

# Stochastic Deconvolution

James Gregson Felix Heide Matthias Hullin Mushfiqur Rouf Wolfgang Heidrich  
The University of British Columbia

## Abstract

*We present a novel stochastic framework for non-blind deconvolution based on point samples obtained from random walks. Unlike previous methods that must be tailored to specific regularization strategies, the new Stochastic Deconvolution method allows arbitrary priors, including non-convex and data-dependent regularizers, to be introduced and tested with little effort. Stochastic Deconvolution is straightforward to implement, produces state-of-the-art results and directly leads to a natural boundary condition for image boundaries and saturated pixels.*

## 1. Introduction

Image deconvolution or deblurring has applications in astronomy, microscopy, GIS and photography among other disciplines. As such it has seen considerable research in graphics and vision.

This paper presents *Stochastic Deconvolution*, a new framework for non-blind image deconvolution based on stochastic random walks. Stochastic Deconvolution is based on an adaptation of a recent stochastic optimization method for solving computed tomography problems [6] to the problem of deconvolution. The resulting algorithm amounts to a variant of coordinate-descent optimization, where the descent direction is chosen using a random walk that utilizes spatial coherence. By solving the image deblurring problem in this fashion, the Stochastic Deconvolution framework directly addresses several issues inherent in developing deconvolution algorithms:

- **Ease of Implementation.** Both the basic algorithm and its regularized variants are very straightforward to implement, and is based on only two very simple operations: splatting of the point spread function (PSF) and point-evaluation of the regularization term.
- **Regularization Research.** Because of the simplicity of implementing new regularizers, Stochastic Deconvolution enables research into new regularization terms and image priors for deconvolution through rapid experimentation. We demonstrate that the methods works for a large array of regularizers, including ones that

are smooth, non-smooth but convex, non-convex, discontinuous, and even data-dependent.

- **Boundary Conditions.** When capturing blurred images, information is propagated into the captured image where no data is captured. Deblurring in these regions requires some condition on scene content outside the captured region. An additional benefit of Stochastic Deconvolution is that it naturally handles these boundary conditions and can use a near-identical process to deal with saturated regions.
- **Shift-variant Kernels.** Finally, Stochastic Deconvolution generalizes naturally to deblurring problems with spatially varying kernels such as the synthetic camera shake example depicted in Figure 1.

The remainder of this paper is structured as follows: in the next section we discuss related work while providing an introduction to the deconvolution problem. We then introduce Stochastic Deconvolution in Section 3. Results are presented in Section 4 after which we conclude with a discussion of future research directions in Section 5.

## 2. Background and Related Work

In this section, we introduce the notation for the deconvolution problem and summarize the optimization framework from Stochastic Tomography [6], which we modify to solve deconvolution problems.

### 2.1. Image Deconvolution

Image deconvolution attempts to remove the blurs introduced when images are captured with real optical systems, including motion blur (e.g. [4, 15, 21, 21, 8, 7]) and depth-of-field blur (e.g. [12, 9, 3]). These artifacts are effectively captured by a point-spread-function (PSF)  $k$  that measures the projection of a point-light source on the captured image for a fixed set of camera parameters.

In general, the PSF is a function of the projected coordinate of the source  $x$ , the distance of the source from the camera  $d$ , and the chromaticity of the image point (i.e.  $k = k(x, d, \lambda)$ ). However in many scenarios the PSF is assumed to be spatially invariant, (i.e. independent of image position). The captured image  $q$  is then represented as the *intrinsic* (deblurred) image  $p$  convolved with the PSF:



Figure 1. Left: Snapshot of the algorithm in progress showing stochastic random walks that form the basis of Stochastic Deconvolution. Green points represent energy added while blue correspond to energy subtracted from the reconstruction. The algorithm automatically focuses sampling effort in regions where the largest improvements to the system energy are obtained. Right: Example of deblurring with a spatially-varying (per-pixel) PSF simulating strong motion-blur. Input image (center-left) and a sampling of per-pixel PSFs at full-scale (center-right). Deblurred result using Stochastic Deconvolution algorithm (right). The Stochastic Deconvolution algorithm naturally handles PSFs with strong spatial variation, including rotations around the optical axis, without resorting to patch-based approximations.

$$q = k \otimes p. \quad (1)$$

The goal of deconvolution is to invert Equation 1 to obtain an estimate of the intrinsic image. In this paper, we are focusing on the *non-blind* version of this problem, where the PSF is assumed to be provided either by calibration or some form of PSF estimation (e.g. [4, 10]).

