CHAPTER 15

Digital Deconvolution of Fluorescence Images for Biologists

Yu-li Wang
Cell Biology Group
Worcester Foundation for Biomedical Research
Shrewsbury, Massachusetts 01545

I. Introduction

The performance of optical microscopes is limited both by the aperture of the lens, which causes light from a point source to spread (or diffract) over a finite volume, and by the cross-contamination of light that originates from out-of-focus planes. To overcome these limitations, approaches have been developed in recent years both to improve the microscope design, as exemplified by confocal scanning microscopy, and to reverse mathematically the degrading effects of the conventional microscope. This latter approach is commonly referred to as “deconvolution,” since the degradation effects of the microscope can be described mathematically as the convolution of input signals by the point spread function of the optical system (see below; Young, 1989; Russ, 1994).

A number of deconvolution algorithms have been tested for restoring fluorescence images. The most straightforward approach, 3D inverse filtering (Agard et al., 1989; Holmes and Liu, 1992), attempts to reverse the effects of image degradation through direct calculations. Unfortunately, it usually suffers from...
excessive computational artifacts. The two methods currently in wide use are constrained iterative deconvolution (Agard, 1984; Agard et al., 1989; Shaw, 1993; Holmes and Liu, 1992) and nearest-neighbor deconvolution (Castleman, 1979; Agard, 1984; Agard et al., 1989). The purpose of this chapter is to introduce the basic rationale of these two methods in languages easily understood by biologists. It will also provide details for the implementation of nearest-neighbor deconvolution using readily available hardware and software. Readers who wish to have a concise introduction of convolution and Fourier transformation for imaging are referred to the book by Russ (1994).

II. Rationale of Constrained Iterative Deconvolution

The two-dimensional convolution operation is defined mathematically as:

\[
i(x, y) \otimes s(x, y) = \sum_{u,v} i(u, v) s(x - u, y - v)
\]  

(1)

This equation can be easily understood when one looks at the effects of convolving a matrix \( i \) with a 3 × 3 matrix:

\[
\begin{array}{cccccc}
... & ... & ... & s_1 & s_2 & s_3 \\
... & ... & ... & s_4 & s_5 & s_6 \\
... & ... & ... & s_7 & s_8 & s_9 \\
\end{array}
\]

Following the calculation, the element \( i_5 \) is replaced by \((i_1 \times s_9) + (i_2 \times s_8) + (i_3 \times s_7) + (i_4 \times s_6) + (i_5 \times s_5) + (i_6 \times s_4) + (i_7 \times s_3) + (i_8 \times s_2) + (i_9 \times s_1)\). That is, each element is now “contaminated” by contributions from the surrounding elements to an extent specified by the values in the \( s \) matrix. In an optical system, the degree of “contamination” is measured as the point spread function (the output image of a point source).

The process of image formation in a microscope can be described as the original distribution of intensities convolved by the 3D point spread function of the optical system (Agard et al., 1989):

\[
o(x, y, z) = i(x, y, z) \otimes s(x, y, z)
\]  

(2)

where \( i(x, y, z) \) is a 3D matrix describing the signal originating from the sample and \( s(x, y, z) \) is a matrix describing the 3D point spread function.

Alternatively, it is equally valid to write the equation using a series of 2D point spread functions (Agard et al., 1989):

\[
o(x, y) = i_0(x, y) \otimes s_0(x, y) + i_{-1}(x, y) \otimes s_{-1}(x, y) + i_{-2}(x, y) \otimes s_{-2}(x, y) + \cdots
\]  

(3)

\[
+ i_{-2}(x, y) \otimes s_{-2}(x, y) + i_{-3}(x, y) \otimes s_{-3}(x, y) + \cdots
\]

where \( i(x, y) \)s are 2D matrices describing the signal originating from the plane of focus \( i_0 \) and from planes above \((i_{-1}, i_{-2}, \ldots)\) and below \((i_{+1}, i_{+2}, \ldots)\), \( s(x, \ldots)\).
15. Digital Deconvolution of Fluorescence Images

307

...s are matrices of 2D point spread functions that describe how point sources on the plane of focus \((s_0)\) or planes above \((s_1, s_2, \ldots)\) and below \((s_{-1}, s_{-2}, \ldots)\) spread out when they reach the image plane.

