IMAGE RESTORATION USING NON-HOMOGENOUS FIELD MODELING

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

1 University of Illinois at Urbana Champaign, 2 Siemens Corporate Research, Princeton NJ.

mdgupta@ifp.uiuc.edu, rajaram1@ifp.uiuc.edu, nemanja@scr.siemens.com, huang@ifp.uiuc.edu

Abstract

In this paper, we present a novel learning based framework for restoring 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 non-parametric kernel densities. The underlying field is assumed to be a non-homogenous field with separate compatibility functions for different regions in the image. Initially, the compatibility functions are learned by non-parametric kernel density estimation. Next, we solve the inference problem by using an extended version of the non-parametric belief propagation algorithm.

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. 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 images which, when blurred, will generate the original image.

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, then Wiener filtering can be used as it is the optimal filter. In the presence of additive noise the Weiner filter method gives the MMSE solution. 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.

In recent years, time domain methods for superresolution have been principle research fields. Among the time domain methods, the two broad sections are iterative methods and learning based methods. Iterative methods [7, 9, 12, 6] mostly use a Bayesian framework. 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 [1, 3, 10, 4, 2], to name a few. In [4], an example based learning method was employed for super resolving natural images up to a zoom factor of 8. The learning based methods can be made more powerful and robust if the images are restricted to be of a specific type, as in [1] or [5] where face images are hallucinated. We note that hallucinated images need not be realistic, though. The unique feature of our work is non-homogenous field assumption. We slide an overlapping window over the image and learn the potential functions based on some similarity measure of the contents of the window with our training data set.

The organization of this paper is as follows. In Sec. 2, we describe the image model and review the details of the non-parametric belief propagation (NBP) algorithm. In Sec. 3, we elaborate our method for image restoration. We introduce the features, potential functions and the non-homogenous field assumption. In Sec. 4, we present experimental results of restoration on synthetic digit images and license plate images. We conclude in Sec. 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 $\{(\mathbf{X}^1, \mathbf{Y}^1), (\mathbf{X}^2, \mathbf{Y}^2), \ldots, (\mathbf{X}^n, \mathbf{Y}^n)\}$. 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) [4]. MRF is a factorable distribution defined by the graph $G = \{V, E\}$ where each node represents a random variable $\mathbf{x}_i, i \in [1 \ldots N]$ 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 (Fig. 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

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

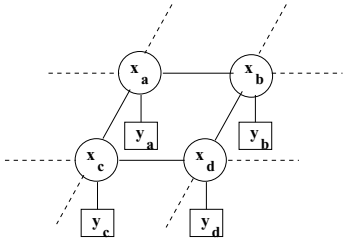

Fig. 1. Image Model. $\mathbf{x}_i$'s are the non-overlapping hidden image patches, $\mathbf{y}_i$'s are the observed patches. The potential functions $\psi(\mathbf{x}_i, \mathbf{x}_j)$ model the interactions between spatial neighbors and $\phi(\mathbf{x}_i, \mathbf{y}_j)$ model the association between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

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. Non-parametric Belief Propagation

For acyclic graphs, the conditional distributions can be calculated exactly by a local message passing algorithm known as Belief Propagation(BP) [11]. Non-parametric BP a simple extension of this algorithm. 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:

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

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

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

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,

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

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 $\psi(\mathbf{x}_i^m, \mathbf{x}_j)$. The message update equation Eqn. 2 can be solved in two phases. The first phase involves computing the product $\boldsymbol{\pi}_{i,j}^n(\mathbf{x}_i)$ of messages $\boldsymbol{m}_{hi}^{n-1}(\mathbf{x}_i), h \in \Gamma(i) \setminus j$, the association potential $\phi(\mathbf{x}_j, \mathbf{y}_j)$ in which $\mathbf{y}_j$ is the local observation corresponding to the j^{th} patch and the marginal influence term $\xi(\mathbf{x}_i)$.

$$\boldsymbol{\pi}_{i,j}^n(\mathbf{x}_i) := \boldsymbol{\phi}(\mathbf{x}_i, \mathbf{y}_i) \xi(\mathbf{x}_i) \prod_{h \in \Gamma(i) \setminus j} m_{hi}^{n-1}(\mathbf{x}_i). \tag{5}$$

