Robust deconvolution of natural images using multiple captures.

Anonymous Author(s)
Affiliation
Address
email

Abstract

Blur in images can appear in a wide range of imaging applications. In consumer photography images are often degraded by motion blur due to camera shake, in medical and aerospace imaging imperfections of the lens system used are significant causes of blur. Commonly the blur is modeled as a 2D-convolution of the underlying sharp image with a 2D blur kernel resulting in the observed blurred image. Assuming the blur kernel to be known (or estimated) various methods exist that reconstruct an estimate of the sharp image from a captured blurred image. Commonly only one blurred image is used for this deconvolution. Now, in many practical applications image sequences of the same scene are available. For example a sequence of low exposure captures of the scene or subsequent video frames. This paper presents an approach of how to incorporate this multiple capture information in the deconvolution of natural scene images. It is shown that using multiple captures can significantly improve the results. During a capture strong noise (e.g. shot noise in a low-exposure image) may corrupt the image and also estimated blur kernels might be noisy due to an imperfect estimation method. The proposed new approach is robust to noise in the captures and blur estimate.

Figure 1: Left: Original sharp image. Middle: Two synthetically blurred noisy images (blur kernel in top left of first blurred image). Right: Deconvolution result of the proposed method using the two blurred images and a noisy blur kernel estimate.

1 Related work

1.1 Model for the blur

Consider a simplified camera model that consists of an image sensor and a lens between the sensor and the object to capture. The object is illuminated by some light source and rays of light hitting the object surface are reflected on each surface point, traveling through the lens until finally hitting the image sensor. In a perfect imaging system the lens changes the direction of incoming rays so that all the rays coming from one object point in the scene converge to exactly one point on the image sensor, thus we get a perfect image (one-to-one mapping) from our scene (points). Blur can now be defined as the effect that rays of one object point do *not* converge in exactly one sensor point, that is they are scattered over a region on the image sensor.

Imaging sensors consist of an array of sensor elements (pixels) that count the number of incoming photons, thus the pixel value in a captured image depends linearly on the number of measured photons (assuming that the image is not gamma corrected).

Using these two general properties of imaging systems discussed in the paragraphs above, we can formulate blurring with the following linear model. Be $B \in \mathbb{R}^{n\times m}$ the blurred image, $k \in \mathbb{R}^{d\times e}$ the blur kernel and $F \in \mathbb{R}^{n\times m}$ the underlying sharp image, then it is:

$$B = k \otimes F + N \quad \text{that is}$$
$$B(r,c) = \sum_{i=-d/2}^{d/2} \sum_{j=-e/2}^{e/2} k(i,j)F(r-i,c-j) + N(r,c) \quad \text{with} \quad \begin{array}{l} 0 \le r \le n-1 \\ 0 \le c \le m-1 \end{array} \tag{1}$$

where $\otimes$ is the 2D convolution operator and $N \in \mathbb{R}^{n\times m}$ noise corrupting the blurred image during the capture. The blur kernel k describes here the scattering of the rays, which would normally converge in a single sensor point in an optimal imaging system, to its neighboring (defocused) positions. k is centered around its spatial position $(0,0)$. By using a convolution Eq. (1) makes implicitly the assumption that the blur does not vary spatially over the image which holds for many practical applications. A spatially invariant blur kernel (as discussed throughout this paper) is commonly called point-spread-function (PSF).

1.2 Known deblurring approaches

The "deconvolution problem" is now defined as solving the inverse problem of Eq. (1) given B,k to determine F. It can be shown that this problem is ill-posed in the sense of Hadamar [1]. An intuitive understanding of the ill-posedness can be gained with trying to solve Eq. (1) for B in the fourier domain:

$$\text{With} \quad \hat{B} = \hat{k} \cdot \hat{F} + N \quad \text{it is} \quad \frac{1}{\hat{k}}\hat{B} = \hat{F} + \frac{\hat{N}}{\hat{k}} \,. \tag{2}$$

where the hat denotes the Fourier transform. Now assume $\hat{k}$ to be zero or small in some spatial frequencies (e.g. that happens for a motion-blur kernel [1]). Since the noise N does not also need to be zero in these frequencies it is strongly amplified and corrupts the solution.