Traditional methods for solving deconvolution problems include Fourier-space division, the Wiener Filter [18], as well as iterative methods such as Richardson-Lucy [16, 14]. All these methods produce significant artifacts in cases where certain image frequencies are completely eliminated by the blur, which is common especially in defocus blur.

Although the results can be improved significantly with variations of the original Richardson-Lucy algorithm (e.g. [21, 22, 5]), most state-of-the art deconvolution methods take a slightly different approach. The basic problem from Equation 1 can be seen as an linear inverse problem that is usually ill-posed, since the PSF filters out some frequency components. General deconvolution methods define a quadratic fitting energy (either in the Fourier or image domain) that is minimized when the solution estimate convolved by the PSF equals the captured image, e.g. when defined in the image domain:

$$F_{fit} = ||q - k \otimes p||_2^2 \quad (2)$$

Since the system is ill-posed, infinitely many solutions weakly minimize the fit energy (Equation 2). To address this, a prior or regularizer  $\Gamma(p)$  is typically added, weighted by  $\lambda$ , to give the system energy, Equation 3.

$$F = F_{fit} + \lambda \Gamma(p) \quad (3)$$

The regularizer penalizes solutions that do not conform to prior expectations on the solution such as smoothness or sparsity. Good regularizers suppress ringing and noise without introducing other undesirable artifacts.

However a problem arises because the regularizer typically changes the mathematical structure of the problem. In particular, priors favoring piecewise smooth solutions cannot be expressed as linear systems, making it necessary to develop highly specialized, regularizer-specific solvers (e.g. [12, 11, 19]). Developing such solvers is a demanding task, complicated further by problems with millions of unknowns.

The goal of our work is to design a simple, reasonably efficient, *general-purpose* deconvolution algorithm capable of handling effectively arbitrary priors. To do so, we adapt the random walk optimization strategy from Stochastic Tomography [6] and modify it to solve deconvolution problems. The result is a straightforward method for image deconvolution that allows the use of arbitrary priors with no change to the underlying algorithm. Another benefit of our method is natural handling of boundary conditions and saturated pixels.

## 2.2. Review of Stochastic Tomography

Recently, Gregson et al. [6] presented a stochastic random walk algorithm for solving tomographic reconstruction problems. The method minimizes a convex objective function  $F$  by continuously placing discrete point samples in a volume that each improve the objective. The change to the objective can be evaluated efficiently due to the small support of each sample. A local sample mutation strategy inspired by Metropolis-Hastings then focuses the sampling efforts in regions with high payoff, i.e. regions where samples have recently been placed successfully, leading to a method that makes many ( $10^7$ - $10^9$ ) low-cost incremental solution updates. However, as their work pointed out, the method deviates from Metropolis-Hastings in a number of key ways, including the fact that the random walk depends on the full history of the sampling process and does thus not represent a Markov Chain, but rather a stochas-

tic coordinate-descent method that employs a Metropolis-Hastings style heuristic for picking the next coordinate axis to descend along. The final result of the tomographic reconstruction is given by the volume density of the placed samples. Algorithm 1 reproduces the full method for completeness sake.

---

**Algorithm 1** Stochastic Optimization Algorithm, from [6]

---

```

// Start the walk from a random point
 $\mathbf{x}_0 \leftarrow \text{random}()$ 
for  $k = 1$  to  $N$  do
  // New sample from  $\mathbf{x}_{k-1}$  using
  // transition PDF  $t(\mathbf{x}_k|\mathbf{x}_{k-1})$ 
   $\mathbf{x}_k \leftarrow \text{sample}(\mathbf{x}_{k-1}, t(\mathbf{x}_k|\mathbf{x}_{k-1}))$ 
   $a \leftarrow \Delta F(\mathbf{x}_k)/\Delta F(\mathbf{x}_{k-1})$ 
  if  $\Delta F(\mathbf{x}_{k-1}) < 0$  or  $\text{random}() \leq a$  then
    // Record only samples that reduce the objective fn.
    if  $\Delta F(\mathbf{x}_k) > 0$  then
      // Incorporate the sample into the output
       $\text{record}(\mathbf{x}_k)$ 
    end if
  else
    // Keep exploring space from previous sample
     $\mathbf{x}_k = \mathbf{x}_{k-1}$ 
  end if
end for

```

---

A key advantage of Stochastic Tomography over other tomographic solvers is that the objective function  $F$  may contain arbitrary convex regularizers without a change in the fundamental algorithm, allowing for easy experimentation and testing of new priors and regularizers. Using  $L_1$  regularizers on several captured and synthetic examples, Gregson et al. demonstrated that Stochastic Tomography can be an effective method for regularized tomographic reconstruction.

