Blind image deconvolution by neural recursive function approximation

Jiann-Ming Wu, Hsiao-Chang Chen, Chun-Chang Wu, and Pei-Hsun Hsu

***Abstract*—**This work explores blind image deconvolution by recursive function approximation based on supervised learning of neural networks, under the assumption that a degraded image is linear convolution of an original source image through a linear shift-invariant (LSI) blurring matrix. Supervised learning of neural networks of radial basis functions (RBF) is employed to construct an embedded recursive function within a blurring image, try to extract non-deterministic component of an original source image, and use them to estimate hyper parameters of a linear image degradation model. Based on the estimated blurring matrix, reconstruction of an original source image from a blurred image is further resolved by an annealed Hopfield neural network. By numerical simulations, the proposed novel method is shown effective for faithful estimation of an unknown blurring matrix and restoration of an original source image.

***Keywords*—**blind image deconvolution, linear shift-invariant (LSI), linear image degradation model, radial basis functions (RBF), recursive function, annealed Hopfield neural networks.

I. INTRODUCTION

THE general image deconvolution is to reconstruct the original image from given degraded observations and a blurring matrix. Some authors [1] preferred to address cases for semi-blind deconvolution, since the blurring matrix in many applications was unknown or partially known. Blind deconvolution refers to a task that is expected to reconstruct an original source image from only degraded observations. Moreover, the degradation is generally nonlinear (due to saturation, quantization, etc.) and spatially varying (lens imperfections, nonuniform motion, etc.). However in many imaging applications, degraded observations could be estimated by a linear spatially invariant (LSI) blur, which are also known as the point-spread function (PSF). For medical imaging and astronomical imaging, the PSF could be unknown or partially known. Reviews on major existing approaches for blind and semi-blind deconvolution can be found in [2] and [3].

Blind deconvolution is related to blind equalization as applied to process single channel temporal observations for digital communications. The goal of blind equalization is to reconstruct transmitting signals from given single-channel observations. The degradation is linear convolution through an unknown filter. Typical approaches to blind equalization are to develop an inverse filter. Convolving single channel observations through an effective inverse filter is expected to reconstruct transmitting signals. In [4] the RBF (radial basis functions) neural network was employed to emulate non-linear dynamics of transmitting signals. Installed at a receiver, it helps to predict the inverse filter output and provide targets to adapt an inverse filter. In [5] blind deconvolution is approached by supervised learning of multilayer neural networks. The authors applied learning multilayer neural networks to extract an embedded nonlinear recursive function from observations and employed it to estimate an unknown transmitting filter. This work further extends recursive function approximation for blind image deconvolution based on supervised learning of multilayer neural networks.

In two-dimensional cases, some existing approaches to blind image deconvolution consider PSF partially known. The method addresses on blur identification based on assumption that elements in a blurring matrix are sampled from a two-dimensional Gaussian distribution. A given degraded image was transformed to frequency domain and factorized by minimizing the Kullback-Leibler (KL) divergence [6] for blur identification. With partially known PSF some variational methodologies to solve the blind deconvolution problem in Bayesian formulation have been presented in [7]-[11].

To estimate unknown PSF, this work explores neural network based methodologies for blur estimation and source restoration. The supervised learning of neural networks are applied to retrieve spatially variant nonlinear recursive structures from a blurred image, then employing them to estimate a globally invariant linear convolutive structure. At restoration phase, Hopfield neural networks are employed to restore a source image from a blurred image based on an estimated blurring matrix. The advantages of our approach for estimation of an unknown blurring matrix leaning are mainly oriented from powerful supervised learning of multilayer neural networks for recursive function approximation. Blind image deconvolution addressed here is without given prior informations about PSF. Numerical simulations show Hopfield neural networks are effective for restoration of binary and gray source images based on an estimated blurring matrix.

J.-M Wu is with the Department of Applied Mathematics, National Dong Hwa University, Hualien 941, Taiwan. Phone: +886-3-8633531 (e-mail: jmwu@mail.ndhu.edu.tw).

H.-C Chen was with the Department of Applied Mathematics, National Dong Hwa University, Hualien 941, Taiwan.

C.-C Wu is with the Department of Applied Mathematics, National Dong Hwa University, Hualien 941, Taiwan (e-mail: d9611002@mail.ndhu.edu.tw).

P.-H. Hsu is with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei 100, Taiwan.

II. A NOVEL NEURAL APPROACH FOR BLIND IMAGE DECONVOLUTION

A. Problem statement

Let $S_{M\times N}$ and $X_{M\times N}$ respectively represent an original source image and a degraded image (observed image, or blurred image) with elements denoted by $s[m,n]$ and $x[m,n]$. The blurring matrix $H_{(2\tau+1)\times(2\tau+1)}$ with elements denoted by $h[i,j]$ is linear and shift-invariant. Under assumption that a blurred image is linear convolution of a source image through a blurring matrix, induces the following equation

$$\begin{aligned} x[m,n] &= h[m,n]*s[m,n]+r[m,n] \\ &= \sum_{i=-\tau}^{\tau}\sum_{j=-\tau}^{\tau} h[i,j]s[m-i,n-j]+r[m,n], \text{ for } n,m\geq\tau \end{aligned} \quad (1)$$

where $*$ denotes the two-dimensional linear convolution operator, $h[m,n]$ denotes the blurring matrix of the degrading system and $r[m,n]$ denotes noise. Equation (1) states a linear image degradation model. Without provided informations about H, blind image deconvolution aims to reconstruct S for given X. Therefore, the original image $s[m,n]$ must be estimated directly from the degraded image $x[m,n]$. The process of the linear image degradation model is depicted in Fig. 1.

