A Blind Source Separation Perspective on Image Restoration

Lidan Miao and Hairong Qi
The University of Tennessee
1508 Middle Dr., Knoxville, TN 37996
{lmiao1, hqi}@utk.edu

Abstract

This paper re-investigates the physical image formation process leading to a new interpretation of the classic image restoration problem from a blind source separation (BSS) perspective. The observed distorted image is considered as a linear combination of a set of shifted version of the point spread function (PSF) with the weight coefficients determined by the actual image. The new interpretation brings two immediate benefits to the practice of image restoration. First, we can utilize the rich set of BSS methods to solve the blind image restoration problem. Second, the new formulation in terms of matrix product has the equivalent merit as the conventional matrix-vector notation in theoretical study of restoration algorithms. We develop a smoothness and block-decorrelation constrained nonnegative matrix factorization method (termed CNMF) to blindly recover both the PSF and the actual image. The experimental results compared to one of the state-of-the-art methods demonstrate the merit of the proposed approach.

1. Introduction

Image restoration is a process of recovering the original image from a degraded observed version. Denote the image domain by Ω on which the intensity function is defined, and the discrete sampling grid by $\mathbf{s} = \{(x, y); x, y = 0, 1, \ldots, M-1\}$, where M is the number of rows and columns (we assume square images for notation simplicity). Assuming the imaging system is linear and space-invariant, the degradation procedure is well described by the following model

$$g(x, y) = f(x, y) \otimes h(x, y) + n(x, y) \tag{1}$$

where $g(x, y)$, $f(x, y)$, and $n(x, y)$ denote, respectively, the observed blurred and noisy image, the actual undistorted image, and the additive white Gaussian noise. $h(x, y)$ is the point spread function (PSF) of the imaging system, and $\otimes$ denotes the convolution operator. Let $\mathbf{g}$, $\mathbf{f}$, and $\mathbf{n}$ denote the $M^2 \times 1$ column vectors obtained by lexicographically stacking $g(x, y)$, $f(x, y)$, and $n(x, y)$ into $M^2 \times 1$ vectors, and let $\mathcal{H}$ denote the $M^2 \times M^2$ block-circulant matrix with the bases determined by the elements of h, the observation model can then be expressed as

$$\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n} \tag{2}$$

Although this model has been well investigated from a mathematical point of view since the seventies [2], its physical interpretation has not been given much attention.

It is well known that recovering the image f with known PSF is a mathematically ill-posed inverse problem. A number of regularization techniques have been studied in the past decades [9, 16, 17] to tackle this problem. However, image restoration becomes more difficult if the PSF is unknown, since the problem is ill-posed with respect to not only the image but also the blur kernel [4]. To find reliable estimates, You and Kaveh [18] introduced the concept of double regularization and proposed to jointly recover the image and the blur kernel with $\mathbb{H}^1$ norm regularization of both. Following the same idea, Chan and Wong [7] investigated the TV regularization to replace the $\mathbb{H}^1$ norm and proposed a fast alternative minimization (AM) algorithm based on the lagged diffusivity fixed point scheme [17]. The convergence proof of AM algorithm for $\mathbb{H}^1$ norm is given in [8]. This method was then integrated with the Mumford-Shah segmentation in [4] to perform image segmentation and deconvolution simultaneously. Latter, Burger and Scherzer [6] showed the existence of solution for a wide class of function spaces except for the $\mathbb{L}^1$, $\mathbb{L}^2$, and $\mathbb{H}^1$ space, and introduced the notion of minimum norm (MN) solution. Recently, Justen and Ramlau [11] proved the uniqueness of solution under a weak smoothness condition as well as the existence of solution. They also derived an explicit form of minimum norm solution without any iterative learning processes and made an important observation that blind deconvolution is less ill-posed than non-blind deconvolution.

Mathematically, the formulation of Eq. 2 is a blind source separation (BSS) problem if we regard $\mathbf{g}$ as a mixture of signal $\mathbf{f}$ generated by the mixing matrix $\mathcal{H}$. Therefore, multivariate data analysis, such as independent component analysis (ICA) [1, 5], might be used to solve a restoration

1-4244-1180-7/07/$25.00 ©2007 IEEE

problem. However, the multichannel image required by BSS algorithms is not always available. Therefore, the key to the single-frame image blind deconvolution using BSS methods is the creation of multichannel data. One such approach was proposed in [12], which uses Gabor filters to produce multichannel filtering and decompose an image into sparse images. Then, a sparseness constrained nonnegative matrix factorization (NMF) method [10] is used to find the optimal estimate. Due to the absorption of the blur kernel into the mixing matrix, the algorithm cannot recover the actual blur kernel. In addition, the multiple filtered images dramatically increase the computational burden.