Because of the ill-posedness of the deconvolution problem, many different approaches exist that try to cure this ill-posedness by using different regularizations. Common approaches are discussed in the following paragraphs:

1.2.1 Tikhonov regularization

The Tikhonov regularization places an ℓ_2-norm prior on the unknown sharp image F. The following minimization problem is solved:

$$\operatorname*{argmin}_{\vec{f}} \; \|\vec{b} - K\vec{f}\|_2^2 + \lambda^2 \|L\vec{f}\|_2^2 \,, \tag{3}$$

where $\lambda \in \mathbb{R}$ is a regularization factor that weights the two terms in the minimization. Eq. (3) is formulated with the images now as $nm \times 1$-vectors consisting of the columns of the respective image stacked in sequence. $K \in \mathbb{R}^{nm\times nm}$ is the blurring matrix according to k that defines in each row the blur kernel for the considered pixel indiced by the number of that row. L is a convolution matrix usually set to the identity matrix. Using the vector equality $\|x\|_2^2 = x^T x$, deriving with respect to F and setting to zero gives:

$$(K^T K + \lambda^2 L^T L)\vec{f} = K^T \vec{b} \tag{4}$$

1.2.2 Regularization using natural image statistics

The ℓ_2-norm prior of the Tikhonov regularization is a heuristic approach which only uses the prior assumption that the solution F has low energy. Significantly better results can be achieved by using priors that are tailored to the statistics of natural images, see [2] and its application in [3]. From research on natural image statistics it is known that images of the real-world follow a gradient magnitude distribution that is heavy-tailed [4], [5] (although the color distribution may vary significantly in different images). Intuitively that is because the real-world consists of many large smooth surfaces (without strong discontinuities and thus small gradient magnitude which causes the mass of the gradient magnitude distribution to be centered around 0) and the small area of their bounds (strong gradient magnitude which causes heavy tails).

Now in [2] the MAP estimate F_{opt} for F given B (and given k assumed to be fixed) is computed using a heavy tailed gradient magnitude prior for F (again the image forming from Eq. (1) is assumed):

$$F_{opt} = \underset{F}{\operatorname{argmax}}\ P(F|B) = \underset{F}{\operatorname{argmax}}\ P(B|F)P(F) \quad \text{, since} \quad P(F|B) = \frac{P(B|F)P(F)}{P(B)} \tag{5}$$

The heavy-tailed gradient magnitude prior is modeled in [2] as:

$$P(F) \propto e^{-\alpha \sum_k \rho(g_k \otimes F)} \tag{6}$$

with g_k as derivative filters and $\alpha \in \mathbb{R}$ as a weight for the prior. The prior is made sparse by using $\rho(z) = \|z\|^{0.8}$ (using an ℓ_1 norm as in the lasso-method would be also a possible choice). Assuming that the noise in the image capture (N from Eq. (1)) is gaussian distributed with variance σ^2 the likelihood can be formulated as $P(B|F) \propto \exp\left(-\frac{1}{2\sigma^2}\|B - k \otimes F\|_2^2\right)$. Having the prior and likelihood defined as above, Eq. (5) becomes after taking the log:

$$F_{opt} = \underset{F}{\operatorname{argmin}}\ \|B - k \otimes F\|_2^2 + \lambda \sum_k \rho\left(g_k \otimes F\right) \tag{7}$$

where $\lambda = 2\alpha\sigma^2$ weights the two minimization terms and is typically specified by the user. The minimization is done by an IRLS method (discussed in detail later in this paper).

2 Overview

This paper presents a novel approach to exploit multiple capture data in the deconvolution of natural images. The method is robust to noise in the images and PSF. It is assumed here that the true PSF is not known in the deconvolution, but rather a noisy PSF estimate (e.g. estimated by the method in [6]).