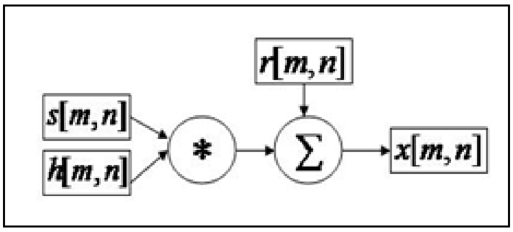

Fig. 1: The process of the linear image degradation model. X is generated by linear convolution of an original source image S through a LSI blurring matrix.

B. Nonlinear recursive function embedding

Let $\tilde{x}[m,n]$ denote collection of elements represented by $x[m+u,n+v]$ with u and v running from $-\tau$ to τ except for $u=v=0$. It covers elements within a $(2\tau+1)\times(2\tau+1)$ window centered at pixel (m,n) on image X. Blind image deconvolution is realized by applying variant RBF learning methods to construct nonlinear recursion embedded within given X. The details of the two methods to train a RBF neural network are given in Appendix. The nonlinear recursive function is expressed by

$$x[m,n]=G\left(\tilde{x}[m,n]\middle|\theta\right)+e[m,n] \quad (2)$$

where G denotes a mapping from $R^{(2\tau+1)^2-1}$ to R, θ denotes collection of its hyper parameters and $e[m,n]$ denotes a non-deterministic component. When G is linear, (2) reduces to describe the traditional linear auto-regression model.

Let $E(\theta)$ denote the mean square error of approximating $x[m,n]$ by $\tilde{x}[m,n]$ for all m and n. $E(\theta)$ is minimized with respect to θ to determine nonlinear recursion embedded within X. Minimization of $E(\theta)$ with respect to θ is the goal of supervised learning of RBF neural networks.

Let $G(\tilde{x}|\theta)$ denote the output of an RBF neural network and θ denote collection of centers and variances of radial basis functions and posterior weights. Let $D=\{(\tilde{x}[m,n],x[m,n])\}$ denote paired training data for supervised learning of an RBF network. Effective methods for supervised learning of an RBF network refer to previous works [15] [16]. If the supervised learning method is effective for minimizing $E(\theta)$, the obtained optimal network parameters,

$$\theta_{opt}=\min_{\theta} E(\theta)$$

will induce an optimal recurrence relation that can be applied to extract deterministic components from X. Let

$$\hat{x}[m,n]=G(\tilde{x}[m,n]|\theta_{opt}) \quad (3)$$

$\hat{x}[m,n]$ represents a deterministic component of $x[m,n]$. This component comes from two possibilities. One is an embedded recurrent relation within the original source S and the other is oriented from linear convolution described by (1). Subtracting $\hat{x}[m,n]$ from $x[m,n]$ attains an approximation to the non-deterministic component in (2). The approximation can be expressed by

$$\hat{e}[m,n]=x[m,n]-\hat{x}[m,n] \quad (4)$$

The estimated non-deterministic component is partly contributed by transmitting noise $r[m,n]$ as described by (1) and partly oriented from S. In other words $\hat{e}[m,n]$ must cover those belonging source S that could not be deterministically approximated by neighboring pixels. By the argument, all $\hat{e}[m,n]$ must contribute to an blurred image X through the convolutive structure described in (1). This provides a cure to estimate the unknown convolutive structure H faithfully. The process of deriving nonlinear recursion embedded within X is depicted in Fig. 2.

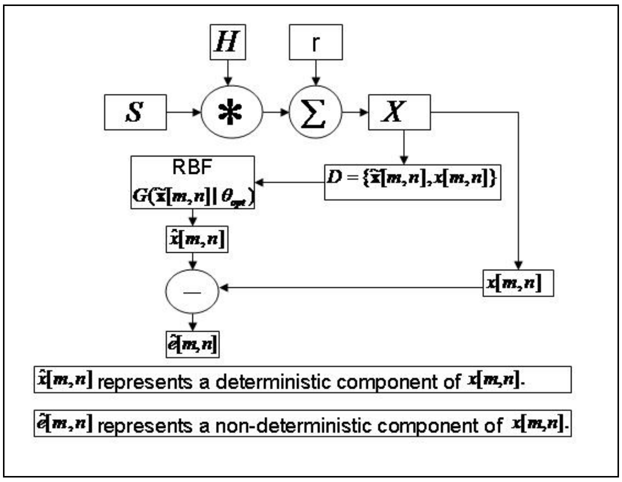

Fig. 2: X is generated by linear image convolution of an independent source. Its deterministic component is estimated by recursive function approximation based on RBF learning. The difference between the networks output and the desired output approximates the independent source from which X is oriented.

