems. Given a blurred image, there can be more than one sharp natural image which, when blurred, will generate the original image. In several important applications like surveillance, tracking, and license plate recognition systems, images may be severely blurred. Hence, the recognition strongly depends on the restoration performed either as an independent step or jointly with some other computer vision or learning tasks. In this paper, we present a method which can handle image restoration with super-resolution. Super-resolution is a specific process where the output image is of higher spatial resolution than the input image.

From a restoration point of view, a reasonable estimate of the restored image may be obtained if we have a priori knowledge about the blurring kernel. If no additive noise is present, then Wiener filtering can be used as it is the optimal filter. In the presence of additive noise the Wiener filter method gives the mean square error (MSE) minimized solution. In [1], Basclé et al. showed that restoration can be made relatively easy if multiple images are taken into account. Further, image restoration can be thought of as a special case

of super-resolution. Image deblurring and super-resolution have been treated concurrently by many authors, so we will intermittently (and erroneously) use the terms “sharp” and “high resolution” synonymously.

In recent years, time domain methods for super-resolution have been principle research fields. Among time domain methods, the two broad sections are iterative methods and learning-based methods. Iterative methods [2–6] mostly use a Bayesian framework, wherein an initial guess about the high-resolution frame is refined at each iteration. The image prior is usually assumed to be a smoothness prior. Another direction was proposed by [7] in which an edge based method was derived to estimate the local blurs for image super-resolution. The principle idea of the machine learning approach is to use a set of high-resolution images and their corresponding low-resolution images to build a compatibility model. The images are stored as patches or as coefficients of other feature representations. Recently, impressive amount of work has been reported in this field, [8–12], to name a few. In [10], face images were broken into regions or subspaces, and priors were imposed to penalize the high-resolution image to be close to the learnt means for the subspaces. In [12], PCA based techniques were used to capture the relationship between the high-resolution and low-resolution patches and then nonparametric modeling was used to estimate the missing details. In [11], an example-based learning method was employed for super resolving

natural images up to a zoom factor of 8. Along the same lines, Bishop et al. [9] performed video super-resolution by considering additional priors to account for the temporal consistencies between the successive frames. The learning-based methods can be made more powerful and robust if the images are restricted to be of a specific type, as in [8, 13] where face images are hallucinated. In [8], Baker and Kanade proposed a hallucination technique based on the recognition of generic local features. The local features are then used to predict a recognition-based prior rather than a smoothness prior as is the case with most iterative techniques. However, we notice that hallucinated images need not be realistic. The spirit of our work [14] is in close synchronization with the work of Freeman et al., however it differs from [11] in using partial message passing and restoration-recognition loop.

The two unique features of our work are partial messages and the restoration-super-resolution-recognition loop. Restoration is the key block since without restoration, the other modules lack the robustness and accuracy which any vision-based system working with real-life images requires. Our restoration algorithm is built on the notion of *partial message propagation* wherein we propose that any given image patch is only partially influenced by its neighbors, depending on the spatial orientation. Moreover the difference in sizes of the observed and inferred image patches results in spatial zooming or super-resolution. The model allows us to infer a zoomed version of the original blurred low-resolution image. The recognition step is performed inside the loop as it helps in localization of search space. For example, the search space for an image of the digit “8” can be greatly minimized if we can introduce information which reduces the search space to the set {0, 3, 6, 8, 9}, which is 50% minimization from the actual search space 0–9. With subsequent iterations, the search space is further reduced until only images of the digit “8” are remaining. Further, our method is not an example-based method, which means that the reconstructed image is not limited to one of the candidates from the training set. It removes the restriction of [11] that reconstructed high-resolution image must be a candidate from the training set whose low-resolution version is the most similar to the input low-resolution patch. Since our method is a learning-based method, the basic assumption is that we have seen the kinds of blurs in the training set which we want to restore in our test cases. In other words, if the training set has examples of Gaussian blurs only, then this set is not adequate for learning potentials when the test cases have motion blur as the primary blur. However, we can combine the training data sets comprising of examples of different kinds of blurs and then use the combined training set to learn the potentials to handle a wide range of blurs in the test cases. We employ sampling techniques to approximate the message products and only partial messages are propagated based on the spatial alignment of the image patches. To target a real-life application, we develop a fully automated license plate recognition system.

The organization of this paper is as follows. In Section 2, we describe the image model and review the details of the

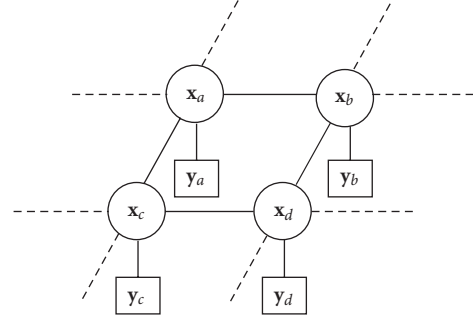


FIGURE 1: Image model. \mathbf{x}_i 's are the nonoverlapping hidden image patches and \mathbf{y}_i 's are the observed patches. The potential functions $\psi(\mathbf{x}_i, \mathbf{x}_j)$ model the interactions between spatial neighbors and $\phi(\mathbf{x}_i, \mathbf{y}_j)$ model the association between a patch and its observation.

nonparametric belief propagation (NBP) algorithm. In Section 3, we present our framework for image segmentation and then elaborate our method for restoration, super-resolution, and recognition in a loop. We introduce the features and potential functions and describe the application of NBP algorithm for restoration and super-resolution. In Section 4, we present experimental results of super-resolution and recognition on synthetic digit images and license plate images. We conclude in Section 5, with a discussion about our work and directions for future research.