One of the contributions of our work is to recognize that this framework for stochastic optimization with a random walk is in fact more general, and can be adapted to inverse problems other than tomography. This is significant since frequency content in measured quantities can differ significantly between deblurring and tomography, leading to more aggressive, often non-convex priors that are more difficult to optimize. The key features required of problems is to have i) a strong geometric structure in which many degrees of freedom can be explored by walks in a low dimensional space, and ii) to have small stencils, so local updates can be performed efficiently.

Deconvolution fits this definition nicely since the PSF links the intrinsic and captured images geometrically in 2D and has relatively compact support allowing efficient local updates. To apply this random walk framework, we only need to derive problem-specific functions for sample mutation, i.e. a transition probability  $t(\mathbf{x}_k|\mathbf{x}_{k-1})$  for choos-

ing sample  $\mathbf{x}_k$  based on the previous sample location  $\mathbf{x}_{k-1}$ , a method for keeping track of the change  $\Delta F(\mathbf{x}_k)$  of the objective function when placing a new sample  $\mathbf{x}_k$ , and finally a method for accepting and recording a new sample  $\text{record}(\mathbf{x}_k)$ . The next section describes how to derive methods for these tasks in the case of deconvolution problems.

### 3. Stochastic Deconvolution

Stochastic Deconvolution begins from Equation 2, which is used as the data-fitting term of the system objective function (Equation 3), and from an initial estimate of the  $p^{(0)}$  of the intrinsic image. We create a random walk of pixel locations  $\mathbf{x}_k$  at which we add or remove an energy quantum  $e_d$ , thus generating a sequence  $p^{(k)}$  of estimates of the intrinsic image:

$$p^{(0)} = q \quad (4)$$

$$p^{(k)} = p^{(k-1)} \pm e_d \cdot \delta_{\mathbf{x}_k}, \quad (5)$$

where  $\delta_{\mathbf{x}_k}$  is the characteristic function (Kronecker Delta) for pixel  $\mathbf{x}_k$ . Both positive and negative energies are tested for each sample location  $\mathbf{x}_k$  in Algorithm 1 but only the sign causing the greatest improvement kept.

**Evaluating  $\Delta F$ .** The quantity  $\Delta F(\mathbf{x}_k)$  measures the change in the objective function if a given sample  $\mathbf{x}_k$  with value  $\pm e_d$  were to be accepted and added to the solution. In order to efficiently compute  $\Delta F(\mathbf{x}_k)$ , we also keep track of a second sequence of images  $q^{(k)} = k \otimes p^{(k)}$ , which represents the observed image we would expect if  $p^{(k)}$  was the intrinsic image.  $q^{(k)}$  can be efficiently updated during the random walk:

$$q^{(0)} = k \otimes p^{(0)} = k \otimes q \quad (6)$$

$$q^{(k)} = q^{(k-1)} \pm e_d (k \otimes \delta_{\mathbf{x}_k}) \quad (7)$$

In other words,  $q^{(k)}$  can be updated by splatting  $k \otimes \delta_{\mathbf{x}_k}$ , a shifted and mirrored copy of the PSF at the sample location  $\mathbf{x}_k$ . With this second image, the change in the data term  $F_{fit}$  can be computed efficiently through local updates.

The change in the regularization energy is evaluated in an analogous manner, but is specific to the chosen regularizer. For example, if the total-variation (TV) function is used, then the change in  $\lambda\Gamma$  is simply the sum of differences of affected gradient magnitudes, scaled by  $\lambda$ .

**Mutation Strategy.** The mutation function generates a new sample  $\mathbf{x}_k$  from the previously accepted sample  $\mathbf{x}_{k-1}$ . We use a simple, symmetric strategy where new samples are generated by a Gauss-distributed offset from the most recently accepted sample. The width of the Gaussian is a user parameter typically set to 50-100% of the PSF width, with the choice not being critical. Using a Gaussian distribution ensures ergodicity; this ensures the sampling process does not erroneously ‘miss’ areas.

We also add a Russian-roulette chain terminating mutation where the sample is simply moved anywhere in the image domain with uniform probability. This mutation is applied with 1% probability, leading to sample chains with expected length of 100. We have found this helps to overcome any start-up bias while also contributing to ergodicity.

**Convergence.** In each iteration, the Stochastic Deconvolution framework picks a single pixel in the image and checks if the objective can be improved by depositing energy in this pixel. This corresponds to picking a single degree-of-freedom and descending along that axis, making it a form of Coordinate Descent. What distinguishes Stochastic Deconvolution from other Coordinate Descent methods [13, 17] is that we use the random walk process to exploit spatial coherence in the deconvolution problem, and focus the computational effort on regions with sharp edges, where most work is to be done in deconvolution.