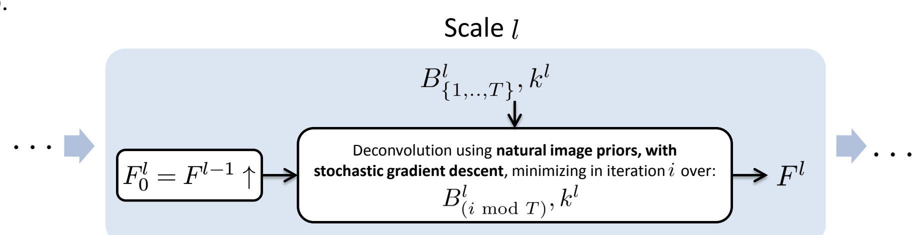

Figure 2: Overview of the proposed deconvolution method in scale space.

To be robust to noise in the PSF and blurred image data the deconvolution is done progressively in scale-space from coarse to fine (similar to [7]). That is, first image pyramids $\{B^l_{\{1,..,T\}}\}^L_{l=0}$, $\{k^l\}^L_{l=0}$ for each image B_i (with $i \in \{1, .., T\}$) of the T blurred multiple capture images and for the (noisy) PSF estimate k are generated by repeated bicubic downsampling of $B_{\{1,..,T\}}$, k with the scale factor $1/\sqrt{2}$. The downsampling is done until the size of the downsampled k becomes smaller or equal to 3×3, which implicitly determines the number of scales L. Scale L denotes here the finest scale and 0 the coarsest scale.

The bicubic downsampling is essentially a low-pass filtering and resampling. Thus, by each downsampling a part of the (high-frequency) noise is removed from the image $B_{\{1,..,T\}}$ and kernel k which enables to solve the deconvolution problem more robustly (follows from Eq. (2)). However, along with noise, also (high-frequency) image details and PSF details are discarded during the low-pass filtering in the downsampling. Therefore, the idea is to solve the deconvolution problem successively on each scale from coarse to fine using the result of a coarser scale as initialization of the next finer scale (see Fig. 2). On each scale l the blurred images $B^l_{\{1,..,T\}}$ are deconvolved with

k^l yielding the sharp image estimate F^l on that scale. The upsampled deconvolution result F^{l-1} of the next coarser scale $l-1$ is used as the initialization $F_0^l = F^{l-1}\uparrow$ of the deconvolution. Here the operator $\uparrow$ denotes (bicubic) upsampling to the next scale. Using the scale-space as described enables a robust (to noise) deconvolution that is not stuck in possible local minima (see [6]).

The deconvolution in each scale is done using the natural image prior from [2] that was described in Sec. 1.2.2. So, Eq. (7) is solved, where the minimization is now done jointly over multiple blurred images captures. This is achieved by doing a stochastic gradient descent iteration where the minimization is done alternately over the images in a round-robin fashion. This approach is described in detail below.

With the outlined deconvolution procedure, the sharp image pyramid $\{F^l\}_{l=0}^{L}$ is recovered successively from coarse to fine. The final result is the estimate F^L on the finest level.

3 Deconvolution using natural image priors on a single scale

To solve the minimization problem from Eq. (7), in [2] an iteratively reweighted least squares approach (IRLS) (see [8] for an introduction) is proposed. The following minimization problem is solved by the IRLS method:

$$\vec{f}_{opt} = \underset{\vec{f}}{\operatorname{argmin}} \sum_j \rho\left(A_j \vec{f} - b_j\right) \tag{8}$$

with the same matrix notation for the images and kernels as used in Eq. (3). The matrices A_j are simply all the convolution kernels k,g_k in matrix form and the b_j chosen as the according $\vec{b}$ or 0 to make Eq. (8) consistent with Eq. (7). The minimization problem Eq. (8) is then solved by the following IRLS iteration (as described in [8]):

Algorithm 1. (IRLS)

$w_j^0 = 1, i = 0$ *(Initialization of weights w_j)*

do

$\quad i = i + 1$

$\quad \overline{A} = \sum_j A_j^T w_j^{i-1} A_j,\ \ \overline{b} = \sum_j A_j^T w_j^{i-1} b_j$