Constrained iterative deconvolution uses a trial and error process to look for signal distribution \(i(x,y,z)\) that satisfies equation (2). It usually starts with the assumption that \(i(x,y,z)\) equals the measured stack of optical sections \(o(x,y,z)\). As expected, when \(o(x,y,z)\) is plugged into the right hand side of equation (2) in place of \(i(x,y,z)\), it generates a matrix \(o'(x,y,z)\) that deviates from \(o(x,y,z)\) on the left-hand side. To decrease this deviation, adjustment is made to the initial matrix \(o(x,y,z)\), voxel by voxel, based on the deviation of \(o'(x,y,z)\) from \(o(x,y,z)\) and on constraints such as nonnegativity of voxel values. Various approaches have been developed to determine how adjustments should be made to the trial image and how voxel values should be “constrained” (Agard et al., 1989; Holmes and Liu, 1992). The modified \(o(x,y,z)\) is then plugged back into the right-hand side of equation (2) to generate a new matrix, \(o''(x,y,z)\), which resembles more closely \(o(x,y,z)\). This process is repeated at least 20–30 times until there is no further improvement or until the calculated image matches closely the actual image.

III. Rationale of Nearest-Neighbor Deconvolution

The nearest-neighbor algorithm uses equation (3) as the starting point. The equation is simplified by introducing three assumptions:

1. Out-of-focus light from planes other than those adjacent to the plane of focus is negligible (i.e., terms containing \(s_{-2}, s_{+2}\), and beyond are insignificant).

2. Light originating from planes immediately above or below the plane of focus can be approximated by images taken while focusing on these planes (i.e., \(i_{-1} \approx o_{-1}\) and \(i_{+1} \approx o_{+1}\)).

3. Point spread functions for planes immediately above and below the focal plane, \(s_{-1}\) and \(s_{+1}\), are equivalent (hereafter denoted as \(s_1\)).

Together, these approximations simplify equation (3) into:

\[
o = i_0 \otimes s_0 + (o_{-1} + o_{+1}) \otimes s_1 \tag{4}
\]

Rearranging the terms and taking advantage of the mathematical fact that if \(a \otimes b = c\), then \(F(a) \times F(b) = F(c)\), where \(F\) represents Fourier transformation and \(\times\) represents multiplication of corresponding elements in the matrices, it can be shown that

\[
i_0 = [o - (o_{-1} + o_{+1}) \otimes s_1] \otimes F^{-1}(1/F(s_0)) \tag{5}
\]

where \(F^{-1}\) represents reverse Fourier transformation. This equation can be understood in a simple, intuitive way: it states that the unknown signal distribution,
\(i_0\), can be obtained by taking the in-focus image, \(\sigma\), subtracting out estimated contributions from planes above and below the plane of focus, \((\sigma_{-1} + \sigma_{+1}) \otimes s_1\), followed by convolution with the matrix \(F^{-1}(1/F(s_0))\), which reverses diffraction-induced spreading of signals on the plane of focus.

Among the three approximations, the second is the most serious. On the one hand, since images taken from planes immediately above or below the plane of focus can include significant contributions of signals from the plane of focus, the use of these images leads to oversubtraction and erosion of structures. On the other hand, due to the diffraction of light, these out-of-focus images also somewhat underrepresent the true contribution from the corresponding planes.

In practice, nearest neighbor deconvolution is performed with a modified form of equation (5):

\[
i_0 = [\sigma - (\sigma_{-1} + \sigma_{+1}) \otimes (c_1 \cdot s_1)] \otimes F^{-1}(F(s_0)/(F(s_0)^2 + c_2)),
\]

where constants \(c_1\) and \(c_2\) are empirical factors. \(c_1\) is used to offset errors caused by oversubtraction as described above. \(c_2\) is required to deal with problems associated with the calculation of reciprocals at the end of equation (5): the error could become astronomical when the value of the matrix element is small. The use of constant \(c_2\) keeps the reciprocal value from getting too large when the matrix element is small compared to \(c_2\). However it does not significantly affect the outcome when the matrix element is large compared to \(c_2\).

### IV. Implementation of Nearest-Neighbor Deconvolution

Nearest-neighbor deconvolution can be performed with readily available equipment: a conventional fluorescence microscope, a stable light source, a stepping motor coupled to the microscope focusing mechanism, a cooled slow-scan CCD camera, and a personal computer. According to equation (6), the calculation of \(i_0\) requires the collection of only in-focus image \(\sigma\) and images immediately above and below the focal plane (Fig. 1). These images are convolved with two matrices, \(c_1 \cdot s_1\) and \(F^{-1}(F(s_0)/(F(s_0)^2 + c_2))\), which are determined by the point spread functions of the microscope system (alternatively, similar calculations can be performed in the frequency space, by replacing matrices in equation (6) with corresponding Fourier transformations and convolution operations with element-by-element multiplication of the matrices).