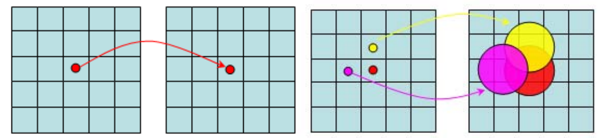
In this paper, we re-investigate the physical image formation process and interpret the optical blur as a signal mixing procedure; that is, the observed image is considered as a linear combination of a set of shifted PSF with the weight coefficients determined by the actual image. We create multichannel data based on a block-based interpretation and formulate the imaging model as a matrix product. Considering the nonnegative property of image contents, we adopt the NMF technique to recover both the actual image and the PSF. Smoothness and block-decorrelation constraints are imposed to regularize the solution. We refer to the proposed method as the constrained NMF (CNMF).

The remainder of this paper is organized as follows. In Section 2, we present the BSS formulation of the image restoration problem. Section 3 details the constrained NMF restoration algorithm. The effectiveness of the proposed method is demonstrated in Section 4 by comparing with one of the advanced restoration technique. Section 5 concludes the paper.

2. Problem Formulation

For an ideal imaging system, the response at a point (x, y) depends only on the value of input at the corresponding point as illustrated in Fig. 1(a), where the left diagram denotes the source domain, and the right diagram is the observation. However, due to the quality of optical systems, the point source will be spreaded to a certain degree represented by the PSF as demonstrated in Fig. 1(b), where the PSF is assumed to be circularly symmetric and have finite support. Thus, the measurement at a single point (e.g. the center pixel) is a mixture of several points of input. If we consider the 5×5 pixel space shown in the figure as an image block, the observed image block can then be represented as a linear combination of a set of shifted PSF weighted by the actual image values. In an extreme case, if we treat the entire image as one block, then the mixing model is exactly the same as the model in Eq. 2, which has strong and well-established mathematical foundations.

To generate multivariate data, we partition the image domain Ω into small rectangular blocks of size B. Then, we lexicographically stack the columns of each block into a



(a) (b)

Figure 1. Physical interpretation of optical blur (a) Ideal optical system (b) Practical optical system.

vector, and create a matrix by ordering all the block vectors as its columns, thus the degradation model reduces to

$$\mathbf{G} = \mathbf{HF} + \mathbf{N} \tag{3}$$

where $\mathbf{G}$, $\mathbf{F}$, and $\mathbf{N} \in \mathbf{R}^{B^2 \times (\lfloor \frac{M}{B} \rfloor)^2}$ are the block vector representations of the degraded image, the actual image, and the additive white Gaussian noise. The matrix $\mathbf{H} \in \mathbf{R}^{B^2 \times B^2}$ is a block-circulant matrix.

For example, if we have a simple 6×6 image g, and assume the blur kernel h is of size 2×2, and the block size is 3×3:

$$g = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{16} \\ g_{21} & g_{22} & \cdots & g_{26} \\ \vdots & \vdots & \vdots & \vdots \\ g_{61} & g_{62} & \cdots & g_{66} \end{bmatrix}, \quad h = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}$$

the degraded image matrix is then given by

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{21} & g_{31} & g_{12} & g_{22} & g_{32} & g_{13} & g_{23} & g_{33} \\ g_{14} & g_{24} & g_{34} & g_{15} & g_{25} & g_{35} & g_{16} & g_{26} & g_{36} \\ g_{41} & g_{51} & g_{61} & g_{42} & g_{52} & g_{62} & g_{43} & g_{53} & g_{63} \\ g_{44} & g_{54} & g_{64} & g_{45} & g_{55} & g_{65} & g_{46} & g_{56} & g_{66} \end{bmatrix}^T$$

The actual image $\mathbf{F}$ is constructed in the same way, and the block-circulant matrix has the following structure,

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{0} & \mathbf{H}_2 \\ \mathbf{H}_2 & \mathbf{H}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \mathbf{H}_1 \end{bmatrix}, \quad \mathbf{H}_i = \begin{bmatrix} h_{1i} & 0 & h_{2i} \\ h_{2i} & h_{1i} & 0 \\ 0 & h_{2i} & h_{1i} \end{bmatrix}$$

Each column of $\mathbf{H}$, when reshaped into a square matrix, is a shifted version of the PSF. Compared with the blur matrix $\mathcal{H} \in \mathbf{R}^{M^2 \times M^2}$ in Eq. 2, they have the same structure, but $\mathbf{H}$ is more dense with less number of zeros and a much reduced matrix size ($M \gg B$). Fig. 2 illustrates an example of 144 shifted PSFs of size 12×12. Each block corresponds to one column vector of $\mathbf{H}$.