C. Estimation of an unknown blurring matrix

Non-deterministic components in S detected by $\hat{e}[m,n]$ contribute to X through an unknown convolutive structure. By the argument, the unknown convolutive structure can be estimated by minimizing the following mean square error,

$$C(h)=\frac{2}{(M-\tau+1)(N-\tau+1)}\sum_{m\geq\tau}^{M}\sum_{n\geq\tau}^{N}\{x[m,n]-\sum_{i=-\tau}^{\tau}\sum_{j=-\tau}^{\tau}h[i,j]\hat{e}[m-i,n-j]\}^2$$

where $\hat{e}[m,n]$ equals $x[m,n]-\hat{x}[m,n]$. Since C quadratically depends on elements in H, its derivative with respect to each $h[i,j]$ linearly depends on elements in H. Setting

$$\frac{dC}{dh[p,q]}=0 \qquad (5)$$

where p and q run from $-\tau$ to τ, forms a linear system whose solution estimates parameters of the linear image degradation model. The proposed novel method for reconstruction of the nonlinear recursion embedded within a given degraded image and estimation of the linear image degradation model is summarized by the following stepwise procedure.

1. Input $X_{M\times N}=\{x[m,n]\}$.
2. Form paired data $(\tilde{x}[m,n], x[m,n])$ for all m,n.
3. Train an RBF neural network to minimize $E(\theta)$ subject to given paired data and set the obtained network parameter to θ_{opt}.
4. Set $\hat{e}[m,n]=x[m,n]-\hat{x}[m,n]$ for all m and n, where $\hat{x}=G(\tilde{x}|\theta_{opt})$.
5. Solve the linear system defined by equation (5) to determine all $h[i,j]$.

With an estimated blurring matrix and a degraded image, the restoration of an original source image is explored in the upcoming sections. In above procedure, if X is a large image, it can be decomposed to several sub-images. At step 3 each sub-image can be processed to attain its local recursive relation. At step 5 over the whole image all $\hat{e}[m,n]$ are employed to estimate a global convolutive structure.

X is generated by linear convolution of source S through blurring matrix H and added with error r. There are two cases of source S. If the pixel of S is independent of its nearby pixels, it is composed of non-deterministic components (NDC) only. Otherwise the pixel of S could have contributions of its nearby pixels through shift invariant linear convolution. In the occasion, S can be decomposed into non-deterministic components (NDC) and deterministic components (DC). The linear convolutions of these two types of sources are respectively depicted in figures 3 and 4, where deterministic components are assumed as results through linear or nonlinear convolution structures. The result of the process in fig. 2 denoted by $\hat{e}[m,n]$ can be employed to approximate NDC in S.

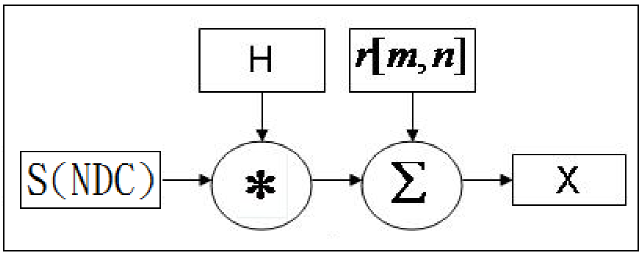

Fig. 3: The original source image S is composed of NDC only.

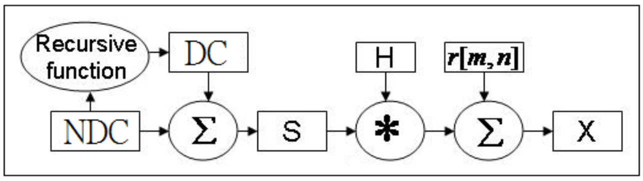

Fig. 4: Linear convolution of an original source image S through a LIS blurring matrix H where S can be decomposed into NDC and DC.

III. Image Restoration by Hopfield Neural Networks

For blind equalization, given an estimated convolutive structure A, the independent source could be retrieved from single channel observations by the following matlab built-in function,

$$[Q] = \mathrm{fdeconv}(signal, A)$$

where Q, signal and A respectively represent the estimated independent source, single channel observations and an estimated convolutive structure. For blind image deconvolution, given an estimated convolutive structure, there exists no public matlab function to estimate an original source from a blurred image. This section develops two methods to solve this problem. Fig. 5 gives an example to illustrate determining S for given H and X. Row major enumeration attains three column vectors for representing matrices S, H and X. Each entry in X induces a linear equation, which is expressed by matlab instructions, matrix inner product and matrix sum over all elements,

$$sum(\bar{s}[m,n] \cdot H) = x[m,n] \tag{6}$$

, where $\bar{s}[m,n]$ denotes a $(2\tau+1)\times(2\tau+1)$ sub-matrix of S with center at the entry $s[m,n]$. As shown in Fig. 5, dummy entries of S are filled with zeros. The source image S is extended to 7×7 matrix in Fig. 5 for its restoration based on the blurring matrix estimated by (5).