2. THE MODEL

2.1. Problem statement and notation

Consider a training set of pairs of images of size n given by $\{(\mathbf{X}^1, \mathbf{Y}^1), (\mathbf{X}^2, \mathbf{Y}^2), \dots, (\mathbf{X}^n, \mathbf{Y}^n)\}$. Let there be an unknown kernel $f(\mathbf{X}^i)$ that maps from \mathbf{X}^i to \mathbf{Y}^i . The objective of the learning algorithm is, given the training set, to learn a model which can be used to infer the image \mathbf{X} from an observed image \mathbf{Y} which is not present in the training set. We model the image \mathbf{X} as an undirected graphical model or more specifically a Markov random field (MRF) [11]. MRF is a factorable distribution defined by the graph $G = \{V, E\}$ where each node represents a random variable \mathbf{x}_i , $i \in [1 \dots N]$, corresponding to a patch in the unknown, sharp image, which is associated with an observation node \mathbf{y}_i which represents the corresponding patch in the observed image (Figure 1). An edge between node \mathbf{x}_i and node \mathbf{x}_j indicates that they are spatial neighbors. The interaction between neighboring patches \mathbf{x}_i and \mathbf{x}_j is modelled using a potential function represented as $\psi(\mathbf{x}_i, \mathbf{x}_j)$ and commonly called the interaction potential. The association between the image patch \mathbf{x}_i and its observed blurred version \mathbf{y}_i is modelled as a pairwise potential represented by $\phi(\mathbf{x}_i, \mathbf{y}_i)$ —called the association potential. The probability distribution over the particular image and its blurred observation $p(\mathbf{X}, \mathbf{Y})$ can now be expressed in a factorized form as

$$p(\mathbf{X}, \mathbf{Y}) = \frac{1}{Z} \prod_{\{i,j\} \in E} \psi(\mathbf{x}_i, \mathbf{x}_j) \prod_{i \in V} \phi(\mathbf{x}_i, \mathbf{y}_i). \quad (1)$$

Learning with an MRF involves two phases, namely, the learning and the inference phase. In the learning phase, the potentials that model the interactions and associations are learned from the training data. The inference phase computes the marginals of posterior distribution $p(\mathbf{x}_i | \mathbf{Y})$, for all the nodes $i \in V$. In the next few sections, we review the BP algorithm that will be used in the restoration part of our restoration-recognition loop.

2.2. Belief propagation

For acyclic graphs, the conditional distributions can be calculated exactly by a local message-passing algorithm known as belief propagation (BP) [15]. The message propagated from node i to a node j in the n th iteration represented as $m_{ij}^n(\mathbf{x}_j)$ is given by

$$m_{ij}^n(\mathbf{x}_j) = \alpha \int_{\mathbf{x}_i} \psi(\mathbf{x}_i, \mathbf{x}_j) \phi(\mathbf{x}_i, \mathbf{y}_i) \prod_{h \in \Gamma(i) \setminus j} m_{hi}^{n-1}(\mathbf{x}_i), \quad (2)$$

where $\Gamma(i)$ indicates the neighborhood of the node \mathbf{x}_i and α represents an arbitrary proportionality constant. The messages computed can be combined to obtain the beliefs

$$b_i(\mathbf{x}_i) = \alpha \phi(\mathbf{x}_i, \mathbf{y}_i) \prod_{h \in \Gamma(i)} m_{hi}^n(\mathbf{x}_i). \quad (3)$$

For tree structured graphs, the beliefs converge to the actual marginal distributions once the messages from each node have been propagated to every other node. Therefore, marginals $p(\mathbf{x}_i | \mathbf{Y})$ are given by

$$p(\mathbf{x}_i | \mathbf{Y}) = \alpha \phi(\mathbf{x}_i, \mathbf{y}_i) \prod_{h \in \Gamma(i)} m_{hi}^n(\mathbf{x}_i). \quad (4)$$

In the case of graphs with cycles, the BP algorithm is not exact. The iterative version of BP algorithm produces the beliefs which do not converge to true marginals. But it was empirically shown that loopy BP produces excellent results for several hard problems. Recently, Yedidia et al. [16] established the link between the fixed points of BP algorithm and stationary points of ‘‘variational free energy’’ defined on the graphical model. This important result sheds more light on convergence and optimality properties of loopy BP approximation.

Loopy BP cannot be discretely applied to our model since the messages computed using (2) are mixtures of Gaussians and computing a message $m_{ij}^n(\mathbf{x}_j)$ involves the product of the interaction potential $\psi(\mathbf{x}_i, \mathbf{x}_j)$, the association potential $\phi(\mathbf{x}_j, \mathbf{y}_j)$, and the messages $m_{hi}^{n-1}(\mathbf{x}_i)$ $h \in \Gamma(i) \setminus j$ where each term is a *mixture of Gaussians*. Hence, in order to evaluate (2), the mixture components in the potentials and the messages have to be pruned so that the number of components in the product is within tractable limits to solve the integral (Figure 2). Such an approximation is unsuitable for the restoration problem and alternatively, we use the nonparametric extension of belief propagation proposed by Sudderth et al. [17] and independently invented by Isard [18], which we briefly review next. For the next section, we assume that



FIGURE 2: The individual Gaussians in the product can be approximated by fewer Gaussians.

we know the form of the association and interaction potentials. The form of the messages as well as the potentials are essentially the same so that the product term can be evaluated. Details about learning the potentials are discussed in Section 3.

2.3. Nonparametric belief propagation

We note that the interaction potential can be decomposed into a marginal influence term given by $\xi(\mathbf{x}_i) := \int_{\mathbf{x}_j} \psi(\mathbf{x}_i, \mathbf{x}_j)$ and a conditional interaction term $\psi(\mathbf{x}_j | \mathbf{x}_i)$. The message update (2) can be written as

$$m_{ij}^n(\mathbf{x}_j) = \alpha \int_{\mathbf{x}_i} \psi(\mathbf{x}_i | \mathbf{x}_j) \pi_{ij}^n(\mathbf{x}_i), \quad (5)$$

$$\pi_{ij}^n(\mathbf{x}_i) := \phi(\mathbf{x}_i, \mathbf{y}_i) \xi(\mathbf{x}_i) \prod_{h \in \Gamma(i) \setminus j} m_{hi}^{n-1}(\mathbf{x}_i). \quad (6)$$

The modified message update (5) can be solved in two phases. The first phase involves computing the term $\pi_{ij}^n(\mathbf{x}_i)$ and the second phase involves integrating the combination of $\pi_{ij}^n(\mathbf{x}_i)$ with the conditional interaction term $\psi(\mathbf{x}_j | \mathbf{x}_i)$. In [17, 18], Gibbs sampling technique was used to solve the first phase and the second phase was handled using stochastic integration. The messages were represented nonparametrically using kernel density estimate as

$$m_{ij}(\mathbf{x}_j) = \sum_{m=1}^M w_j^m \mathcal{N}(\mathbf{x}_j; \boldsymbol{\mu}_j^m, \boldsymbol{\Lambda}_j^m), \quad (7)$$

where, w_j^m , $\boldsymbol{\mu}_j^m$, $\boldsymbol{\Lambda}_j^m$ correspond to the weight, mean, and covariance associated with the m th kernel. In the following sections, we will elaborate on the procedure to obtain the message updates.