The easiest way to obtain the two matrices, \(s_0\) and \(s_1\), is to take serial optical sections, at a spacing equal to that used for collecting images of the sample, of fluorescent beads of \(-0.1\ \mu m\) diameter. The image with the highest intensity at the center of the bead is identified as the in-focus image. To obtain matrix \(s_0\), this image is trimmed to an appropriate size, made radially symmetric by averaging, and normalized such that the sum of all elements equals 1 (Fig. 1). We found that the optimal size of the matrix lies between \(11 \times 11\) and \(17 \times 17\). This matrix is used as the input for Fourier transformation and matrix multiplication/
division to generate $F^{-1}(F(s_0)/(F(s_0)^2 + c_2))$. The calculations can be streamlined using readily available PC programs such as Mathcad (Fig. 1). The generation of $s_1$ is more straightforward. Images of beads immediately above and below the plane of sharp focus are averaged, followed by trimming, symmetrization, and normalization as for the generation of $s_o$. The constant $c_1$ is then incorporated into the $s_1$ matrix as shown in the last equation of Fig. 1.

The performance of nearest-neighbor deconvolution is highly sensitive to the values $c_1$ and $c_2$ in equation (6). In general, $c_1$ falls in the range of 0.45 to 0.50: the optimal value varies with the optical condition and the separation between adjacent optical slices. Too large a value causes erosion and discontinuity of in-focus structures (Fig. 3c), while too small a value would lead to high residual background due to the incomplete removal of out-of-focus noises (Fig. 3a). The value of $c_2$ is generally in the range of 0.0001 to 0.001. Too large a $c_2$ value causes the loss of details, yielding blurry images (Fig. 3e), while too small a value would cause the amplification of random noises into bright spots, rings, or patches (Fig. 3d). The optimal values for $c_1$ and $c_2$ can be found only through systematic trials.

The convolution matrices and sample images are then fed into equation (6) to generate the desired image $i_o$. This requires a computer program that can perform image/matrix subtraction and convolution with floating-point matrices (or fast Fourier transformation for calculations in the frequency space). These functions again can be programmed into common mathematical packages such as Mathcad 6.0 Plus. To improve the speed of calculation, these operations are performed in this laboratory with a DSP board installed in a personal computer (AL860-40MHz, single processor; Alacron, Inc., Nashua, NH), which with dedicated software can perform convolution of a $512 \times 384$ image with a $17 \times 17$ floating-point matrix within 2 sec. The entire computations of equation (6) can be completed within 10 sec. After computation, the image needs to be scaled properly to yield gray values suitable for display on a computer monitor. A stack of processed optical sections can then be used for 3D reconstruction or visualization (Fig. 4).

V. Evaluation of Digital Deconvolution Methods

For both constrained iterative deconvolution and nearest-neighbor deconvolution, images are collected with a conventional microscope, which has advantages such as limited photobleaching and versatility in the choice of excitation wavelengths. While it is essential to use a high-quality cooled CCD camera, the rest of the system can be installed with limited cost using readily available equipment and software. In addition, it is possible to implement both nearest-neighbor and constraint iterative programs in the same system, using nearest-neighbor for preliminary evaluation of images and constrained iterative deconvolution for more precise restorations.
File “onfile” contains the in-focus intensity distribution of fluorescent beads in a 17x17 matrix, obtained by trimming and averaging a number of in-focus images.

ON = READPRN(onfile)

```
26 24 27 28 27 29 28 29 29 28 26 24 25 21 22 21 20 20
27 28 28 31 32 33 35 34 32 29 27 25 24 25 21 22 22
26 27 29 34 36 39 41 42 41 36 31 30 26 25 24 23 23
28 30 33 35 43 48 53 56 53 47 41 32 28 25 23 22 22
29 31 35 42 54 59 69 82 86 79 60 39 33 26 24 22 22
32 33 37 49 60 71 104 157 190 163 107 61 36 29 27 26 23
33 36 41 55 63 90 172 296 345 294 176 90 45 34 28 22 23
32 37 44 60 68 107 225 383 451 370 220 102 56 33 28 26 24
34 36 44 59 70 98 202 332 395 322 193 93 50 35 29 26 23
33 36 43 55 68 82 129 204 236 188 122 66 44 35 30 27 23
34 36 40 50 63 69 81 99 104 92 70 51 44 33 29 25 23
35 35 40 45 52 58 62 59 57 52 47 43 37 33 28 25 25
32 35 36 38 46 47 48 49 45 43 39 37 33 30 28 25 23
29 31 34 37 37 38 40 40 35 35 35 32 27 27 25 24 22
27 29 32 32 33 32 34 31 31 29 30 30 26 26 23 24 22
27 26 28 29 27 30 31 30 30 28 26 26 24 25 23 23 22
```

File “offile” contains the out-of-focus intensity distribution of fluorescent beads in a 17x17 matrix, obtained by trimming and averaging a number of images immediately above and below the plane of focus.