$\quad x^i = \overline{A}^{-1}\overline{b}$ *(Solving the lin. system $\overline{A}x^i = \overline{b}$ for x^i.)*

$\quad u_j = A_j x^i - b_j$

$\quad w_j^i(u_j) = \frac{1}{u_j}\frac{\partial\rho(u_j)}{\partial u} \approx \max(|u_j|, \epsilon)^{0.8-2}$

while $\left\|(x^i - x^{i-1})\right\|_2^2 / \left\|x^i\right\|_2^2 > \delta_{IRLS}$

$\vec{f}_{opt} = x^i$

with δ_{IRLS} as a termination constant. In [2] the linear system $\overline{A}x_i = \overline{b}$ in each iteration is solved by conjugate gradient method.

The approximation $|u_j| \approx \max(|u_j|, \epsilon)$ with ϵ close to 0 is made in the last step of each iteration to avoid the division by very small values that would generate infinite weights. This is a common approach to stabilize the IRLS process. It is important to note that this does not solve exactly Eq. (7), but is only an approximation (which depends on the choice of ϵ). In [9] a very similar approach of stabilizing IRLS with ϵ as a damping weight is discussed. The authors show that for compressed sensing a significantly higher signal recovery rate can be achieved by iteratively repeating algorithm 1 (the do-while-loop) for a successively decreasing ϵ. This behavior was reproduced for the application here as well - high fixed values of ϵ generate a large approximation error, while low values make the IRLS more instable. Therefore, the approach from [9] is adopted here. Starting with $\epsilon = 10^{-2}$, the IRLS-loop in algorithm 1 is repeated decreasing ϵ by a factor of 10 after every repetition until $\epsilon <= 10^{-5}$.

The linear system $\overline{A}x^i = \overline{b}$ in each iteration of algorithm 1 is now solved by the following stochastic gradient descent algorithm which jointly minimizes over the multiple captures.

Algorithm 2. (Stochastic gradient descent)

$z^0 = \frac{1}{T}\sum_{v=1}^{T}\overline{b}_v, j = 0$

do

$j = j + 1$

$r^j = \overline{b}_{(j \bmod T)} - \overline{A}z^{j-1}$

$\alpha^j = \frac{(r^j)^T r^j}{(r^j)^T \overline{A} r^j}$

$z^j = z^{j-1} + \alpha^j r^j$

while $\left\|(z^j - z^{j-1})\right\|_2^2 / \left\|z^j\right\|_2^2 > \delta_{SGD}$

$x^i = z^j$

Since we have T multiple captures $B_{\{1,..,T\}}$, also T according multiple $\overline{b}_{\{1,..,T\}}$ exist. Algorithm 2 is a modification of the steepest descent iteration as defined in [10] (see the paper for details). The modification is here that the residual of the current solution z^j is minimized alternately between all captures $\overline{b}_{\{1,..,T\}}$ in a round robin fashion. The final result x^i is a joint minimizer for all $\overline{b}_{\{1,..,T\}}$. As an initialization the mean over all $\overline{b}_{\{1,..,T\}}$ is chosen. In this work no formal proof of the convergence of algorithm 2 is given. However, in all practical tests done for this project the iteration converged.

4 Results

To evaluate the proposed approach synthetic data is used. Since the ground truth sharp image F is known, a quantitative analysis of the quality of the deconvolution results can be presented (that would not be possible for real-world images where F is unknown). Figure 3 shows a result of the proposed method and further methods for comparison.