2.3.1. Products of Gaussians

Consider the product of L Gaussian distributions, $\prod_{l=1}^L \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_l, \boldsymbol{\Lambda}_l)$. The mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$ of the resulting product, which is a Gaussian $\mathcal{N}(\mathbf{z}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$, are

$$\begin{aligned} \boldsymbol{\Lambda}^{-1} &= \sum_{l=1}^L \boldsymbol{\Lambda}_l^{-1}, \\ \boldsymbol{\mu} &= \boldsymbol{\Lambda} \sum_{l=1}^L \boldsymbol{\Lambda}_l^{-1} \boldsymbol{\mu}_l. \end{aligned} \quad (8)$$

2.3.2. Products of Gaussians with different dimensions

As an introduction to our partial message-passing algorithm, let us consider the product of L Gaussian distributions where the first $L - 1$ of them are distributions over the random vector \mathbf{z} of dimension d and the L th one is a distribution over a subset of the components of the random vector $\tilde{\mathbf{z}}$ of dimension $\tilde{d} < d$. Without loss of generality, let us consider the case when $\tilde{\mathbf{z}}$ corresponds to the first \tilde{d} components of \mathbf{z} . The resulting product of such L distributions is a Gaussian, whose mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$ can be computed in closed form, and is given by

$$\begin{aligned} \boldsymbol{\Lambda}^{-1} &= \sum_{l=1}^{L-1} \boldsymbol{\Lambda}_l^{-1} + \begin{pmatrix} (\boldsymbol{\Lambda}_L^{-1})_{\tilde{d} \times \tilde{d}} & \mathbf{0}_{\tilde{d} \times (d-\tilde{d})} \\ \mathbf{0}_{(d-\tilde{d}) \times \tilde{d}} & \mathbf{0}_{(d-\tilde{d}) \times (d-\tilde{d})} \end{pmatrix}, \\ \boldsymbol{\Lambda}^{-1} \boldsymbol{\mu} &= \sum_{l=1}^{L-1} \boldsymbol{\Lambda}_l^{-1} \boldsymbol{\mu}_l \\ &+ \begin{pmatrix} (\boldsymbol{\Lambda}_L^{-1})_{\tilde{d} \times \tilde{d}} & \mathbf{0}_{\tilde{d} \times (d-\tilde{d})} \\ \mathbf{0}_{(d-\tilde{d}) \times \tilde{d}} & \mathbf{0}_{(d-\tilde{d}) \times (d-\tilde{d})} \end{pmatrix} \begin{pmatrix} (\boldsymbol{\mu}_L)_{\tilde{d} \times 1} \\ \mathbf{0}_{(d-\tilde{d}) \times 1} \end{pmatrix}. \end{aligned} \quad (9)$$

The above approach can be extended to handle products of L Gaussian distributions, where each term is a distribution over a different subset $\tilde{\mathbf{x}}$ of the components of \mathbf{x} , where the sets satisfy the property that each component of the random vector \mathbf{x} appears in at least one of the L different sets.

2.3.3. Parallel sampling

The first phase of computing the messages corresponds to evaluating the product $\pi_{i,j}^n(\mathbf{x}_i)$. We observe that each term in the product is a mixture of Gaussians and, for instance, 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 as explained in Section 2.3.1, however, it is not feasible because of the $O(M^L)$ 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 [19] and importance weighting were used in [17, 18] to generate M asymptotically unbiased samples without explicitly computing the product.

In this work, we use alternating Gibbs sampling [20] to obtain samples from the product $\pi_{i,j}^n(\mathbf{x}_i)$. The procedure for alternating Gibbs sampling to sample from a product of the form $\prod_{l=1}^L \sum_{m=1}^M w_{l,m} \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{l,m}, \boldsymbol{\Lambda}_{l,m})$ is as follows.

- (1) Pick a data vector \mathbf{z} randomly.
- (2) Compute the posterior probability $P_{l,m} = w_{l,m} \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{l,m}, \boldsymbol{\Lambda}_{l,m})$ for each of the M mixture components in every term of the product, given the data vector \mathbf{z} .
- (3) Pick a mixture component m_l for each term in the product based on the posterior probability distribution.
- (4) Compute the resulting distribution obtained by multiplying the picked mixture components, that is, $\prod_{l=1}^L \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}_{l,m_l}, \boldsymbol{\Lambda}_{l,m_l})$, using (8).

- (5) Sample from this resulting distribution to obtain the new data vector \mathbf{z} .
- (6) Go back to step 2.

The above technique can be used to obtain asymptotically unbiased samples $\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^M$ from $\pi_{i,j}^n(\mathbf{x}_i)$. Further, the same sampling approach can be used to obtain samples for the posterior $p(\mathbf{x}_i | \mathbf{Y})$ given by (4) after each iteration of the message passing algorithm.

2.3.4. Message updates

The second phase of obtaining the message update is to integrate the combination of the samples obtained from alternating Gibbs sampling and the conditional interaction potential. This is performed using stochastic integration, where every sample \mathbf{x}_i^m is propagated to node j by sampling \mathbf{x}_j^m from $\psi(\mathbf{x}_i = \mathbf{x}_i^m, \mathbf{x}_j)$. Now, nonparametric density estimation (7) is used to obtain the messages $m_{i,j}^n(\mathbf{x}_j)$, 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 [21].

3. LICENSE PLATE RECOGNITION (LPR) SYSTEM: AN APPLICATION FOR SEGMENTATION, SUPER-RESOLUTION, AND RECOGNITION

In this section, we will elaborate the algorithmic modules of the entire LPR system. Segmentation is performed outside the loop. First, we elaborate the algorithm for segmenting out images of digits from clustered license plates. For image segmentation, we use a dissection method to cut out the images of the digits from the blurred license plates. Then we will elaborate the application of the NBP algorithm for super-resolving and restoring a blurred and downsampled image of a digit. The recognition phase is explained thereafter, which is the last module in the loop.