It can be seen that the multichannel data are obtained from the observed image alone without any filtering, which results in a small data volume making fast computation possible. When written in the matrix product form, the restoration problem has been converted to a problem of blind source separation; that is, given the observed image $\mathbf{G}$, the objective is to find the underlying source $\mathbf{H}$ and the mixing matrix $\mathbf{F}$ such that $\mathbf{G} \approx \mathbf{HF}$. Both $\mathbf{H}$ and $\mathbf{F}$ are nonnegative matrices.

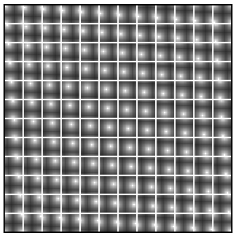


Figure 2. Illustration of a 144×144 $\mathbf{H}$ matrix. Each block, a reshaped version of the column of $\mathbf{H}$, is a shifted PSF.

3. Constrained NMF for Image Restoration

The solutions to a BSS problem can be found in a number of ways, such as ICA [1, 5], NMF [14, 15], and independent factor analysis (IFA) [3]. For ICA and IFA, the recovered sources are not guaranteed to be nonnegative, and a critical condition for them to work is that the sources have to be statistically independent. Moreover, ICA requires the blur kernel to be non-stationary as the blur caused by the atmospheric turbulence but not the case for the out-of-focus blur. NMF is a popular matrix factorization method, which aims at rendering part-based and nonnegative representations [13]. It has found a wide range of applications in data analysis, dimensionality reduction, feature extraction, and target recognition.

From the aspect of NMF, the columns of $\mathbf{H}$ are considered as the basis vectors, which are the intrinsic structures of the underlying data, and the columns of $\mathbf{F}$ are the corresponding weight coefficients to form the blurred image. To measure the quality of the factorization approximation $\mathbf{G} \approx \mathbf{HF}$, a cost function between $\mathbf{G}$ and $\mathbf{HF}$ needs to be optimized subject to the nonnegative constraint. Besides the widely used Euclidean distance, another popular measure is the divergence between $\mathbf{G}$ and $\mathbf{HF}$ [14] defined as

$$D(\mathbf{G}||\mathbf{HF}) = \sum_{ij} \left(g_{ij} \log \frac{g_{ij}}{\sum_k h_{ik} f_{kj}} - g_{ij} + \sum_k h_{ik} f_{kj} \right) \tag{4}$$

A standard NMF problem is to minimize the above cost function under the nonnegative constraint.

3.1. Regularization

One hurdle of solving the NMF problem is the existence of local minima due to the non-convexity of the objective function. To stabilize the solution, extra constraints need to be imposed, which could wrap the original objective function to approach a convex function so that the solution converges to the global optimum. As in most restoration algorithms, a roughness penalty will be added to the recovered image to achieve local smoothness. In this paper, we consider a nonlinear constraint based on the negative Gaussian distribution. The energy function is defined as

$$U_1(\mathbf{F}) = -\sum_{ij} \exp(-\frac{[\mathbf{CF}]_{ij}^2}{2\tau^2}) \tag{5}$$

where $\mathbf{C}$ is also a block-circulant matrix constructed from a high pass filter in the same way as the blur matrix $\mathbf{H}$. In our implementation, a Laplace filter is adopted. The parameter τ is the variance of the Gaussian model. The superiority of this nonlinear constraint is that it treats high frequency information differently based on their magnitudes. While small τ penalizes only small variations and considers most variations as edge details, large τ penalizes most variations and only keeps strong edge transitions. In other words, we implicitly make the assumption that a large value of the variation corresponds to a true edge while a small value of the variation is an effect of noise. This assumption is not necessarily satisfied in some real world applications.

Besides the local smoothness, another regularization we investigate is the correlation between image blocks, which is a direct benefit brought by the block-based formulation. We assume that different image blocks are uncorrelated especially for images corrupted by strong noise, then the correlation matrix $\mathbf{FF}^T$ should be diagonal dominant. For this purpose, we choose to minimize the block correlation, which is formulated as

$$U_2(\mathbf{F}) = \frac{1}{B^2} \sum_{i,j \neq i}^{B^2} [\mathbf{FF}^T]_{ij} - \frac{1}{B^2} \sum_{i=1}^{B^2} [\mathbf{FF}^T]_{ii} \tag{6}$$

To provide a reliably recovered PSF, additional constraints on $\mathbf{H}$ need to be imposed besides the nonnegativity. Considering the block-circulant structure of $\mathbf{H}$ and the unit $\mathbb{L}^1$ norm property of PSF, we know that the sum of columns and the sum of rows of $\mathbf{H}$ equal 1, that is

$$\sum_j h_{ij} = 1, \ \sum_i h_{ij} = 1, \quad i, j = 1, 2, \ldots, B^2 \tag{7}$$

Moreover, the block-circulant structure needs to be incorporated into the learning process. For this purpose, we start from a $B^2 \times B^2$ block-circulant matrix constructed based on a kernel estimator, then we develop a multiplicative rule to update $\mathbf{H}$ so that the block circulant nature can be preserved.