Coordinate Descent methods provably converge for *smooth* objective functions for a fixed step length so long as all possible descent directions (i.e. pixels) are examined with a finite probability. In our framework, this condition is met by the ergodicity of the sampling process in the limit of number of samples.

For *general, non-smooth* objectives, no proof of convergence is available for Coordinate Descent, although convergence has been shown for *specific, separable*  $L_1$ -regularized problems such as basis pursuit [17, 13]. In this paper, we show empirical evidence of the convergence of Stochastic Deconvolution for convex objectives, in particular a total variation (TV) regularized deconvolution problem (Section 4).

As with other optimization strategies, no theoretical results are available for the use of non-smooth, non-convex objectives with Coordinate Descent. Our results in Section 4 empirically show that Stochastic Deconvolution is competitive for such regularizers and even for a simple discontinuous and data-dependent prior.

**Boundary Conditions and Saturated Pixels.** The issue of boundary handling is difficult in deconvolution algorithms, since the process of capturing an image necessarily cuts off some of the data needed to deconvolve at the image boundaries. Stochastic Deconvolution naturally handles this situation by padding the input image by the PSF width and creating a mask that indicates which pixels are from the captured region versus from the boundary region.

During the sampling process, samples are allowed to be placed anywhere within the image or padded regions, but when evaluating the change in system energy due to a sample, only samples flagged as interior have their data-fitting term  $F_{fit}$  evaluated, since the saturated and padding pixels have no valid captured value.

The same strategy can be used for other pixels where the measurements in the observed image are invalid, for example excessively bright pixels, where the image sensor has been saturated. Ignoring the data term for these regions while enforcing the regularization term causes the method to perform a simple form of inpainting in the padded and saturated regions to improve the fit to the valid measurements.

**Choosing  $e_d$ .** To choose the deposition energy  $e_d$ , an initially large value is assigned, e.g.  $e_d = 0.05$  (assuming pixel values in  $[0, 1]$ ). An outer iteration of the sampling procedure from Listing 1 is then started with a total of one mutation per-pixel, and the percentage of accepted samples computed. If this value falls below 7.5%,  $e_d$  is scaled by 0.75 before starting the next outer iteration. Outer iterations are continued in this manner until a set number of iterations is exhausted or convergence stalls. This simple adaptive choice for  $e_d$  works well in practice and frees the user from specifying a specific value.

**Comparison with Stochastic Tomography.** While Stochastic Deconvolution uses the same basic random walk as Stochastic Tomography [6], there are also a number of differences that are worth pointing out. First, adapting the method to deblurring requires very specific modifications to handle boundaries and saturation, while switching from continuously placed samples to discrete pixel locations.

Perhaps more significantly, deblurring can be thought of as redistributing the energy from the blurred image to form the sharp intrinsic image. This makes the need for negative energy samples obvious since both negative and positive samples are needed near edges. For Stochastic Tomography, such samples were only needed to prevent the algorithm from stalling due to startup-bias.

### 3.1. Regularization

We have implemented a host of different regularization strategies in the Stochastic Deconvolution framework but summarize here several that highlight the flexibility of the method.

**Total Variation.** Total variation (TV) regularizers corresponds to an assumption of *sparse gradients*, that is, of piecewise-smooth solutions with occasional step discontinuities. This is incorporated by adding the one of the following regularization energies at each pixel:

$$\Gamma_{TV}(\mathbf{x}) = \|\nabla p(\mathbf{x})\|_2 \quad (8)$$

$$\Gamma_{ATV}(\mathbf{x}) = \|\nabla p(\mathbf{x})\|_1 \quad (9)$$

$$\Gamma_{MTV}(\mathbf{x}) = \left\| \sum_{i=1}^3 \left( \|\nabla p^{(i)}(\mathbf{x})\|_2^2 \right) \right\|_2 \quad (10)$$



where Equation 8 is the standard TV, Equation 9 is a simple, anisotropic variant and Equation 10 is an anisotropic adaptation to color images [20]. The gradient terms are evaluated with first-order finite differences. TV regularizers are simple and generally effective regularizers that have the benefits of being convex.

**Sparse 1<sup>st</sup> and 2<sup>nd</sup> Order Derivatives.** We have also implemented a version of the regularizer introduced by Levin et al. [12], which uses a fractional (0.8) norm to enforce a heavy-tailed distribution for first and second order derivatives. We refer to that paper and the code posted on the corresponding project page for details.