From our modeling perspective, the observation \mathbf{Y} corresponds to a blurred and downsampled version of the original image \mathbf{X} with an unknown blurring kernel function $f(\mathbf{X})$. The training set comprises of several instances of image and blurred-downsampled version pairs $\{\mathbf{X}, \mathbf{Y}\}$. As elaborated in Section 2.1, \mathbf{X} is modelled as an MRF over the patch-based representation \mathbf{x}_i , $i \in [1 \dots N]$. The choice of patch size is often a critical issue, as a small sized patch does not capture enough information which makes the restoration very ill-posed and a bigger sized patch captures too much information thereby resulting in computational problems. In this work, after performing several experiments, we chose the observation patches to be of size 4×4 and the patches in high-resolution images to be of size 8×8 .

3.1. Unsupervised image segmentation

The first step in the LPR system is to segment the blurred image of the license plate into individual images of digits. We follow a ‘‘dissection’’ scheme similar to [22]. In this work, we make a few assumptions based on the properties inherent to registration plates. We assume that the digits in a plate are

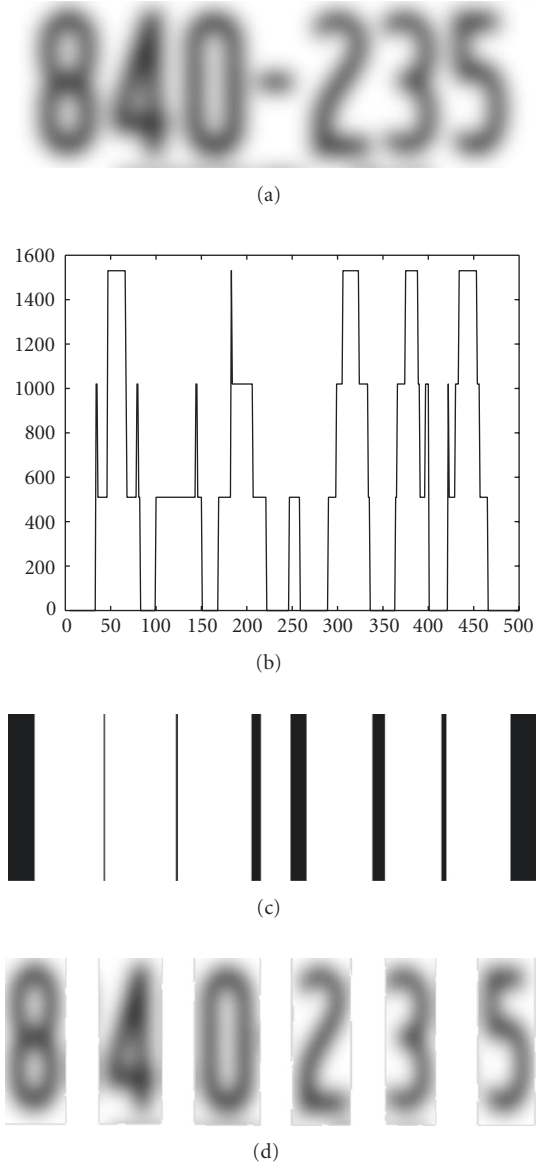


FIGURE 3: (a) blurred input image, (b) variance along a given column, (c) the thresholded variance plot, (d) segmented digits.

always dark and the background is of a lighter shade. Next we assume that the smallest rectangular box binding the digits in a plate has the maximum area compared to all other contents of the plate. The segmentation scheme is illustrated in Figure 3. The variance along a column is estimated and then compared against a threshold to obtain the thresholded variance plot in Figure 3. The final segmentation is done by cutting the initial blurred image along the black regions as shown in the fourth image in Figure 3.

3.2. Learning the association and interaction potentials

One of the novelties of this work is in using *nonparametric kernel density estimation for learning the potentials* to avoid

the averaging effects in the parametric methods which is against the spirit of a restoration problem. Namely, one could try to fit the high-resolution portion of the training set with continuous parametric distribution that is tractable for BP algorithm (mixture of Gaussians, e.g.). From a generative model perspective, sampling from such parametric distribution would produce samples that are averaged versions of the images in the training set, and hence the generated images would be blurred. A parametric model would not generate samples that are similar to the training data and hence, we use a nonparametric modeling approach for restoration.

We model the association potential $\phi(\mathbf{x}_i, \mathbf{y}_i)$ as a function over the vectorized patch association, as shown in Figure 4, and with the form

$$\phi(\mathbf{x}_i, \mathbf{y}_i) = \frac{1}{M} \sum_{m=1}^M \mathcal{N}([\mathbf{x}_i, \mathbf{y}_i]^t; \boldsymbol{\mu}_m, \boldsymbol{\Lambda}_m), \quad (10)$$

where M is the number of components and $\mathcal{N}([\mathbf{x}, \mathbf{y}]^t; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$ over the random vector $[\mathbf{x}, \mathbf{y}]^t$. From the training images, the patch association vectors $[\mathbf{x}, \mathbf{y}]^t$ corresponding to the image and its blurred-downsampled version are constructed. The patch association vectors are pruned to avoid redundancy. In other words, patches that are similar to each other are represented by a single patch. The potential is constructed by considering a kernel with the mean chosen as the patch association vector and the covariances are chosen using the leave-one outcross validation technique [21]. The interaction potential $\psi(\mathbf{x}_i, \mathbf{x}_j)$ is a function over the vectorized two pixel-thick nonoverlapping patch boundary, as shown in Figure 5, and learned using the above mentioned nonparametric estimation technique. The interaction potential is of the form

$$\psi(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{N} \sum_{n=1}^N \mathcal{N}([\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i]^t; \boldsymbol{\mu}_n, \boldsymbol{\Lambda}_n), \quad (11)$$

where N is the number of components and $\mathcal{N}([\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i]^t; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$ over the random vector $[\tilde{\mathbf{x}}_j, \tilde{\mathbf{x}}_i]^t$. The notation $\tilde{\mathbf{x}}_i$ has been used to denote the boundary pixels which actually interact while passing messages between two neighboring nodes. In our partial message-passing algorithm, only the pixels which are close to the boundary interact with the neighbor as shown in Figure 5. The interaction potential vectors are pruned to avoid redundancy just as in the case of the association potential.