OFF = READPRN(offile)

```
24 27 27 28 28 30 31 30 27 25 25 25 26 22 22 20 20 20
26 27 28 30 30 35 34 32 31 31 26 25 22 22 21 21 20 20
26 28 31 34 37 36 38 36 35 30 32 27 25 24 24 23 24 24
28 28 34 39 41 41 43 45 42 40 35 33 28 25 21 25 22
29 30 34 36 44 50 56 59 59 56 45 35 30 27 26 23 22
30 32 38 41 51 59 74 93 100 90 70 49 37 28 27 25 24 24
34 34 40 46 58 72 113 168 195 172 120 70 45 32 28 26 25 25
33 36 44 51 59 88 166 276 312 269 177 98 54 36 31 25 24
32 38 44 54 62 97 195 306 368 314 200 110 59 38 32 28 27 27
35 37 43 55 61 88 162 261 304 256 172 100 57 38 32 29 26
35 34 43 51 60 69 103 152 178 160 115 77 50 37 32 29 28
36 37 39 49 55 61 67 80 87 83 69 54 43 37 30 30 27 27
34 37 38 41 47 53 51 52 53 53 46 45 39 35 29 28 25 25
32 34 38 38 41 41 43 43 40 39 37 34 31 28 27 25 25 25
32 31 32 33 35 37 38 35 35 36 34 31 29 27 24 24 24 24
28 28 30 32 31 32 32 33 34 31 30 30 28 28 25 24 25 25
27 26 29 29 29 30 31 30 29 28 28 26 25 24 23 24 24
```

Fig. 1 Mathead program used for the calculation of $s_0$ and $s_1$. The averaged in-focus image of fluorescent beads was stored in an ASCII file “onfile.” Out-of-focus images collected above and below the plane of sharp focus were averaged and stored in an ASCII file “offile.” These images are shown as two $17 \times 17$ matrices near the beginning of the figure. The calculations then symmetrize and normalize the matrices as indicated in the comments. The resulting matrices, denoted as “ON” and “OFF,” are then used for the calculation of $s_0$ and $s_1$, which are output as ASCII files at the end of the program. Note that parameter $c_1 (0.50)$ is incorporated into the $s_1$ matrix near the end of the figure.
Symmetrize the matrices

\[ \begin{array}{cccc}
ON_{i,j} & ON_{i+16-j,j} & ON_{i,j+16} & ON_{i+16,j+16} \\
ON_{i+16-j,j} & ON_{i,j+16} & ON_{i+16,j+16} & ON_{i,j} \\
ON_{i,j+16} & ON_{i+16,j+16} & ON_{i,j} & ON_{i+16-j,j} \\
ON_{i+16,j+16} & ON_{i,j} & ON_{i+16-j,j} & ON_{i,j+16} \\
\end{array} \]

Use the corner value as background intensity, subtract from all elements

\[ \begin{align*}
i = 0 \ldots 16 \\
j = 0 \ldots 16 \\
\text{Back} &= \min(ON_{i,j}, OFF_{i,j}) \\
\text{ON}_{i,j} &= \text{ON}_{i,j} - \text{Back} \\
\text{OFF}_{i,j} &= \text{OFF}_{i,j} - \text{Back} \\
\end{align*} \]

Normalize the matrices

\[ \begin{align*}
\sum_{i,j} ON_{i,j} &= \sum_{i,j} OFF_{i,j} \\
ON_{i,j} &= \frac{ON_{i,j}}{\sum_{i,j} \text{ON}_{i,j}} \\
OFF_{i,j} &= \frac{OFF_{i,j}}{\sum_{i,j} \text{OFF}_{i,j}} \\
\end{align*} \]

Calculate the Fast Fourier Transform, assuming \( c_2 = 0.0005 \)

\[ \begin{align*}
\text{FTON}_{i,j} &= \text{FFT}(\text{ON}_{i,j}) \\
\text{FTOFF}_{i,j} &= \text{FFT}(\text{OFF}_{i,j}) \\
\end{align*} \]