**Gamma-corrected Sum of Absolute Differences.** Finally, we introduce a new regularizer that is designed to better deal with dark image regions. A standard problem with deconvolution algorithms is that the deconvolution has to be performed in linear intensity space, but the results have to be gamma corrected for viewing. The gamma curve, however, stretches the low intensity regions of the image disproportionately, thus amplifying noise in the solution. Together with the already low signal values, this results in poor signal-to-noise ratio in dark image regions.

Our approach is to introduce a regularizer that minimizes the data term in linear space, but ensures sparse gradients in the gamma-corrected image. To achieve this, we apply an gamma curve to the signal *before* evaluating a sum of absolute differences (SAD) regularizer in a  $3 \times 3$  window  $W$  centered at  $\mathbf{x}$ :

$$\Gamma(\mathbf{x}) = \sum_{i \in W(\mathbf{x})} |p(\mathbf{x}_i)^{\frac{1}{\gamma}} - p(\mathbf{x})^{\frac{1}{\gamma}}|, \quad (11)$$

with  $\gamma \approx 2$  to simulate a regular display gamma.

This regularizer is non-convex and would be non-trivial to design and implement a custom solver for, but is easily added to the Stochastic Deconvolution framework.

**Discontinuous and Data-Dependent Regularizers.** In Section 4 we demonstrate the flexibility of the Stochastic Deconvolution framework by experimenting with a data-dependent regularizer.

## 4. Results

The following sections present results comparing different regularization strategies and objective functions, as well as comparing to several existing methods. Runtimes vary based on PSF and image size but are typically only a few minutes. As an example, a 0.7 megapixel monochrome image with a  $21 \times 21$  PSF took 126 seconds with our unoptimized implementation.

**Comparison with Existing Methods.** Figures 2 and 3 show comparisons with the Coded Exposure Photography method of Raskar et al. [15]. With the addition of priors, Stochastic Deconvolution produces results with less noise and chromatic artifacts. However we note that this is expected given that their method is effectively unregularized. To illustrate the effect of different regularizers we show results for an enlarged area of the *train* image using the convex Total-Variation (TV) prior, the prior from Levin et al. [12], as well as the Gamma prior described in Section 3.1.

All three priors reduce the noise and chromatic artifacts present in the original results, however the two non-convex priors, (Figure 2(d) and 2(e)), provide the smoothest results. We note that our Gamma prior accomplishes its intended aim of reducing noise levels in darker regions, as can be seen by zooming in on the window and roof regions of Figures 2(d) and 2(e). We stress that it was straightforward to implement all of these priors in our common framework, while developing specialized solvers for each method would have taken significantly more effort.

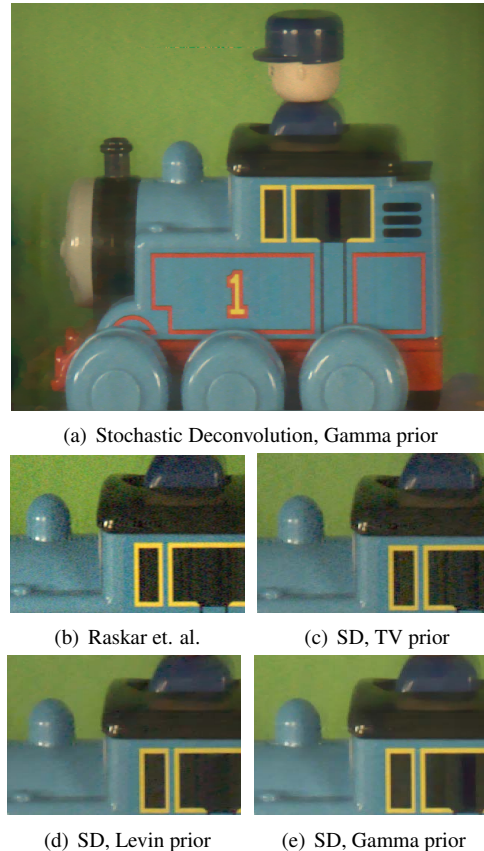


Figure 2. Comparison of Raskar et. al. (left) vs. Stochastic Deconvolution (right) using the regularizer of Levin et. al. Incorporation of the regularizer significantly reduces the noise in the reconstructed image while preserving image detail.

Figure 3 shows a comparison between Raskar et al. and Stochastic Deconvolution for the *white-car* image. We use the Gamma prior which reduces the noise and chromatic artifacts in dark regions such as the wheels and windows, while slightly improving the legibility of the text on the cab. We conclude that the Gamma prior is effective for preserving details and improving overall image quality.