The drawback of using the nonparametric approach is that the number of components M are equal to the total number of samples which results in computational issues while performing inference. Thus, we are restricted to using only a few images from the training set for learning the potentials. The novelty of this work is in overcoming the above problem by an *iterative loop which alternates between recognition/super-resolution and restoration* better illustrated in Figure 6. The recognition block generates the confidence scores which are then used to sample from the training data set and these samples are then used to learn better potential

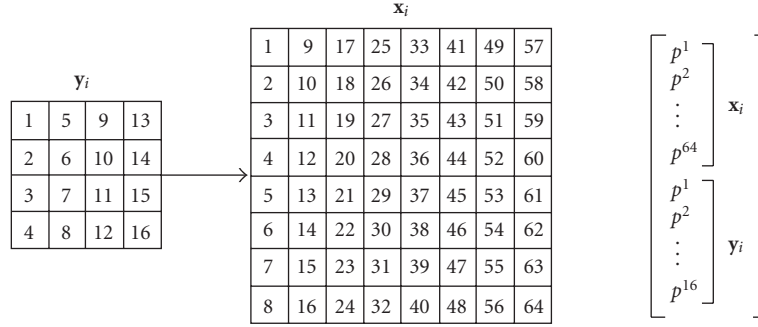


FIGURE 4: The vectorized pixels in patch \mathbf{x}_i are appended onto the vectorized pixels in patch \mathbf{y}_i to obtain a feature vector. The association potential $\phi(\mathbf{x}_i, \mathbf{y}_i)$ is a function over this feature vector.

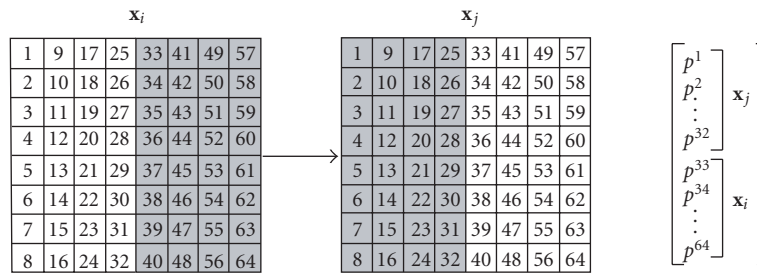


FIGURE 5: The vectorized boundary pixels of patches \mathbf{x}_i and \mathbf{x}_j are appended to obtain a feature vector. The interaction potential $\psi(\mathbf{x}_i, \mathbf{x}_j)$ is a function over this feature vector.

functions. The recognition approach is further elaborated in Section 3.4. During the first iteration, the association as well as the interaction potentials are learnt from patches obtained from a randomly sampled set of images in the training set. The learning-based method enables us to restore and hence super-resolve a test image even if we have not seen *exactly* the same image in the training ensemble. If a priori knowledge was available about the digit present in the image, then we can learn the potentials only from images of the same digit class. Since we do not have this information, we go for random training set where each digit is present at least once. By iterating the confidence, the true digit increases and so we select more and more training images from the true digit class. Optimality is ensured in the sense that a particular digit is more or less the same across different font families, provided other factors like size are held constant.

3.3. Restoration and super-resolution using nonparametric belief propagation

The first step in restoration involves iterating the message-passing algorithm several times after which the posterior distribution $p(\mathbf{x}_i | \mathbf{Y})$, $i \in [1 \dots N]$, can be obtained using alternating Gibbs sampling.

As mentioned earlier, the message update $m_{i,j}^n(\mathbf{x}_j)$ is performed in two different phases, where the first phase involves sampling from the product $\pi_{i,j}^n(\mathbf{x}_i)$, as in (6), using the alternating Gibbs sampling method and the second

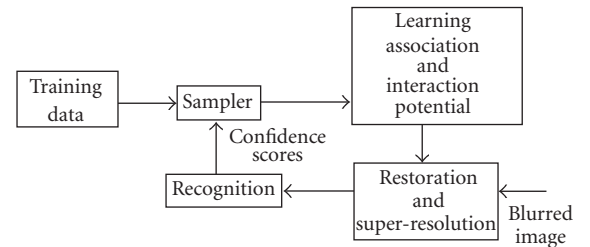


FIGURE 6: Block diagram illustrating our framework for performing recognition and restoration in a loop.

phase corresponds to stochastic integration by propagating the samples obtained from sampling to node j based on the conditional interaction, which is followed by non parametric estimation of the message as a kernel density estimate. Now, we note that the message $m_{i,j}(\mathbf{x}_j)$ is a function of the four pixel-thick boundary pixels because of the structure of the interaction potential. Another novel contribution of this work is the notion of passing *partial messages* to a node. The idea is better illustrated in Figure 7, where we have indicated the partial influence of the left neighbor on the central patch. This modeling is crucial to ensure good interaction between adjacent patches and it is based on the intuition that neighboring patches are more likely to have influences on the boundary pixels rather than on the whole patch. Hence, we introduce the notation $\tilde{\mathbf{x}}_{i,j}$ which represents