The second phase involves integrating the combination of the above product $\boldsymbol{\pi}_{i,j}^n(\mathbf{x}_i)$ with the conditional interaction term. In [14], [8] Gibbs sampling technique was used to solve the first phase and the second phase was handled using stochastic integration and further, represented the messages non-parametrically using a kernel density estimate as,

$$m_{ij}(\mathbf{x}_j) = \sum_{m=1}^{M} w_j^m \mathcal{N}\left(\mathbf{x}_j; \boldsymbol{\mu}_j^m, \boldsymbol{\Lambda}_j^m\right) \tag{6}$$

where, w_j^m, $\boldsymbol{\mu}_j^m$, $\boldsymbol{\Lambda}_j^m$ correspond to the weight, mean and covariance associated with the mth kernel.

3. IMAGE RESTORATION WITH NON-HOMOGENOUS FIELD ASSUMPTION

In this section, we will elaborate the motivation behind the non-homogenous field assumption. First we will elaborate our method for learning the association and interaction potentials.

3.1. Learning the Association and Interaction Potential

We model the association potential $\phi(\mathbf{x}_i, \mathbf{y}_i)$ as a function over the vectorized patch association as shown in Fig. 2 and with the form

$$\phi_{\mathbf{x}_i, \mathbf{y}_i} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{N}([\mathbf{x}_i, \mathbf{y}_i]; \boldsymbol{\mu}_m, \boldsymbol{\Lambda}_m). \tag{7}$$

where M is the number of components and $\mathcal{N}([\mathbf{x}, \mathbf{y}]; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ 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 version are constructed. The patch association vectors are pruned to avoid redundancy. 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 out cross validation technique [13]. The interaction potential $\psi(\mathbf{x}_i, \mathbf{x}_j)$ is a function over the vectorized two pixel thick non overlapping patch boundary as shown in Fig. 3 and learned using the above mentioned non-parametric estimation technique.

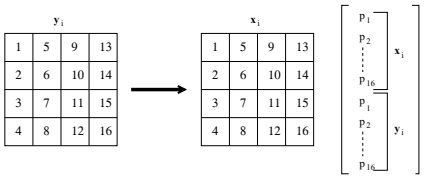

Fig. 2. The vectorized pixels in patch $\mathbf{x}_i$ are appended onto the vectorized pixels in patch $\mathbf{y}_i$ to obtain a feature vector. The association potential $\phi(\mathbf{x}_i, \mathbf{y}_i)$ is a function over this feature vector.

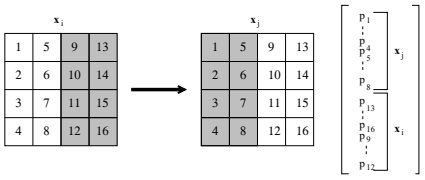

Fig. 3. The vectorized boundary pixels of patches $\mathbf{x}_i$ and $\mathbf{x}_j$ are appended to obtain a feature vector. The interaction potential $\boldsymbol{\psi}(\mathbf{x}_i, \mathbf{x}_j)$ is a function over this feature vector.

3.2. Non-homogenous field model

For the restoration problem if we want to learn the association and interaction potentials assuming a homogenous field, then we need to generate the feature vectors for the entire training data and for each hidden node in the field we need to search this huge feature space. This is the prime motivation behind a non-homogenous field model. If we assume that we have a particular digit in an image, then we can learn the compatibility functions from a particular subset of our training data set, namely those image pairs which have that particular digit. This essentially means segmenting the images before restoration, assuming there are only digits in the image, which is the case with most vehicle registration plates. The non-homogenous field modelling does exactly this in a more subtle way. We slide a suitably sized window over the blurred image and then by a suitable similarity measure ascertain which images in the training data set are closer to the image inside the window. Then we use this subset of the training data set to learn the compatibility functions. This is a tremendous saving since we are left with only a few sharp-blurred image pairs to work with instead of the entire training data set. In Fig. 4 the top image shows the full plate. The bottom image shows the sliding window and the similar digits found from the training data set.