3.2. Learning Algorithms

Combining the objective of minimizing the fitness error to the observation and incorporating the local smoothness and block-decorrelation constraint, we arrive at the following optimization problem

$$\begin{aligned} \text{min.} \quad & \mathcal{L}(\mathbf{H}, \mathbf{F}) = D(\mathbf{G}||\mathbf{HF}) + \lambda_1 U_1(\mathbf{F}) + \lambda_2 U_2(\mathbf{F}) \\ \text{s.t.} \quad & \mathbf{H1} = \mathbf{1}, \mathbf{H}^T\mathbf{1} = \mathbf{1} \\ & \mathbf{H} \succeq \mathbf{0}, \mathbf{F} \succeq \mathbf{0} \end{aligned} \tag{8}$$

where the energy function $U_1(\mathbf{F})$ and $U_2(\mathbf{F})$ are given by Eq. 5 and Eq. 6, and λ_1, λ_2 are the corresponding regularization parameters, which control the tradeoff between the confidence on the observation and the regularity of solutions. The automatic selection of these parameters has been an active research area. In this paper, an empirical constant will be used. The symbol $\succeq$ denotes *componentwise inequality*, e.g., $\mathbf{H} \succeq \mathbf{0}$ means $h_{ij} \geq 0$ for $i = 1, 2, \dots, B^2$, $j = 1, 2, \dots, B^2$.

To consider the row-sum-to-one constraint, we augment the observed and the actual image by a column of constant denoted by

$$\mathbf{G}^+ = \left[\begin{array}{cc} \mathbf{G} & \theta\mathbf{1} \end{array}\right], \quad \mathbf{F}^+ = \left[\begin{array}{cc} \mathbf{F} & \theta\mathbf{1} \end{array}\right] \tag{9}$$

where $\mathbf{1}$ is a $B^2 \times 1$ column vector with all elements being 1s. θ is a positive number to control the effect of the sum-to-one constraint. The learning of $\mathbf{H}$ takes these two augmented matrices as inputs. Then, the columns of $\mathbf{H}$ are normalized to satisfy the column-sum-to-one constraint.

To find the optimal solution, we resort to the technique of auxiliary function [14] to derive the learning rules. For a given objective function $\mathcal{L}(\mathbf{X})$, its auxiliary function is defined as a function $\Phi(\mathbf{X}, \mathbf{X}')$ which satisfies $\Phi(\mathbf{X}, \mathbf{X}) = \mathcal{L}(\mathbf{X})$ and $\Phi(\mathbf{X}, \mathbf{X}') \geq \mathcal{L}(\mathbf{X})$, where $\mathbf{X}'$ is a variable different from $\mathbf{X}$. The auxiliary function $\Phi(\mathbf{X}, \mathbf{X}')$ is very helpful when minimizing the corresponding objective $\mathcal{L}(\mathbf{X})$ in the sense that $\mathcal{L}(\mathbf{X})$ is non-increasing under the update

$$\mathbf{X}^{k+1} = \arg\min_{\mathbf{X}} \Phi(\mathbf{X}, \mathbf{X}^k) \tag{10}$$

The proof is easy to see, as $\mathcal{L}(\mathbf{X}^{k+1}) \leq \Phi(\mathbf{X}^{k+1}, \mathbf{X}^k) \leq \Phi(\mathbf{X}^k, \mathbf{X}^k) = \mathcal{L}(\mathbf{X}^k)$ [14].

We employ the popular alternating minimization (AM) scheme to update $\mathbf{H}$ and $\mathbf{F}$ to minimize the objective function. AM is an iterative scheme, which updates one matrix while holding the other one fixed; that is, starting from an initial guess $\mathbf{F}^0$, the optimal $\mathbf{H}^1$ is calculated by minimizing $\mathcal{L}$ with fixed $\mathbf{F}^0$. Then $\mathbf{F}^1$ is obtained by minimizing $\mathcal{L}$ with fixed $\mathbf{H}^1$,

$$\mathbf{H}^{k+1} = \arg\min_{\mathbf{H}\succeq 0} \mathcal{L}(\mathbf{H}, \mathbf{F}^k), \mathbf{F}^{k+1} = \arg\min_{\mathbf{F}\succeq 0} \mathcal{L}(\mathbf{H}^{k+1}, \mathbf{F})$$

This process is continued until the desired stop conditions are satisfied. By employing the auxiliary function technique, we derive the iterative update rules given in the following theorem.