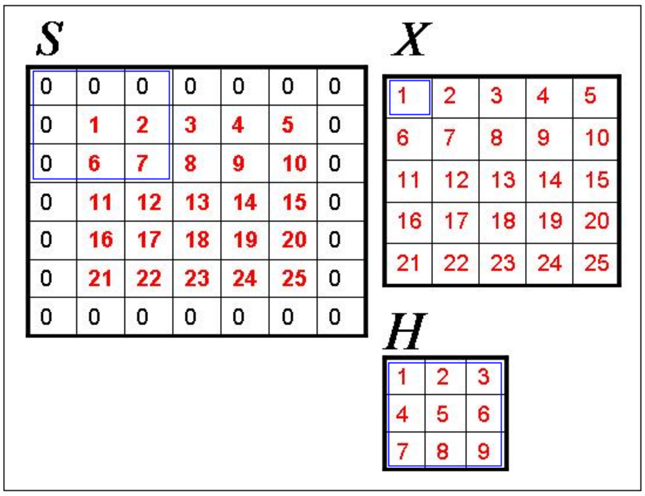

Fig. 5: S is extended to have surrounding zeros before for its convolution through H to form X. For example, the first element of X is generated by the product of upper left 3×3 sub-matrix of S to H.

There are M×N entries in S and induce M×N linear equations, which form a linear system whose solution estimates M×N unknowns in S. If the blurring matrix H is given, image deconvolution is to solve the linear system defined by (6). But this method needs an accurate estimation to H and has low tolerance for slightly perturbed H. With only an estimation to H instead of an actual H that has been used to form X, solving the linear system defined by (6) may result in disqualified restoration of H.

An annealed Hopfield neural network (12) and (13) is devised to restore an original source image from an estimated convolutive structure and a blurred image. Let $S_{mn} \in \{\pm 1\}$ denote an element in image S. Let X_{mn} and H_{ij} respectively denote elements in X and H. An estimated original binary source image is expected to minimize the following approximating error,

$$E(S) = \sum_m \sum_n [(\sum_{i=-\tau}^{\tau} \sum_{j=-\tau}^{\tau} H_{ij} S_{m-i,n-j} - X_{mn})^2 - \sum_{i=-\tau}^{\tau} \sum_{j=-\tau}^{\tau} (H_{ij} S_{m-i,n-j})^2]$$

where $S_{mn} \in \{\pm 1\}$. The first term in bracket measures the square error between X_{mn} and its approximation by linear convolution of S through H. Since S_{mn} belongs $\{\pm 1\}$, S_{mn}^2 equal one. All S_{mn}^2 are subtracted to avoid self-connections in an annealed Hopfield neural network. Consider a neural system that is composed of M×N stochastic neural variables S_{mn} whose joint distribution obeys the following Boltzmann distribution

$$\Pr(S) \propto \exp(-\beta E(S))$$

where β is a temperature-like parameter. The free energy can be expressed by

$$F = \langle E(S) \rangle - \frac{1}{\beta} H(S) \tag{7}$$

where the first term denotes the mean energy and $H(S)$ denotes the system entropy,

$$H(S) = -\sum_{\{S\}} \Pr(S) \log \Pr(S)$$

By naive mean field approximation, all binary random variables in S are assumed statistically independent. Let $\langle S_{mn} \rangle$ denote the individual mean of S_{mn}. Then

$$\Pr(S_{mn} = 1) = \frac{1 + \langle S_{mn} \rangle}{2}$$

$$\Pr(S_{mn} = -1) = \frac{1 - \langle S_{mn} \rangle}{2}$$

By the independence assumption, the system entropy equals the sum of all individual entropies, such as

$$H(S) \approx \sum_m \sum_n H(S_{mn})$$

where

$$H(S_{mn}) = -\frac{1+\langle S_{mn} \rangle}{2} \log \frac{1+\langle S_{mn} \rangle}{2} - \frac{1-\langle S_{mn} \rangle}{2} \log \frac{1-\langle S_{mn} \rangle}{2} \tag{8}$$

Furthermore the mean energy can be approximated by substituting each $\langle S_{mn} \rangle$ to S_{mn} in E. The free energy in (7) can be rewritten as

$$F \approx \sum_m \sum_n [(\sum_{i=-\tau}^{\tau} \sum_{j=-\tau}^{\tau} H_{ij} \langle S_{m-i,n-j} \rangle - X_{mn})^2 - \sum_{i=-\tau}^{\tau} \sum_{j=-\tau}^{\tau} (H_{ij} \langle S_{m-i,n-j} \rangle)^2]$$
$$-\frac{1}{\beta} \sum_m \sum_n (-\frac{1+\langle S_{mn} \rangle}{2} \log \frac{1+\langle S_{mn} \rangle}{2} - \frac{1-\langle S_{mn} \rangle}{2} \log \frac{1-\langle S_{mn} \rangle}{2})$$

Let $V_{hk} = \langle S_{hk} \rangle$. The derivative of F with respect to V_{hk} is well defined. Setting

$$\frac{\partial F}{\partial V_{hk}} = 0$$

leads to the following mean field equation,

$$V_{hk} = \tanh(-2\beta[\sum_{m=h-\tau}^{h+\tau} \sum_{n=k-\tau}^{k+\tau} (\sum_{i=-\tau}^{\tau} \sum_{j=-\tau}^{\tau} H_{ij} V_{m-i,n-j} - X_{mn}) H_{m-h,n-k} - \sum_{m=h-\tau}^{h+\tau} \sum_{n=k-\tau}^{k+\tau} H_{h-m,k-n} H_{m-h,n-k} V_{hk}]) \quad (9)$$

IV. Numerical simulations

The proposed method is tested for blur estimation and source restoration. Fig. 6 - 9 shows an original source image, a blurring matrix, a blurred image, an estimated blurring matrix and a restored source image. Numerical simulations are summarized by the following stepwise procedure.