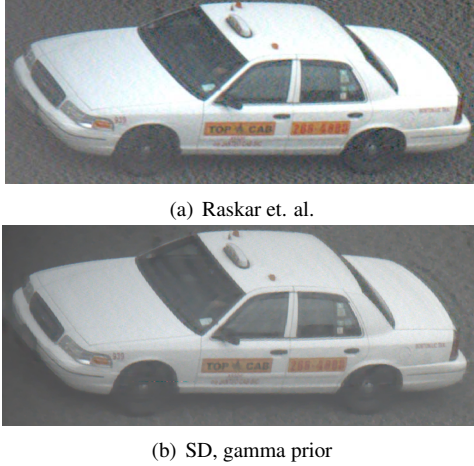


Figure 3. Comparison of Coded Exposure Photography (Raskar et al.) (top) to Stochastic Deconvolution (bottom). Addition of a prior helps to suppress noise and chromatic artifacts present in the original results, while improving the legibility of the text.

Figure 4 shows a comparison of deconvolution results using the method of Fergus et al. [4] with Stochastic Deconvolution. Stochastic Deconvolution produces sharper results with reduced ringing. Stochastic Deconvolution is also able to reconstruct the entire image right up to the image boundary through the use of the stochastic boundary condition.

Finally, we show a comparison of deconvolution results between the relatively recent method for large-blur removal of Xu and Jia [19] with Stochastic Deconvolution using Levin et al.’s prior. Our results are very comparable for this challenging dataset; both methods show minor artifacts throughout the image, however the results are very similar in terms of overall quality. Figure 6 highlights the effect of the stochastic boundary condition for inpainting plausible content in boundary regions, including additional windows and staircase details.

**Defocus Blur and Lens Aberrations.** We have also applied Stochastic Deconvolution to remove defocus blurs and lens aberrations in images taken with standard SLR cameras. Results comparing Stochastic Deconvolution using the Levin prior to the method of Levin et al. are shown in Figure 7. As expected, the results are very similar.

Figure 8 shows a color image blurred by a synthetic, wavelength dependent PSFs. Deblurring using the MTV

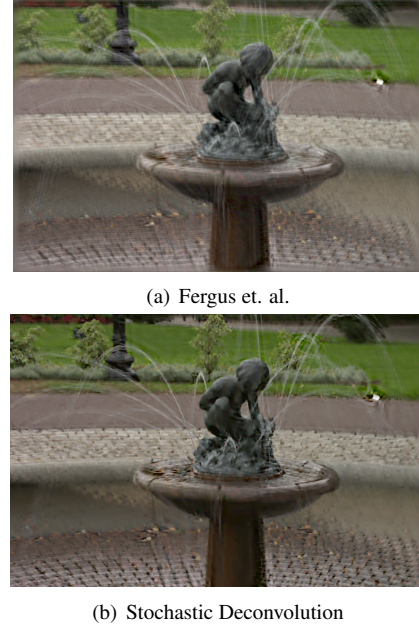


Figure 4. Comparison with the method of Fergus et. al. (top). The Stochastic Deconvolution result (bottom) shows substantially reduced ringing as well as much-improved handling of image boundaries due to the use of the Stochastic boundary condition.

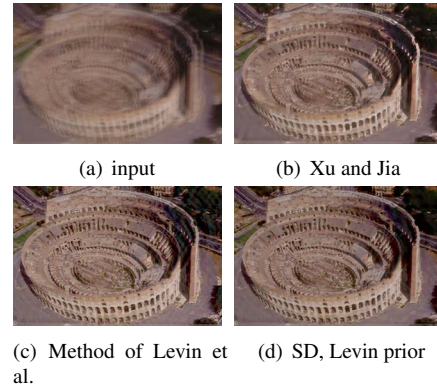


Figure 5. Non-blind deconvolution comparison with Xu and Jia (using kernels estimated by Xu and Jia) for the Roma image.

regularizer results in a slightly sharper image with reduced chromatic artifacts. Optimizing for such priors has been the focus of several papers, e.g. [20, 1], however they are easily implemented within our framework.