Theorem 1 *Given a small positive λ_2, the regularized cost function in Eq. 8 is monotonically nonincreasing under the following update rules:*

$$\begin{aligned} f_{st} &= \frac{-b + \sqrt{b^2 - 4ac}}{2a} \\ h_{st} &= h_{st} \frac{\sum_i f_{ti}^+ g_{si}^+ / [\mathbf{HF}^+]_{si}}{\sum_i f_{ti}^+}, \quad h_{st} = \frac{h_{st}}{\sum_s h_{st}} \end{aligned} \tag{11}$$

where

$$a = -\frac{\lambda_2}{B^2}, \quad c = -\sum_i g_{it} \frac{h_{is} f_{st}}{\sum_r h_{ir} f_{rt}}$$

$$b = \sum_i \left(h_{is} + \frac{\lambda_1}{\tau^2} c_{is} [\mathbf{CF}]_{it} \exp(-\frac{[\mathbf{CF}]_{it}^2}{2\tau^2}) \right) + \frac{\lambda_2}{B} \sum_{i=1, i\neq s}^{B^2} f_{it}$$

f_{ti}^+ and g_{ti}^+ are the elements of the augmented matrices $\mathbf{G}^+$ and $\mathbf{F}^+$ given in Eq. 9.

Proof To find the update of $\mathbf{F}$, we construct an auxiliary function for $\mathcal{L}(\mathbf{H}, \mathbf{F})$ with fixed $\mathbf{H}$ (denoted as $\mathcal{L}(\mathbf{F})$ in the following description),

$$\begin{aligned} \Phi(\mathbf{F}, \mathbf{F}') &= \sum_{ij} g_{ij} \log g_{ij} - \sum_{ij} g_{ij} + \sum_{ijk} h_{ik} f_{kj} \\ &- \sum_{ijk} g_{ij} \frac{h_{ik} f'_{kj}}{\sum_r h_{ir} f'_{rj}} \left(\log(h_{ik} f_{kj}) - \log \frac{h_{ik} f'_{kj}}{\sum_r h_{ir} f'_{rj}} \right) \\ &+ \lambda_1 U_1(\mathbf{F}) + \lambda_2 U_2(\mathbf{F}) \end{aligned} \tag{12}$$

It is easy to verify that $\Phi(\mathbf{F}, \mathbf{F}) = \mathcal{L}(\mathbf{F})$. To show $\Phi(\mathbf{F}, \mathbf{F}') \geq \mathcal{L}(\mathbf{F})$, we note that $\log(\sum_k h_{ik} f_{kj})$ is a convex function, then there exists a set of coefficient μ_{ijk} satisfying $\sum_k \mu_{ijk} = 1$ for all i, j, such that

$$\log(\sum_k h_{ik} f_{kj}) \geq \sum_k \mu_{ijk} \log \frac{h_{ik} f_{kj}}{\mu_{ijk}} \tag{13}$$

Setting $\mu_{ijk} = \frac{h_{ik} f'_{kj}}{\sum_r h_{ir} f'_{rj}}$, we immediately obtain

$$\begin{aligned} &\log(\sum_k h_{ik} f_{kj}) \\ &\geq \sum_k \frac{h_{ik} f'_{kj}}{\sum_r h_{ir} f'_{rj}} \left(\log(h_{ik} f_{kj}) - \log \frac{h_{ik} f'_{kj}}{\sum_r h_{ir} f'_{rj}} \right) \end{aligned} \tag{14}$$

From this inequality and the definition of $\mathcal{L}(\mathbf{F})$, it follows that $\Phi(\mathbf{F}, \mathbf{F}') \geq \mathcal{L}(\mathbf{F})$.

Based on Eq. 10, the update of $\mathbf{F}$ can be obtained by setting $\frac{\partial \Phi(\mathbf{F}, \mathbf{F}')}{\partial f_{st}} = 0$, that is,

$$\sum_i h_{is} - \sum_i g_{it} \frac{h_{is} f'_{st}}{\sum_r h_{ir} f'_{rt}} \frac{1}{f_{st}} + \lambda_1 \frac{\partial U_1(\mathbf{F})}{\partial f_{st}} + \lambda_2 \frac{\partial U_2(\mathbf{F})}{\partial f_{st}} = 0 \tag{15}$$

where

$$\begin{aligned} \frac{\partial U_1(\mathbf{F})}{f_{st}} &= \frac{1}{\tau^2} \sum_i c_{is} [\mathbf{CF}]_{it} \exp\left(-\frac{[\mathbf{CF}]_{it}^2}{2\tau^2} \right) \\ \frac{\partial U_2(\mathbf{F})}{f_{st}} &= \frac{1}{B^2} \sum_{i=1, i\neq s}^{B^2} f_{it} - \frac{1}{B^2} f_{st} \end{aligned} \tag{16}$$