1. Use matlab built-in function imread.m to load an original source image, $S_{M\times N} = \{s[m,n]\}$.
2. Ask the user to define a blurring matrix , $H_{(2\tau+1)\times(2\tau+1)} = \{h[i,j]\}$.
3. Convolve S through H to generate a blurred image, $X_{M\times N} = \{x[m,n]\}$.
4. Form paired data $D = \{(\tilde{x}[m,n], x[m,n])\}$ from X.
5. Apply RBF learning to build an embedded recursive function within X.
6. Calculate $\hat{x}[m,n]$ for all m and n by (3).
7. Set $\hat{e}[m,n] = x[m,n] - \hat{x}[m,n]$ for all m and n.
8. Solve the linear system defined by (5) to determine all $h[i,j]$.
9. Execute the mean field equation (9) under an annealing process for source restoration.

The proposed novel method is tested for case of $\tau = 2$ and $K = 31$, where K denotes the number of hidden units in an RBF network and $(2\tau+1)\times(2\tau+1)$ denotes the size of the blurring matrix. Fig. 7 show numerical results for this example. Fig. 8 shows the blurring matrix estimated by the RBF & AdaBoost_{reg} Software [15] method.

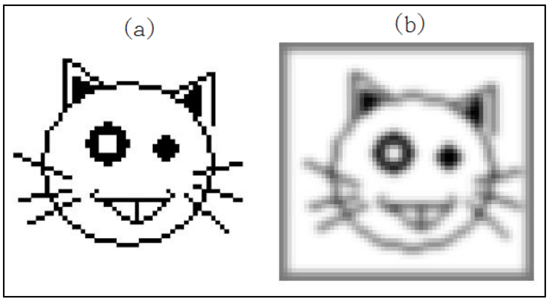

Fig. 6: (a) Original source image. (b) Blurred image is generated by linear image degradation model.

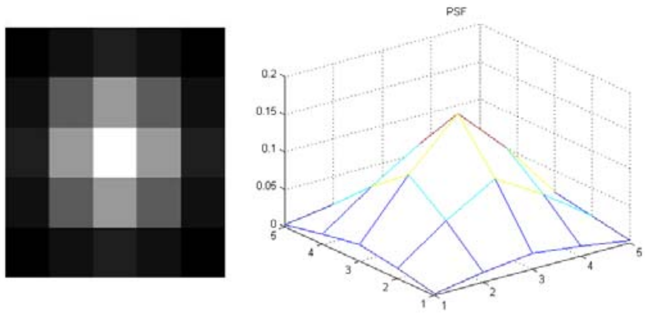

Fig. 7: The image and 3D mesh of a blurring matrix H.

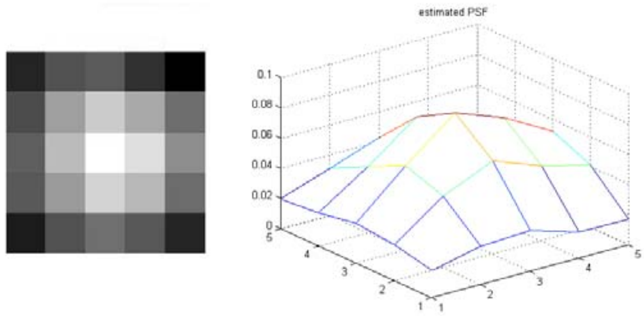

Fig. 8: The image and 3D mesh of an estimated blurring matrix.

Fig. 9: Restored source image by Hopfield neural networks.

With an exact blurring matrix, source restoration by solving (6) is shown effective. But when a given blurring matrix is added with noises or slightly perturbed, the general image deconvolution will be not effective for restoring an original source image as shown in Fig. 10. Numerical simulation show annealed Hopfield neural networks effective for source restoration based on an estimated blurring matrix in Fig.9. Fig. 11 – 14 shows the numerical result of a number plate derived by an annealed Hopfield neural network for image deconvolution.

Based on the blurring matrix estimate by recursive function approximation an annealed Hopfield neural network is applied to restore an original source image. Fig. 8 - 9 shows numerical results of the proposed novel method and Fig. 15 - 16 shows numerical results derived by the matlab built-in function, deconvblind.m, proposed in [17]. The performances of the two methods, are evaluated by the following error rate

$$\lambda = \frac{1}{2MN}\sum_m \sum_n \left| s[m,n] - \phi(\hat{s}[m,n]) \right|$$

where ϕ denotes the sign function and $\{\hat{s}[m,n]\}$ denotes an estimated original image. Table 1 lists error rates derived by relevant RBF learning methods for recursive function approximation.