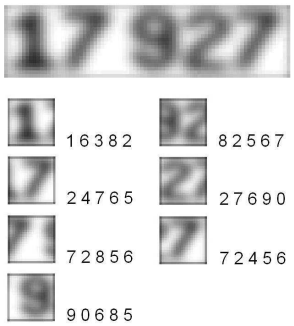

Fig. 4. Top image shows the original blurred registration plate. The lower image shows the sliding window capture, with the similar digits from the training data set.

4. EXPERIMENTS AND RESULTS

For this work we performed several experiments on synthetic images of digits and actual license plate images. For the high-res training data set set we used digit fonts from 20 different font families. The digits are represented by same-size, center-aligned binary images. The low-res training data set consists of corresponding gray-scale images obtained by convolution with a Laplacian kernel. The real data set consists of blurred binary registration plate images. To evaluate the results of the restoration algorithm we subjectively inspect the visual quality of the restored images.

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 Fig. 5 we illustrate the restoration of blurred digits "1" and "2". The reconstructed high-res images were almost indistinguishable from the originals (for digits "1" and "6") or very close to the original (for digit "2").

We use correlation as the similarity measure. The test data is sorted and the top few contenders are selected to learn the potentials. Finally we present test results on real license plates images as shown in Fig. 6. The original license plate, the blurred version of the digits, the deconvolution based deblurring result as well as our results are shown. It

Fig. 5. Left image is the input to the system. Center image is the sharpened image using deconvolution methods. Right image is the output of our algorithm. Testing samples are taken from the training ensemble.

is fairly clear that our method works well for this scenario.

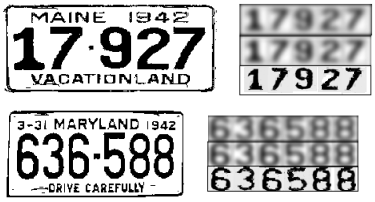

Fig. 6. Left: The original License Plate. Right: (top to bottom) Blurred input, deblurred using deconvolution methods, our method.

5. CONCLUSION AND FUTURE WORK

In this work, we use Non-parametric Belief Propagation to perform restoration of images of digits which have been blurred using an unknown kernel. The main contribution of this work is the framework where we minimize the search space for image restoration by subdividing the images by overlapping windows and then learning the compatibility functions for the individual window images.

A 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 superresolution. Inter-frame relationships can be learnt for applications like frame rate enhancement.

6. REFERENCES

[1] S. Baker and T. Kanade. Limits of super-resolution and how to break them. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, volume 24, pages 1167–1183, 2002.

[2] C. M. Bishop, A. Blake, and B. Marthi. Super-resolution enhancement of video. In *Proc. Artificial Intelligence and Statistics*, 2003.

[3] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1 of *II*, pages 627–634, 2001.

[4] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *Int'l J. of Computer Vision*, 20(1):25–47, 2000.

[5] G.Dedeoglu, T. Kanade, and J. August. High-zoom video hallucination by exploiting spatio-temporal regularities. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 151–158, 2004.

[6] R. C. Hardie, K. J. Barnard, and E. E. Armstrong. Joint map registration and high-resolution image estimation. In *IEEE Trans. Image Processing*, volume 6, pages 1621–1338, 1997.

[7] M. Irani and S. Peleg. Improving resolution by image restoration. In *CVGIP: Graph. Models Image Processing*, volume 53, pages 231–239, 1991.

[8] M. Isard. Pampas: real-valued graphical models for computer vision. In *Proceedings of Computer Vision and Pattern Recognition 2003*, volume 1, pages 613–620, 2003.

[9] S. Kim, N. Bose, and H. Valenzuela. Recursive reconstruction of high resolution image from noisy, undersampled multiframes. In *IEEE Trans. Acoustics, Speech, and Signal Processing*, volume 38, pages 1013–1027, 1990.

[10] C. Liu, H. Y. Shum, and C. S. Zhang. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1 of *I*, pages 192–198, 2001.

[11] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., 1988.

[12] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. In *IEEE Trans. Image Processing*, volume 5, pages 996–1011, 1996.

[13] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. CRC Press, 1986.

[14] E. B. Sudderth, A. T. Ihler, W. T. Freeman, and A. S. Willsky. Nonparametric belief propagation. In *Proceedings of Computer Vision and Pattern Recognition 2003*, volume 1, pages 605–612, 2003.