Multiplying both sides of Eq. 15 by f_{st} and rearranging the terms, we arrive at

$$\begin{aligned} &-\frac{\lambda_2}{B^2} f_{st}^2 + \left(\sum_i h_{is} + \frac{\lambda_1}{\tau^2} \sum_i c_{is} [\mathbf{CF}]_{it} \exp\left(-\frac{[\mathbf{CF}]_{it}^2}{2\tau^2} \right) \right. \\ &\left. + \frac{\lambda_2}{B^2} \sum_{i=1, i\neq s}^{B^2} f_{it} \right) f_{st} - \sum_i g_{it} \frac{h_{is} f'_{st}}{\sum_r h_{ir} f'_{rt}} = 0 \end{aligned} \tag{17}$$

Solving the above quadratic function of f_{st}, we immediately achieve the update rule in (11).

Following the similar procedure, we can prove that

$$h_{st} = h'_{st} \frac{\sum_i f_{ti} g_{si} / [\mathbf{H}'\mathbf{F}]_{si}}{\sum_i f_{ti}} \tag{18}$$

In order to consider the row-sum-to-one constraint, the augmented matrices $\mathbf{G}^+$ and $\mathbf{F}^+$ are used. The columns are then normalized to have unit $\mathbb{L}^1$ norm. These operations together with the update in Eq. 18 result in the learning rule for $\mathbf{H}$ in (11). □

From the above analyses, we conclude that the learning steps in (11) result in a sequence of non-increasing values of $\mathcal{L}(\mathbf{H}, \mathbf{F})$, therefore, it converges to a local minimum. However, the convergence to the global minimum cannot be guaranteed. Another important issue is that we have not explicitly imposed the nonnegative constraint during the algorithm derivation. However, the nonnegativity of solution has been guaranteed by the derived learning rules. It is obvious that the learning of $\mathbf{H}$ follows the multiplicative update rule with a positive factor and a nonnegative initial $\mathbf{H}^0$, thus its elements will never become negative. $\mathbf{F}$ is calculated by solving a quadratic function. As can be seen, $a < 0$, $c < 0$, and $b > 0$. For small parameter λ_2, we have $b^2 - 4ac > 0$ and $\sqrt{b^2 - 4ac} < b$ since $ac > 0$. Therefore, $f_{st} > 0$ is guaranteed.

4. Algorithm Implementation and Results

In this section, we illustrate the performance of the proposed CNMF restoration method. We compare the results obtained by CNMF to the minimum norm blind deconvolution (MNBD) method [11], which deterministically calculates the image and kernel pair without iterative process. The algorithm involves two free parameters that need to be selected for a specific experiment, the regularization parameter γ and the smoothness parameter s; see [11] for more details.

4.1. Practical Implementation Issues

There are several implementation issues that need to be clarified. First, we partition the image domain into partially overlapped instead of discrete non-overlap blocks. The reason for doing this is because the independent deconvolution of individual block has a couple of shortcomings: first, the deconvolution operation induces boundary effects, which result in a blocking artifact; second, the algorithm is sensitive to the target position in the image with respect to the block position. The use of overlapping blocks is able to alleviate these drawbacks, but it unfortunately increases the computational burden due to large data size. To achieve fast computation and still be able to reduce the boundary artifacts, we choose blocks with the minimum amount of overlap needed, which is determined by the kernel size K.

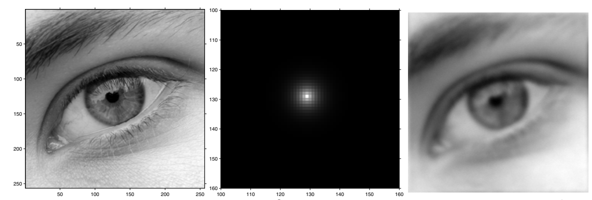


Figure 3. The original image f (left), the actual blur kernel h with $\kappa = 0.4$ (middle) and the degraded image with 2% noise (right).

The size of overlap between blocks is $K-1$. Then, only the center part of the restored block is kept. If we increase the size of overlapping, we would have more than one estimate for each pixel, and the mean of these values can be used as the final estimate.