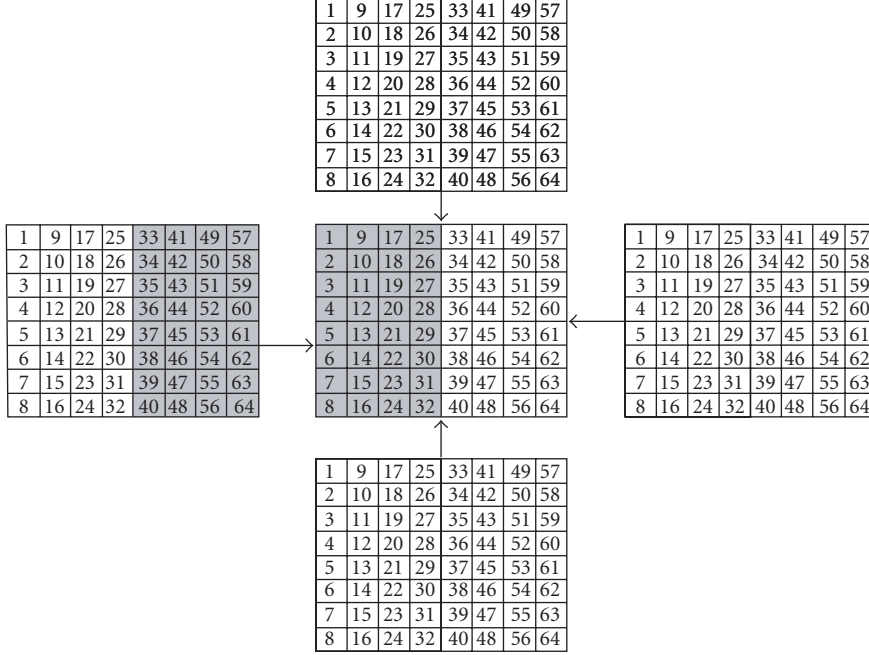


FIGURE 7: Illustration of our partial influence model: pixels 1–32 of the central patch are influenced by pixels 33–64 of the left patch.

the four pixel-thick boundary of the patches corresponding to nodes i and j and further, the messages are denoted as $m_{i,j}^n(\tilde{\mathbf{x}}_{i,j})$. The alternating Gibbs sampling procedure can still be used to generate samples from $\pi_{i,j}^n(\mathbf{x}_i)$ with a modification to step 4 of the algorithm. In the restoration problem setup, the product in step 4 corresponds to

$$\underbrace{\mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}, \boldsymbol{\Lambda})}_{\phi(\mathbf{x}_i, \mathbf{y}_i)} \underbrace{\mathcal{N}(\tilde{\mathbf{x}}_{i,j}; \tilde{\boldsymbol{\mu}}_{i,j}, \tilde{\boldsymbol{\Lambda}}_{i,j})}_{\xi(\tilde{\mathbf{x}})} \prod_{h \in \Gamma(i) \setminus j} \underbrace{\mathcal{N}(\tilde{\mathbf{x}}_{h,i}; \tilde{\boldsymbol{\mu}}_{h,i}, \tilde{\boldsymbol{\Lambda}}_{h,i})}_{m_{h,i}^{n-1}(\tilde{\mathbf{x}}_{h,i})}, \quad (12)$$

where the function indicated under the braces ($\phi(\mathbf{x}_i, \mathbf{y}_i)$, $\xi(\tilde{\mathbf{x}})$, or $m_{h,i}^{n-1}(\tilde{\mathbf{x}}_{h,i})$) is the term in $\pi_{i,j}^n(\mathbf{x}_i)$ from which the component was picked. Except for the first term which is a normal over \mathbf{x}_i , the rest—the component from the marginal influence term and the component from the messages—are normal distributions over subsets of the components of the random vector \mathbf{x}_i . Such a product of Gaussians can be solved using the method discussed in Section 2.3.1 for computing the products of normal distributions over different subsets of the components of a random vector. The rest of the alternating Gibbs sampling procedure remains unchanged and can be used to generate samples from $\pi_{i,j}^n(\mathbf{x}_i)$ and further these samples are propagated to node j as explained in Section 2.3.4. Several iterations of the message update algorithm are performed for all the nodes in the graph and at the end of each iteration, the posterior $p(\mathbf{x}_i | \mathbf{Y})$ given in (4) can be computed using the alternating Gibbs sampling procedure.

One of the key contributions of this framework is handling restoration and super-resolution within the same step. Since we learn the interactions between the downsampled

and blurred image patches and their corresponding high-resolution, sharp image patches, we handle both the inference problems, namely deblurring and super-resolution, within the same model. In other words, this method extends the learning-based algorithms which have been used by numerous researchers in restoration as well as super-resolution.

3.4. Recognition

One of the key contributions of this work is the iterative loop which alternates between restoration/super-resolution and recognition as shown in Figure 6. This feedback feature allows us to perform sampling from the training set in a way which ensures that the data used for learning the potentials are similar to the test data. This technique resembles a boosting procedure wherein the distribution over the class labels is modified in order to boost the performance of the restoration method.

The algorithm used for recognition of the digits is based on the k -nearest-neighbor algorithm. The Euclidean distance metric is used to compute the distances between the test image X and the images in the database. Based on the distance, the top k closest points from the dataset $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k\}$ are picked. Let us denote the distances corresponding to the top k points to be $\{D_1, D_2, \dots, D_k\}$ and denote the class index of the i th closest image to be $c_i, c_i \in [0 \dots 9]$. Now, we arrive at a confidence estimate for each class using

$$C(\text{class} = c | \mathbf{X}) = \frac{1}{Z} \left(\frac{\sum_{i=1}^k \exp(-D_i I_c(c_i))}{\sum_i I_c(c_i)} \right), \quad (13)$$

where Z is a normalizing constant and $I_c(c_i)$ is an indicator

function. The decision rule for recognition is given by

$$c^* = \arg \max_c C(\text{class} = c | \mathbf{X}). \quad (14)$$

The confidence estimate obtained for each class is a multinomial distribution over the class labels. Samples are then generated from the training data set based on this multinomial distribution and these samples are then used to learn the association and interaction potentials.

4. EXPERIMENTS AND RESULTS

For this work we performed several experiments on synthetic images of digits and actual license plate images. For the high-resolution training data set, we used digit fonts from 20 different font families. The digits are represented by same-size center-aligned binary images. The low-resolution training data set consists of corresponding gray-scale images obtained by convolution with a blurring kernel followed by downsampling. Different types of blurring kernels used in this work are Gaussian and Laplacian Kernels. Another neat feature of this work is the fact that different types of blurs can be modeled simultaneously and inference can be performed in a unified framework. Since we have some idea about the target domain, we can train our algorithm to handle such blurring kernels. More severe levels of blur and different types, namely, motion blur, can also be incorporated in this framework.

The real data set consists of blurred binary registration plate images. The registration plates are segmented using the algorithm discussed in Section 3.1. To evaluate the results of the restoration algorithm, we subjectively inspect the visual quality of the restored images. The spatial zoom has been fixed to a factor of 2 in both directions. Zoom factors less than 2 can be easily achieved, but for higher zooming levels, the patch sizes for the sharp images have to be very large and hence block effects would start appearing in the inferred high-resolution image. Additionally, we measure the improvement in the recognition rate and recognition confidence after performing restoration and recognition in a loop.