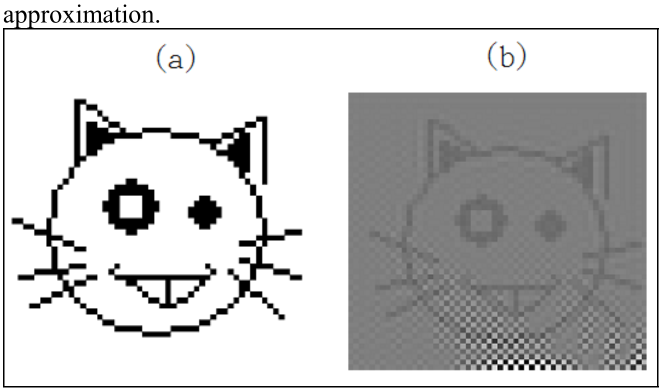

Fig. 10 (a)Restored source image with an exact blurring matrix by general image deconvolution. (b) Restored source image with a blurring matrix added noises or slightly perturbed by general image deconvolution.

Fig. 11: Original source image of a number plate.

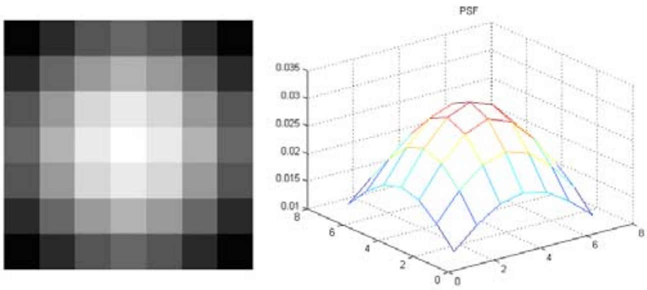

Fig. 12: The image and 3D mesh of a blurring matrix H.

Fig. 13: Blurred image of a number plate is generated by linear image degradation model.

Fig. 14: Restored source image of a number plate by Hopfield neural networks.

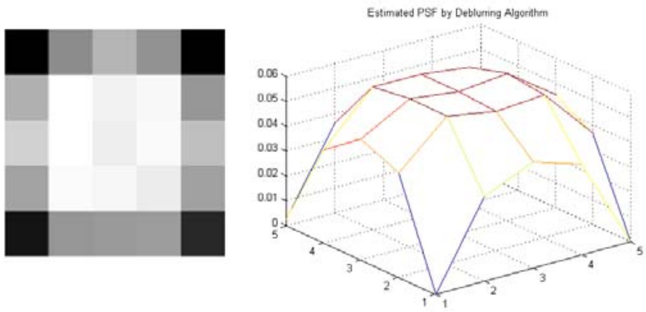

Fig. 15: The image and 3D mesh of an Estimated blurring matrix by the matlab built-in function, deconvblind.m.

Fig. 16: Restored source image by the matlab built-in function, deconvblind.m.

TABLE I
UNITS FOR MAGNETIC PROPERTIES

Method		Error rate of the cat $\lambda=\frac{1}{2MN}\sum_m\sum_n\left\|s[m,n]-\phi(\hat{s}[m,n])\right\|$
Recursive Function Approximation	k-means clustering for σ	0.0975
	k-means clustering for σ_k	0.1355
	RBF and AdaBoost-reg Software	0.0974
	Levenberg-Marquardt method	0.0980
Deblurring images using the blind deconvolution algorithm		0.2031

Error rate derived by relevant RBF learning methods for recursive function approximation and the matlab built-in function, deconvblind.m.

V. CONCLUSION

This work has shown the proposed novel method effective for faithful blur estimation and source reconstruction. The novel method applies supervised learning of RBF networks to construct an embedded recursive function within degraded observations and use them to extract the non-deterministic component of an original source image. The difference between degraded observations and the extracted deterministic component serves as the non- deterministic component. The non-deterministic component is further applied to estimate an unknown blur structure. Moreover, an annealed Hopfield neural network is applied for image deconvolution and its effectiveness is shown for source restoration based on a noisy blurring matrix.

APPENDIX

The RBF network function employed in (2) is defined by

$$\begin{aligned}\hat{x}[m,n] &= G(\tilde{x}[m,n]\,|\,\theta)\\ &= w_0+\sum_{k=1}^{K} w_k \exp\left(-\frac{\left\|\tilde{x}[m,n]-\boldsymbol{\mu}_k\right\|^2}{2\sigma_k^2}\right)\end{aligned} \tag{10}$$

where $\theta=\{w_k\}\cup\{\boldsymbol{\mu}_k\}\cup\{\sigma_k\}$ denotes collection of network parameters.