Perform Inverse FFT

\[ \text{SO} = \text{Real}(\text{FFT}(\text{FTSO})) \]

\[ \text{NormSO} = \text{SO} \text{ such that the sum of elements equals 10. This makes the output decimal numbers more readable but does not affect the final results. Parameter c1 is multiplied into the s1 matrix.} \]

\[ \begin{align*}
\sum_{i,j} \text{SO}_{i,j} &= 10 \\
\text{PRNCOLWIDTH} &= 12 \\
\text{WRITEPRN}(\text{SO}) &= \frac{\text{SO}}{\sum_{i,j} \text{SO}_{i,j}} \\
\end{align*} \]

Fig. 1—Continued
Fig. 2 Original images used for nearest-neighbor deconvolution. NRK epithelial cells were stained with antibodies against β-tubulin and rhodamine-conjugated secondary antibodies. Images were recorded with a Zeiss Axiocert microscope, a 100X/N.A. 1.30 neofluar lens, and a cooled slow-scan CCD camera from Princeton Instruments (Trenton, NJ). The image is 576 x 384 pixels, with each pixel corresponding to a sample area of 0.085 x 0.085 μm. The three images are 0.25 μm apart in focus from one another. During the calculation, b is used as the in-focus image and a and c are used as its nearest neighbor.

Compared to confocal laser scanning microscopy, one disadvantage of both methods is their limited ability to penetrate thick samples. In addition, when the sample is very weak, input images become limited in S/N ratio, which could seriously jeopardize the calculation for image restoration. Both deconvolution methods are also sensitive to instabilities in image intensity, caused by fluctuating lamp intensities or sample photobleaching. Thus samples should be mounted in antibleaching solutions whenever possible and microscope lamps should be driven with stabilized DC power supplies. Another common characteristic of the deconvolution techniques is their sensitivity to the vertical distance among optical sections. The distance should be close enough such that the stack includes focused or close-to-focused images of all structures. However, too close a distance would cause increases in the calculation time for the constrained iterative method, and for the nearest-neighbor method, serious errors due to assumption 2 discussed above. Typically, the optimal distance falls in the range of 0.25–0.50 μm.

Under ideal conditions, constrained iterative deconvolution can achieve a resolution that goes beyond what is predicted by the optical theory and provide 3D intensity distribution of photometric accuracy (Carrington et al., 1995). In addition, it can generate accurate image slices between collected slices. With confocal microscopes or nearest-neighbor deconvolution, such information can be generated only through mathematical interpolation. The main drawback of constrained iterative deconvolution is its computational demand. While it is possible to perform the calculation with a PC, the restoration of one stack of images could take many hours. Thus for practical purposes, constrained iterative deconvolution has to be performed at least with a powerful workstation. A second disadvantage with constrained iterative deconvolution is its requirement for a complete stack of images for deconvolution, even if the experiment requires only one optical slice. Iterative constrained deconvolution is also very sensitive to the condition of the microscope (dust, alignment, etc.) and the quality of the
Fig. 3 Nearest-neighbor deconvolution of images shown in Fig. 2 with different values of $c_1$ and $c_2$. (b) The optimal setting, with $c_1 = 0.50$ and $c_2 = 0.0005$. When $c_1$ is decreased to 0.46 (a), the removal of out-of-focus noises becomes incomplete, resulting in high diffuse background. When $c_1$ is increased to 0.54 (c), out-of-focus noises are overcorrected, resulting in fragmented structures. The parameter $c_2$ controls the sharpness of the final image. When it is too small (d; $c_2 = 0.00004$), the image suffers from excessive spotty noise. In addition, some structures become double (arrow). When $c_2$ is too large (e; $c_2 = 0.01$), the structures become fuzzy and faint.
Fig. 4 A stereo pair constructed from a complete stack of 25 deconvolved images. The images were deconvolved as in Fig. 3b, with values of 0.50 and 0.0005 for $c_1$ and $c_2$, respectively. Although the stereo pair is generated with custom-written software, similar results can be achieved with commercial packages such as Metamorph (Universal Imaging, West Chester, PA).

With the improvements in computer performance, digital deconvolution is becoming increasingly feasible for average cell biology laboratories. Even with
the advancement in confocal scanning microscopy such as two-photon excitation, computation methods will continue to serve their unique purposes. For example, nearest-neighbor deconvolution will remain as an efficient, economical method for samples of discrete structures and limited thickness (5–20 μm). The iterative constraint deconvolution, on the other hand, will remain as the method of choice for obtaining precise 3D fluorescence distribution. Most importantly, since confocal scanning microscopy and computational deconvolution work under independent principles, these methods can be easily combined to obtain resolution and photometric precision far beyond what was feasible with individual approaches.

References