Another issue is also related to the block-based formulation. It is apparent that we have made the assumption of a limited finite kernel support. In the experiment, two types of blurs will be investigated, i.e., the atmospheric turbulence blur and the uniform out-of-focus blur. The atmospheric turbulence blur is caused by long-term exposure through the atmosphere commonly occurred in remote sensing and aerial imaging. The analytical model is given by a Gaussian distribution $h(x, y) = \rho \exp\left(-\kappa\sqrt{x^2 + y^2}\right)$, where ρ is a normalizing constant ensuring that h has unit $\mathbb{L}^1$ norm, and x, y vary from -1 to 1. The parameter κ determines the severity of the blur. The uniform out-of-focus blur is a simple approximation of optical defocus, which has a uniform distribution within a circular disk

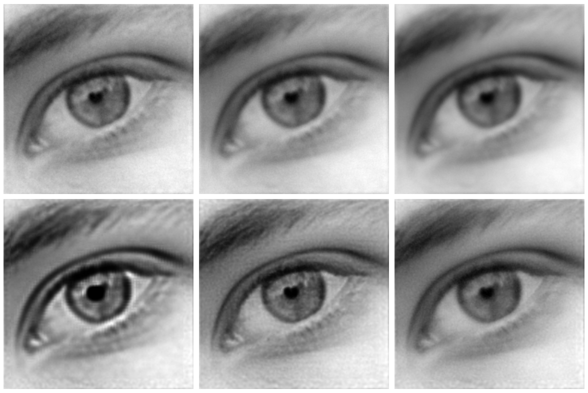
$$h(x, y) = \begin{cases} \frac{1}{\pi R^2}, & \text{if } \sqrt{x^2 + y^2} \le R; \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

For a Gaussian kernel estimator $\hat{h} = \rho \exp\left(-\hat{\kappa}\sqrt{x^2 + y^2}\right)$ with small $\hat{\kappa}$, its nonzero support is too large for block formulation. Therefore, we only keep 80% of the kernel energy and set small values to zero. The remaining elements of $\hat{h}$ is then normalized so that $\sum_{x,y} \hat{h}(x, y) = 1$. The kernel size K is referred to as the support of the truncated kernel. After the determination of the kernel size, the block size B can be selected to reduce the boundary artifacts.

4.2. Experimental Results

In the following experiments, we shall first evaluate CNMF performance for images degraded by the Gaussian blur. In the last experiment, different sizes of out-of-focus blur will be studied. To simulate possible noise corruptions, we add white Gaussian noise to the blurred image. Following [11], the noise level is measured according to $n_l = \|\mathbf{g} - \mathcal{H}\mathbf{f}\| / \|\mathcal{H}\mathbf{f}\|$. In the experiments, four levels of noisy data with $n_l = 1\%, 2\%, 5\%$ and 10% will be tested. The test image, a Gaussian blur kernel with $\kappa = 0.4$, and the degraded image with 2% noise are shown in Fig. 3.

The reconstructed images using MNBD and CNMF with different kernel estimators are demonstrated in Fig. 4. The



$\hat{\kappa} = 0.2$ $\hat{\kappa} = 0.4$ $\hat{\kappa} = 0.6$

Figure 4. Reconstructed images by MNBD (top row) and CNMF (bottom row) using different kernel estimators $\hat{\kappa}$. The degraded image and the actual blur kernel are shown in Fig. 3.

top-row images are results of MNBD with parameters set to $s = 1.3$ and $\gamma = 1^{-10}$ based on the guide in [11]. Note that as the kernel parameter $\hat{\kappa}$ increases, the reconstructed images are getting closer to the blurred version. For the CNMF reconstructions, all the images show a significant quality improvement compared to the degraded images. The best performance is obtained when the actual blur kernel is used. Note the detail information around the eye has been recovered.

We then investigate how the algorithm performs for different degrees of blur. For this purpose, we blur the original image using the kernels with κ equal to 0.2, 0.3, and 0.5. The corresponding kernel estimators are $\hat{\kappa} = 0.1$, 0.2, and 0.4. We again add 2% noise to the blurred images. Fig. 5 elaborates the experimental results, where the first row shows the degraded images. The second and third row correspond to the restorations of MNBD and CNMF, respectively. Both methods have the capability of reconstructing image details and exhibit comparable visual performance. The quantitative measurements in terms of improved signal-to-noise ratio (ISNR) are illustrated in Fig. 6. As can be seen, CNMF yields higher ISNR than MNBD for various blurs.

To demonstrate the effect of noise, we change the noise level to 1%, 5% and 10%. The actual blur kernel is given by $\kappa = 0.4$ and its estimator is chosen to be $\hat{\kappa} = 0.3$. The optimal scaling factor for MNBD is selected as $\gamma = 1.5, 1.2, 1$ and the smoothing parameter is $s = 1^{-10}$ for all three cases. The degraded images and the reconstructions for the three noise levels are demonstrated in Fig. 7. Both MNBD and CNMF show enhanced image quality, and CNMF performs better in recovering more details along the eyelids.