A. K-means based RBF learning

K-means clustering [14] aims to determine K means of given observations by partitioning them into non-overlapping subsets. Let $\{\boldsymbol{\xi}_i\}_{i=1}^n$ denote collection of d-dimensional real vectors. K-means clustering aims to partition them into k subsets, each denoted by Λ_j, where $K<n$, so as to minimize the within-cluster sum of squares (WCSS),

$$\Lambda^* = \arg\min_{\Lambda}\sum_{j=1}^{K}\sum_{x_i\in\Lambda_j}\left\|\boldsymbol{\xi}_i-\boldsymbol{\mu}_j\right\|^2$$

where $\boldsymbol{\mu}_j$ is the mean of elements in Λ_j. K-means clustering can be employed to initialize the network parameter θ of equation (10). Let $D=\{(\boldsymbol{\xi}_i,\mathbf{y}_i)\}_i$ denote paired training data. Two simple approaches are considered to determine network parameters other than determined K means, denoted by $\{\boldsymbol{\mu}_j\}_{j=1}^K$.

1. The first situation considers $\sigma_k=\sigma$ for all k, where σ is a real value. The means $\boldsymbol{\mu}_k$ for all k are determined by k-means clustering. A common variance is employed. Equation (10) for each $\boldsymbol{\xi}_i$ will reduce to a linear system,

$$y_i = w_0+\sum_{k=1}^{K} w_k v_{ik} \tag{11}$$

where $v_{ik}=\exp\left(-\frac{\left\|\boldsymbol{\xi}_i-\boldsymbol{\mu}_k\right\|^2}{2\sigma_k^2}\right)$. There are K unknowns, $\{w_k\}_k$, and n constraints, which constitute a linear system well resolved by the technique of pseudo inverse. Substituting all determined network parameters to

$$E_\sigma=\frac{1}{2N}\sum_{i=1}^{N}(y_i-G(\xi\,|\,\theta))^2 \tag{12}$$

attains an error. This error is minimized by selecting the best common σ.

2. The second situation employs different variances. The K-means clustering method is used to determine all $\boldsymbol{\mu}_k$. Then σ_k is determined by elements that are closest to $\boldsymbol{\mu}_k$ relative to $\boldsymbol{\mu}_j$ with j≠k. Then equation (11) is employed to determine all w_k.

B. Levenberg-Marquardt method

The Levenberg-Marquardt method [16] can be considered as a hybrid of the gradient method and the Newton-Gauss method. Let θ_i denote the network parameter at the ith iteration and $\varepsilon(t,\theta_i)=y_t-y(t\,|\,\theta_i)$ denote an approximating error. The mean square error in (12) can be rewritten as

$$E_D(\theta_i)=\frac{1}{2N}\sum_{t=1}^{N}\varepsilon(t,\theta_i)^2 \tag{13}$$

It follows

$$\nabla(\theta_i)=\frac{dE_D(\theta)}{d\theta}\Big|_{\theta=\theta_i}=-\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta_i)\Psi(t,\theta_i)$$

Where $\Psi(t,\theta)=\frac{dG(x[t]\,|\,\theta)}{d\theta}\Big|_{\theta=\theta_i}$. Moreover,

$$\Delta\theta_i \propto -\frac{dE_D(\theta)}{d\theta}\Big|_{\theta=\theta_i}=-\nabla(\theta_i) \tag{14}$$

$\varepsilon(t,\theta)$ by linear expansion at $\theta=\theta_i$ can be represented as

$$\tilde{\varepsilon}_i(t,\theta) = \varepsilon(t,\theta_i) + (\theta-\theta_i)^T \frac{d\varepsilon(t,\theta)}{d\theta}\Big|_{\theta=\theta_i}$$
$$= \varepsilon(t,\theta_i) - (\theta-\theta_i)^T \Psi(t,\theta_i)$$

So the mean square error (12) can be written as the quadratic form. That is

$$L_i(\theta) = \frac{1}{2N}\sum_{t=1}^{N} \tilde{\varepsilon}_i^2(t,\theta)$$

where $L_i(\theta) \approx E_D(\theta)$. The quadratic form by expansion at $\theta = \theta_i$ can be represented as

$$L_i(\theta) = L_i(\theta_i) + \nabla(\theta_i)^T(\theta-\theta_i) + \frac{1}{2}(\theta-\theta_i)^T R(\theta_i)(\theta-\theta_i)$$

where $R(\theta_i) = \frac{1}{N}\sum_{t=1}^{N}\Psi(t,\theta_i)\Psi^T(t,\theta_i)$ denote Gauss-Newton Hessian matrix and the gradient $\nabla(\theta_i) = -\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta_i)\Psi(t,\theta_i)$.

Moreover, the Newton-Gauss method takes the derivative of the quadratic form. That is

$$\frac{dL_i(\theta_i)}{d\theta} = \nabla(\theta_i) + R(\theta_i)(\theta-\theta_i) = 0$$

let $\Delta(\theta_i) = (\theta-\theta_i)$ can be rewriting as

$$\Delta(\theta_i) = -\left(R(\theta_i)\right)^{-1}\nabla(\theta_i) \qquad (15)$$