In the first experiment, we perform a “sanity check” to determine the restoration performance for a test image, where the potentials are learned using images from the training set which correspond to the digit in the test image. In Figure 8, we illustrate the restoration of blurred digits “5” and “6.” The reconstructed high-resolution images were almost indistinguishable from the originals (for digit “5”) or very close to the original (for digit “6”). The zoom factor is 2 in both directions.

In the second experiment, we trained the potentials on a set of 150 synthetic images consisting of 15 images for each digit. The testing set consists of 50 images consisting of 5 images for each digit. We inspected the improvement of visual quality of reconstructed images after each of five iterations of NBP algorithm (Figure 9). We note that the vertical and horizontal lines of digit “5” become thinner and clearer as the iterations proceed. For the digit “9,” the reconstruction becomes less spiky and the semicircular regions become thinner and smoother.

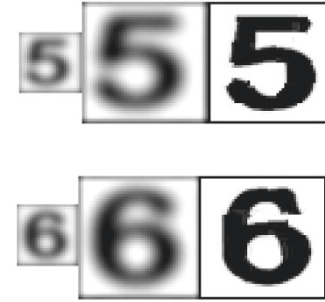


FIGURE 8: Left image is the input to the system. Center image is sharpened using deconvolution methods and then zoomed using interpolation methods. Right image is the output of our algorithm. Testing samples are taken from the training ensemble.



FIGURE 9: Restoration result obtained after second, third, fourth, and fifth iterations of the NBP algorithm for “5” and “9.” Zooming factor is equal to 2.

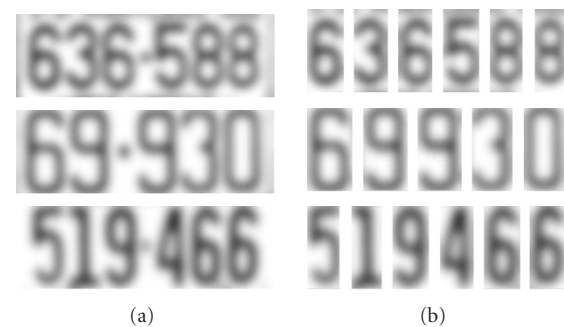


FIGURE 10: Segmentation results for original license plates: (a) original plates, (b) segmented digits.

The segmentation routine is tested on actual license plates as shown in the Figure 10. In spite of the very simple segmentation algorithm, the results were near optimum for most cases. In some cases the digits are a little thinner than necessary, but this can be overcome by appropriate padding with the boundary pixels. The individual digit images are resized before they are input to the restoration/super-resolution recognition loop.

Next, we test the recognition accuracy and confidence of our alternating restoration and resolution algorithm. For

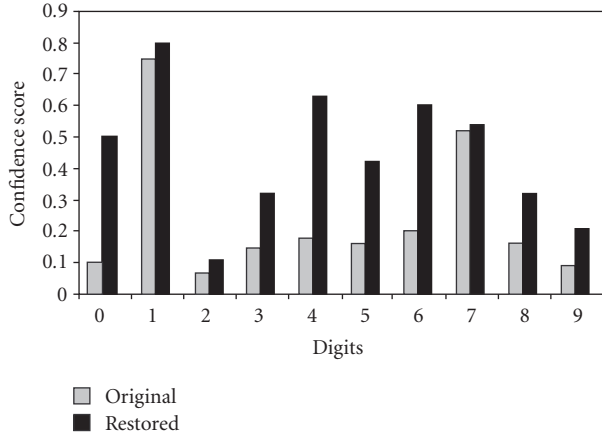


FIGURE 11: Confidence scores versus digits. Original confidence scores denote the recognition confidence for the blurred digits, the darker bars show the recognition confidence for restored digits.

this experiment, the training set consists of 200 synthetic images and the test set is composed of 200 real images from 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/super-resolution and recognition loop. There was a significant improvement in recognition rate from 40% to 92% because of restoration. In Figure 11, 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.

Finally we present test results on real license plates images as shown in Figure 12. The original license plate, the blurred version of the digits, the deconvolution-based deblurring results upsampled by using interpolation methods, as well as our results are shown. It is fairly clear that our method works well for this scenario.

5. CONCLUSION AND FUTURE WORK

In this work, we use nonparametric belief propagation to perform restoration of images of digits which have been blurred using an unknown kernel. We incorporate zooming in our learning-based framework, and hence super-resolution is also achieved along with deblurring. We introduce the notion of partial messages and extend the NBP approach to handle them. The main contribution of this work is the framework where the confidence scores of recognition are fed back to the restoration/super-resolution algorithm. The confidence scores are used to generate samples from the training data set based on which the potentials of the field are learned. This results in a boosting of the restoration performance. We achieve a spatial zooming up to 2 times in both directions. Further, we show significant improvement in recognition performance for synthetic images and digits in

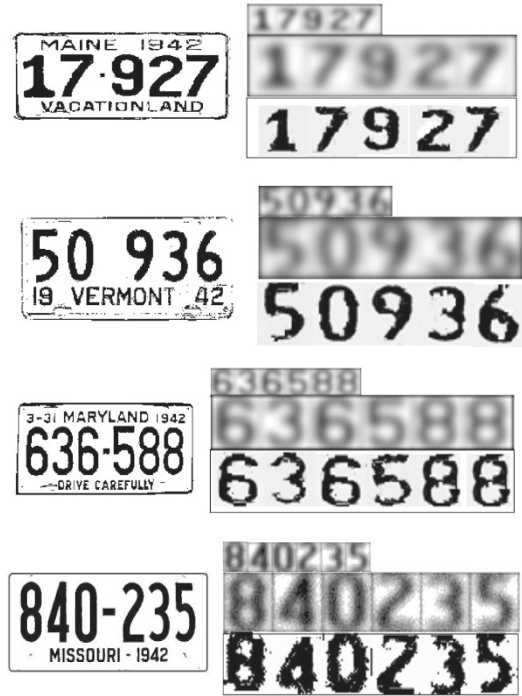


FIGURE 12: Left: the original license plate. Right: (top to bottom) blurred input, deblurred, and zoomed using deconvolution methods followed by interpolation, our method. Note the addition of noise in the 4th test case.

license plates. Another important feature of this work is the integration of an LPR system, which is fully automated, robust, and self boosting.