**Spatially Varying PSFs.** Due to the local nature of Stochastic Deconvolution, the deblurring problem can be relaxed from deconvolution to deblurring with spatially varying kernels. While many other deconvolution methods require subdividing the image into tiles with approximately constant PSF, in Stochastic Deconvolution every pixel can have its own distinct PSF. Figure 1 shows a synthetic example for deblurring results of a strong, spatially varying motion blur with rotational components about the optical

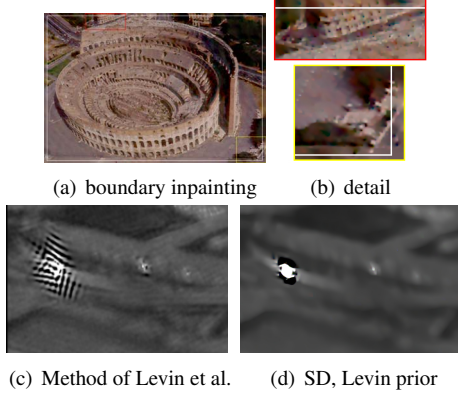


Figure 6. Top row: inpainted details from the stochastic boundary condition, windows are added to a building on the boundary (red outline) and staircase details outside the image are introduced (yellow outline). Zoom in to the top-left figure for additional features. Bottom row: the method of Levin et al. rings for highly saturated pixels, while masking these from the reconstruction produces considerably smaller artifacts.

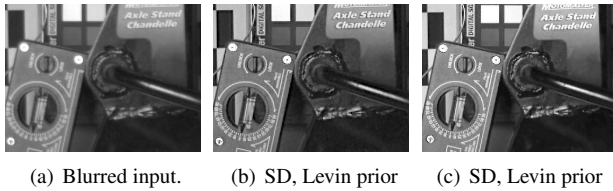


Figure 7. Comparison of method of Levin et al. with Stochastic Deconvolution for defocus blur from a standard SLR.



Figure 8. Comparison of per-channel TV (top) with the multichannel MTV prior (bottom) for a blur kernel with chromatic aberration. Image sharpness is slightly improved and color artifacts reduced around the tree branches.

axis. For real motion blur, one could obtain spatially varying PSFs either using estimation methods such as the one by Hirsch et al. [7], or using IMU sensors that are becoming increasingly available in cellphones and cameras [8].

**Data-Dependent Regularizers** We now provide a simple example of a discontinuous data-dependent regularizer. An image known to consist of only five colors is blurred and corrupted with noise. Deblurring with a TV regularizer yields gives an optimal peak PSNR value of 31.91 dB among all prior weights tested, however by clustering the image colors periodically and adding the  $L_1$  distance to the nearest cluster this can be improved by 0.7 dB while simultaneously reducing the weight on the TV term by an order

of magnitude. Although something of a contrived example, many applications can exploit similar domain-specific knowledge, an example being magnetic resonance imaging (MRI) where given tissue types and machine settings produce gray-values that are known a priori. Exploiting this knowledge can reduce reliance on heuristic priors, e.g. sparsity of gradients, and as illustrated above and in Figure 9, quantitatively improves reconstruction quality. However, such discontinuous, discrete-choice regularizers are problematic to implement effectively in conventional, gradient-based solvers.

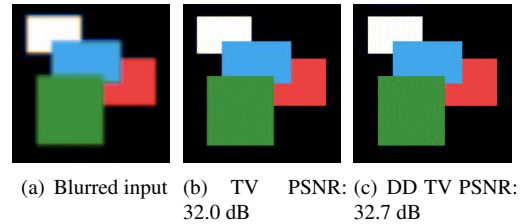


Figure 9. Simple data-dependent TV regularizer. Adding the  $L_1$  RGB distance to the nearest of one of five RGB clusters (computed by K-means) to a standard TV regularizer improves the best PSNR values by 0.7 dB over all parameter values

**Empirical Convergence** As a variant of coordinate-descent, our method has no theoretical convergence guarantees for general, non-smooth objectives. However, we have performed empirical convergence tests for the anisotropic TV regularizer and compared final objective values to the provably convergent method of Chambolle and Pock [2]. For the blurred image in Figure 10(a), the objective value computed by Stochastic Deconvolution after  $300 \times N_{pixels}$  mutations was 26.42, while the objective value by the primal-dual method was 26.69. We attribute the minor discrepancy to differences in boundary handling and termination criteria between the two methods. The objective function history is shown in Figure 10(c), showing a fast initial convergence rate that gradually flattens, as might be expected from a stochastic sub-gradient method. With that said, visual convergence is in practice very quick; by iteration 50 the gray-values are being adjusted by only 0.08% of the maximum range and are indistinguishable from the results at 300 iterations without careful, pixel-level examination.

## 5. Conclusions and Future Work

In this paper we have present Stochastic Deconvolution, a new, general-purpose method for the deconvolution problem based on stochastic random-walks. Stochastic Deconvolution is straightforward to implement, easily incorporates state-of-the-art priors and produces high-quality results.