The last experiment is designed to investigate the algorithm performance for the out-of-focus blur. The kernel radius R is varied to be 4, 6 and 8, and the corresponding kernel estimator changes from 6, 8, to 10. The blurred images and the reconstructions are shown in Fig. 8, where the noise level is again set to $n_l = 2\%$. It can be seen that the MNBD reconstructions display obvious artifacts. The reason for this bad performance has been addressed in [11]. While, the CNMF method demonstrates improved image quality by extracting more details.

Figure 5. Degraded (top row) and reconstructed images by MNBD (middle row) and CNMF (bottom row) for different blur kernels, $\kappa = 0.2, \hat{\kappa} = 0.1$ (first column), $\kappa = 0.3, \hat{\kappa} = 0.2$ (second column), and $\kappa = 0.5, \hat{\kappa} = 0.4$ (third column).

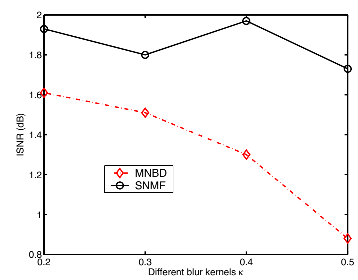


Figure 6. ISNR comparison of reconstructed images by MNBD and CNMF for different blur kernels.

5. Conclusion

In this paper, we have re-investigated the physical image formation process and formulated the classic image restoration as a BSS problem. The new interpretation regards the degraded image as a linear combination of a set of shifted version of PSF weighted by the actual image values. A constrained NMF approach with local smoothness and block-decorrelation regularization is developed to recover the source image. The comparative analysis with one of the state-of-the-art methods is conducted, which demonstrates the merit of the proposed approach.

Figure 7. Degraded (top row) and reconstructed images by MNBD (middle row) and CNMF (bottom row) for three different noise levels, $n_l = 1\%$ (first column), $n_l = 5\%$ (second column), and $n_l = 10\%$ (third column).

Figure 8. Degraded (top row) and reconstructed images by MNBD (middle row) and CNMF (bottom row) for different sizes of out-of-focus blur, $R = 4, \hat{R} = 6$ (first column), $R = 6, \hat{R} = 8$ (second column) and $R = 8, \hat{R} = 10$ (third column).

Acknowledgements

The authors wish to thank Dr. Justen at the University of Bremen for providing the MNBD code and the test image, and the reviewers for useful comments and suggestions.

References

[1] S. I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing System*, 8:757–763, 1996. 1, 3

[2] H. C. Andrews and B. R. Hunt. *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977. 1

[3] H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, May 1999. 3

[4] L. Bar, N. Sochen, and N. Kiryati. Variational pairing of image segmentation and blind restoration. In *European conference omg computer vision (ECCV'04)*, volume 3022, pages 166–177, May 2004. 1

[5] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995. 1, 3

[6] M. Burger and O. Scherzer. Regularization methods for blind deconvolution and blind source separation problems. *Math. Cont. Signals & Systems*, 14:358–383, 2001. 1

[7] T. F. Chan and C. K. Wong. Total variation blind deconvolution. *IEEE Trans. Image Process.*, 7(3):370–375, Mar. 1998. 1

[8] T. F. Chan and C. K. Wong. Convergence of the alternating minimization algorithm for blind deconvolution. *Linear Algebra and its Applications*, 316(1-3):259–285, 2000. 1

[9] D. C. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM J. Appl. Math.*, 56:1181–1198, 1996. 1

[10] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004. 2

[11] L. Justen and R. Ramlau. A non-iterative regularization approach to blind deconvolution. *Inverse Problem*, 22:771–800, 2006. 1, 5, 6

[12] I. Kopriva. Single-frame multichannel blind deconbolution by nonnegative matrix factorization with sparseness constraints. *Optical Letters*, 30(23):3135–3137, Dec. 2005. 2

[13] D. Lee and S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999. 3

[14] D. Lee and S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, volume 13, pages 556–562. MIT Press, 2001. 3, 4

[15] P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994. 3

[16] L. Rudin and S. Osher. Total variation based image restoration with free local constraints. In *Proc. IEEE international conference on image processing*, pages 31–35, Austin, TX, 1994. 1

[17] C. R. Vogel and M. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Trans. Image Process.*, 7:813–824, Jun. 1998. 1

[18] Y. You and M. Kaveh. A regularization approach to joint blur identification and image restoration. *IEEE Trans. on Image Process.*, 5:416–428, 1996. 1