This work primarily focusses on single frame super-resolution. The model can easily be extended to handle traditional multiframe super-resolution [23].

Another promising direction of future research is to extend the above framework for real scenes and other constrained domains like faces for applications like tracking and surveillance. Another challenging field where we can extend this framework is time super-resolution. Inter-frame relationships can be learnt for applications like frame rate enhancement. This learning-based framework can easily be extended to handle different kinds and levels of blurs.

ACKNOWLEDGMENT

This work was supported in part by Advanced Research and Development Activities (ARDA) under Contract MDA904-03-C-1787.

REFERENCES

- [1] B. Bascle, A. Blake, and A. Zisserman, “Motion deblurring and super-resolution from an image sequence,” in *Proceedings of 4th European Conference on Computer Vision (ECCV ’96)*, vol. 2, pp. 573–582, Cambridge, UK, April 1996.
- [2] 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.
- [3] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231–239, 1991.
- [4] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 6, pp. 1013–1027, 1990.
- [5] 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.
- [6] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America. A, Optics and image science.*, vol. 6, no. 11, pp. 1715–1726, 1989.
- [7] M.-C. Chiang and T. E. Boulton, "Local blur estimation and super-resolution," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 821–826, San Juan, Puerto Rico, USA, June 1997.
- [8] 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.
- [9] C. M. Bishop, A. Blake, and B. Marthi, "Super-resolution enhancement of video," in *Proceedings of 9th International Workshop on Artificial Intelligence and Statistics (AISTATS '03)*, Key West, Fla, USA, January 2003.
- [10] D. Capel and A. Zisserman, "Super-resolution from multiple views using learnt image models," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 2, pp. II-627–II-634, Kauai, Hawaii, USA, December 2001.
- [11] 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.
- [12] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: global parametric model and local non-parametric model," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. I-192–I-198, Kauai, Hawaii, USA, December 2001.
- [13] G. Dedeoglu, T. Kanade, and J. August, "High-zoom video hallucination by exploiting spatio-temporal regularities," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II-151–II-158, Washington, DC, USA, June–July 2004.
- [14] M. D. Gupta, S. Rajaram, N. Petrovic, and T. S. Huang, "Restoration and recognition in a loop," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 638–644, San Diego, Calif, USA, June 2005.
- [15] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, Calif, USA, 1988.
- [16] 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.
- [17] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky, "Nonparametric belief propagation," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 1, pp. I-605–I-612, Madison, Wis, USA, June 2003.
- [18] M. Isard, "PAMPAS: real-valued graphical models for computer vision," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, vol. 1, pp. I-613–I-620, Madison, Wis, USA, June 2003.
- [19] 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.
- [20] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
- [21] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 1986.
- [22] R. L. Hoffman and J. W. McCullough, "Segmentation methods for recognition of machine-printed characters," *IBM Journal of Research and Development*, vol. 15, no. 2, pp. 153–165, 1971.
- [23] M. D. Gupta, S. Rajaram, N. Petrovic, and T. S. Huang, "Non-parametric image super-resolution using multiple images," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, Genova, Italy, September 2005.

Shyamsundar Rajaram received the B.S. degree in electrical engineering from the University of Madras, India, in 2000, and the M.S. degree in electrical engineering from the University of Illinois at Chicago in 2002. He is currently working on his Ph.D. thesis at the University of Illinois at Urbana-Champaign under Professor Thomas S. Huang. He has published several papers in the field of machine learning and its applications in signal processing, computer vision, information retrieval, and other domains.



Mithun Das Gupta received his B.S. degree in instrumentation engineering from the Indian Institute of Technology, Kharagpur, in 2001, and his M.S. degree in electrical engineering from the University of Illinois at Urbana-Champaign in 2003. He is currently pursuing his Ph.D. degree under the guidance of Professor Thomas S. Huang at the University of Illinois at Urbana-Champaign. His research interests include learning-based methods for image and video understanding and enhancement.



Nemanja Petrovic received his Ph.D. degree from the University of Illinois, in 2004. He is currently a member of research staff at Siemens Corporate Research, Princeton, New Jersey. His professional interests are computer vision and machine learning. Dr. Petrovic has published 20 papers in the area of graphical models, video understanding, and data clustering and enhancement.



Thomas S. Huang received his B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, China; and his M.S. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts. He was at the faculty of the Department of Electrical Engineering at MIT from 1963 to 1973; and at the faculty of the School of Electrical Engineering and Director of its Laboratory for Information and Signal Processing at Purdue University from 1973 to 1980. In 1980, he joined the University of Illinois at Urbana-Champaign, where he is now William L. Everitt Distinguished Professor of Electrical and Computer Engineering, Research Professor at the Coordinated Science Laboratory, Head of the Image Formation and Processing Group at the Beckman Institute for Advanced Science and Technology, and Cochair of the Institute's major research theme: human-computer intelligent interaction. Dr. Huang's professional interests lie in the broad area of information technology, especially the transmission and processing of multidimensional signals. He has published 20 books and over 500 papers in network theory, digital filtering, image processing, and computer vision. He is a Member of the National Academy of Engineering, a Foreign Member of the Chinese Academies of Engineering and Science, and a Fellow of the International Association of Pattern Recognition, IEEE, and the Optical Society of America, and has received a Guggenheim Fellowship, an A.von Humboldt Foundation Senior US Scientist Award, and a Fellowship from the Japan Association for the Promotion of Science. He received the IEEE Signal Processing Society's Technical Achievement Award in 1987, and the Society Award in 1991. He was awarded the IEEE Third Millennium Medal in 2000. Also in 2000, he received the Honda Lifetime Achievement Award for "contributions to motion analysis." In 2001, he received the IEEE Jack S. Kilby Medal. In 2002, he received the King-Sun Fu Prize, International Association of Pattern Recognition, and the Pan Wen-Yuan Outstanding Research Award. He is a Founding Editor of the International Journal of Computer Vision, Graphics, and Image Processing and Editor of the Springer Series in Information Sciences, published by Springer.