The performance of our unoptimized implementation is currently comparable to other recent methods such as the



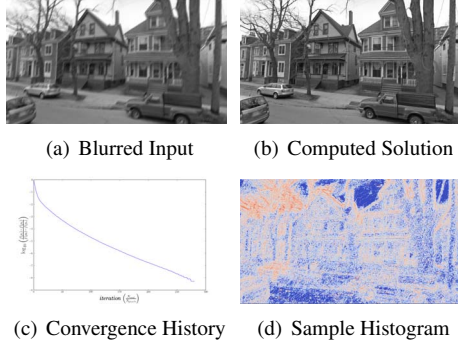


Figure 10. Convergence history of method down to  $e_d < 4 \times 10^{-9}$  for anisotropic TV regularizer with weight  $\lambda = 10^{-3}$ . Note that each Stochastic Deconvolution iteration has an approximately equal computational cost to one gradient-descent step using image-space convolutions but is able to focus sampling effort near details, as shown in the sampling histogram.

one by Levin et. al [12]. It is not currently competitive with methods working in Fourier space, or other highly optimized solvers. On the other hand, we gain the flexibility to not only incorporate arbitrary regularizers, but also to use spatially varying PSFs and modify the solver at boundaries and saturated pixels.

We also note that the algorithm as presented has significant potential for optimization, including parallelization on GPUs or multicore CPUs. For blurs with very large support, one could also adopt a multi-scale approach similar to Yuan et al. [22]. Since the primary cost of our method is the splatting of PSFs to update Equation 7, working in an image pyramid like this could significantly improve performance.

Overall, we believe that Stochastic Deconvolution is an excellent general-purpose method, particularly for evaluating the efficacy of new regularization strategies. Prior to Stochastic Deconvolution, trying a new prior often meant developing an entirely new solver. With Stochastic Deconvolution, new priors can be implemented and tested in a matter of minutes.

## Acknowledgments

This work was supported in part by the GRAND NCE and NSERC.

## References

- [1] P. Blomgren and T. Chan. Color tv: total variation methods for restoration of vector-valued images. *Image Processing, IEEE Transactions on*, 7(3):304–309, 1998.
- [2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
- [3] O. Cossairt and S. Nayar. Spectral focal sweep: Extended depth of field from chromatic aberrations sweep. In *Proc. ICCP*, 2010.
- [4] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. Freeman. Removing camera shake from a single photograph. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 2006.
- [5] B. Goldluecke and D. Cremers. Introducing total curvature for image processing. In *Proc. ICCV*, 2011.
- [6] J. Gregson, M. Krimerman, M. B. Hullin, and W. Heidrich. Stochastic tomography and its applications in 3D imaging of mixing fluids. *ACM Trans. Graph. (Proc. SIGGRAPH)*, pages 52:1–52:10, 2012.
- [7] M. Hirsch, C. Schuler, S. Harmeling, and B. Schlkopf. Fast removal of non-uniform camera shake. In *Proc. ICCV*, pages 463–470, 2011.
- [8] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 2010.
- [9] N. Joshi, C. Lawrence, Z. Richard, S. David, and J. Kriegman. Image deblurring and denoising using color priors. In *Proc. CVPR*, 2009.
- [10] N. Joshi, R. Szeliski, and D. J. Kriegman. PSF estimation using sharp edge prediction. In *Proc. CVPR*, 2008.
- [11] D. Krishnan and R. Fergus. Fast image deconvolution using Hyper-Laplacian priors. *Advances in Neural Information Processing Systems*, 22:1–9, 2009.
- [12] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM Trans. Graph. (Proc. SIGGRAPH)*, 2007.
- [13] Y. Li and S. Osher. Coordinate descent optimization for l1 minimization with application to compressed sensing; a greedy algorithm. *Inverse Problems and Imaging*, 3(3):487–503, 2009.
- [14] L. Lucy. An iterative technique for the rectification of observed distribution s. *The Astronomical Journal*, 79:745+, 1974.
- [15] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. In *ACM Trans. Graph. (Proc. SIGGRAPH)*, pages 795–804, 2006.
- [16] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62:55–59, 1972.
- [17] S. Sardy, A. Bruce, and P. Tseng. Block coordinate relaxation methods for nonparametric wavelet denoising. *Journal of computational and graphical statistics*, 9(2), 2000.
- [18] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.
- [19] L. Xu and J. Jia. Two-phase kernel estimation for robust motion deblurring. In *Proc. ECCV*, pages 157–170, 2010.
- [20] J. Yang, W. Yin, Y. Zhang, and Y. Wang. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM Journal on Imaging Sciences*, 2(2):569–592, 2009.
- [21] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 2007.
- [22] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Progressive inter-scale and intra-scale non-blind image deconvolution. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 2008.