Therefore, the Levenberg-Marquardt method let $\Delta(\theta_i)$ is determined by

$$\left(R(\theta_i) + \lambda I\right)\Delta(\theta_i) = -\nabla(\theta_i) \qquad (16)$$

where λ is control factor and I is identity matrix. If $\lambda = 0$, equation (16) reduce to (15) and it use the Newton-Gauss method. Otherwise λ is sufficiently large, $\Delta(\theta_i)$ is determined by (14) and it use the gradient method. Under the assumption, control λ by following equation

$$\alpha_i = \frac{E_D(\theta_i) - E_D(\theta_i + \Delta\theta_i)}{E_D(\theta_i) - L_i(\theta_i + \Delta\theta_i)}$$

where $E_D(\theta_i)$ is called current parameter, $E_D(\theta_i + \Delta\theta_i)$ is called next parameter, $E_D(\theta_i) - E_D(\theta_i + \Delta\theta_i)$ represent as actual cost reduction and $E_D(\theta_i) - L_i(\theta_i + \Delta\theta_i)$ represent as predicted cost reduction. If α_i is high, λ is reduced. Otherwise λ is increased. So the heuristic adaption could be written as

$$\text{If } \alpha_i > 0.75,\ \lambda \leftarrow 0.5\lambda$$
$$\text{If } \alpha_i < 0.25,\ \lambda \leftarrow 2\lambda$$

The Leveberg-Marquardt method has been applied to learning multilayer neural networks, including multilayer perceptrons (MLP), RBF (radial basis functions) networks, and their hybrid networks.

REFERENCES

[1] L. Bar, N. Sochen and N. Kiryati, "Semi-blind image restoration via Mumford-Shah regularization," IEEE Trans. Image Process., vol. 15, no. 2, pp. 483--493, Feb 2006.
[2] D. Kundur and D. Hatzinakos, "Blind image deconvolution," IEEE Trans. Signal Process. Mag., vol. 13, no. 3, pp. 43--64, Mar. 1996.
[3] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, "Blind image deconvolution: Problem formulation and existing approaches," in Blind Image Deconvolution: Theory and Applications, P. Campisi and K. Egiazarian, Eds. Boca Raton, FL: CRC, ch. 1, 2007.
[4] Nan Xie and Henry Leung, Member, "IEEE Blind Equalization Using a Predictive Radial Basis Function Neural Network," IEEE Trans. Neural Networks., vol. 16, no. 3, May 2005.
[5] J. M. Wu and C. S. Lian, "Blind Deconvolution by Supervised Learning of Neural Networks," Master thesis, Department of Applied Mathematics, National Dong Hwa University, Jul. 2008.
[6] S. Kullback, "Information Theory and Statistics," New York: Dover, 1959.
[7] A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," IEEE Trans. Signal Process., vol. 52, no. 8, pp. 2222--2233, 2004.
[8] K. Z. Adami, "Variational methods in Bayesian deconvolution," in Proc. PHYSTAT ECONF, 2003, vol. CO30908, p. TUGT002.
[9] S. D. Babacan, R. Molina and A. K. Katsaggelos, "Variational Bayesian Blind Deconvolution Using a Total Variation Prior," IEEE Trans. Image Process., vol. 18, no. 1, Jan. 2009.
[10] R. Jirik and T. Taxt, "Two-dimensional blind Bayesian deconvolution of medical ultrasound images," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control., vol 55, no 10, pp. 2140 - 2153, Oct. 2008.
[11] D. G. Tzikas, A. C. Likas and N. P. Galatsanos, "Variational Bayesian Sparse Kernel-Based Blind Image Deconvolution With Student's-t Priors," IEEE Trans. Image Process., vol. 18, no. 4, pp. 753 - 764, Apr. 2009.
[12] J. M. Wu, "Annealing by Two Sets of Interactive Dynamics," IEEE Trans. Systems, Man, and Cybernetics, Part B., vol. 34, no. 3, pp. 1519 - 1525. Jun. 2004.
[13] J. J. Hopfield and D. W. Tank, "Neural computation of decisions in optimizationproblems," Biol. Cybern., vol. 52, pp. 141, 1985.
[14] J. B. MacQueen, (1967) . "Some Methods for classification and Analysis of Multivariate Observations" in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1: 281--297, University of California Press. MR0214227. Zbl 0214.46201. Retrieved on 2009-04-07.
[15] Gunnar Rätsch, "A (Partial) Documentation of RBF & AdaBoostReg Software Packages."
[16] M. Nørgaard, O. Ravn, N. K. Poulsen and L. K. Hansen. "Neural Networks for Modelling and Control of Dynamic Systems," London, U.K.:Springer-Verlag, 2000.
[17] Matlab toolbox "Deblurring Images Using the Blind Deconvolution Algorithm."

Jiann-Ming Wu was born in Taiwan, on November 22,1966. He received the B.S degree in computer science from Nation Chiao Tung University, Hsinchu, Taiwan, in 1988, the M.S and Ph.D. degrees in computer science and information engineering form National Taiwan University, Taipei, Taiwan in 1990 and 1994, respectively.

In 1996, he joined the Faculty as an Assistant Professor at the Department of Applied Mathematics in National Dong Hwa University, Hualien, Taiwan, where he is currently an Associate Professor. His current research interests include neural networks and signal processing. In the past decade, he has developed effective learning methods for neural networks toward solving independent component analysis (ICA), self-organization, classification, data driven function approximation, and density estimation.