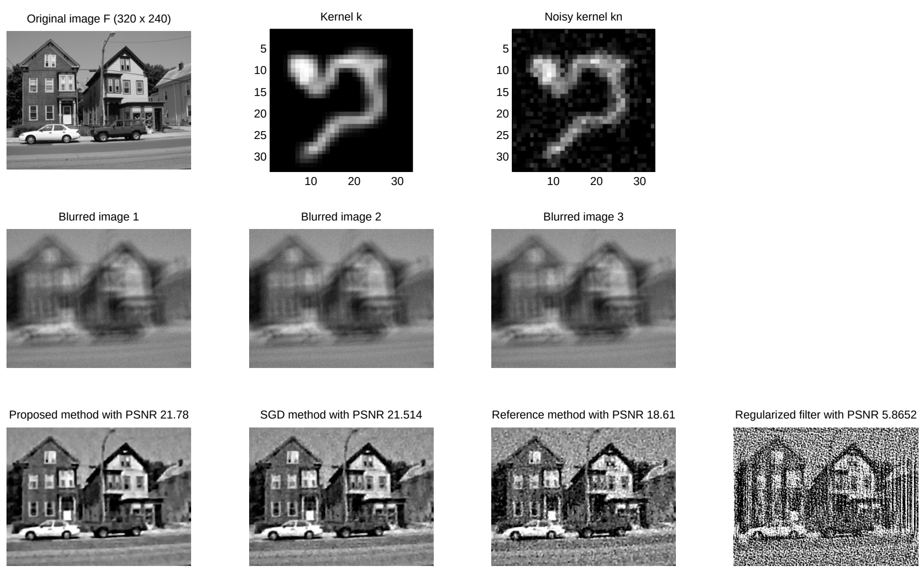

Figure 3: Results for synthetic example data.

In the first row the original ground truth image F and the original (large 33×33) blur kernel k are shown. Three blurred images shown in the second row are generated synthetically by the already discussed Eq. (1) $B = k \otimes F + N$, where here $N \propto \mathcal{N}(0, 5 \cdot 10^{-5})$. Now, to make the test-case presented here as realistic as possible it is assumed that the true kernel k is not known, but only a noisy estimate $kn = k + N_k$ with $N_k \propto \mathcal{N}(0, 5 \cdot 10^{-7})$. Deconvolution results of (one or all) blurred images and kn are shown in the last row of Fig. 3. The quality of the results is measured with the frequently used peak-signal-to-noise-ratio (PSNR), see [1]. It is the ratio between the maximum possible signal power and the mean square error of the reconstruction, interpreted as noise power corrupting the signal. This ratio is expressed using logarithmic decibel scale:

$$\text{PSNR}(F_{opt}) = 10 \log_{10} \frac{1}{\frac{1}{nm}\|\vec{F} - \vec{F}_{opt}\|_2^2} \text{ [dB]} \tag{9}$$

for $n \times m$ images. The third image in the last row of Fig. 3 shows for reference the result of applying the natural image statistics method from [2], discussed in Sec. 1.2.2. Only the first blurred image was used as input data and the parameters proposed in the original paper [2] have been used. The last image shows the reconstruction with a regularized least-squares filter using the Matlab function "deconvreg" with default parameter settings. It is considered here as a standard approach to deconvolution. The regularized least-squares filter minimizes a ℓ_2-norm-based regularization problem similar to Tikhonov regularization (see MATLAB doc). As shown by the PSNR values (and visual comparison) of the reconstructions, using the natural image prior enables to reconstruct significantly more details while achieving lower noise in the reconstructions. The second image in the last row of Fig. 3 shows the proposed stochastic gradient descent method, but without using the scale space. All the three blurred images and the parameters $\delta_{\text{IRLS}} = \delta_{\text{SGD}} = 10^{-3}$ have been used in the deconvolution. To limit the execution time the IRLS loop is terminated after just 2 iterations (that was sufficient to get good results for a set of test-images). Comparing the deconvolved image to the reference deconvolution image shows clearly that using multiple blurred capture images can significantly improve the deblurring results. A visible lower noise level (see also PSNR values) is achieved. The first deblurred image in the last row of Fig. 3 has been computed with the full method as proposed in this paper. Comparing it to the second image shows that exploiting scale-space in the discussed approach can again yield better results. Since the noise is reduced a higher PSNR value is achieved.

5 Conclusion

This paper has introduced a new deconvolution method exploiting multiple capture data. Is has been shown that incorporating the multiple images in the deconvolution operation significantly improves image quality of the reconstructions. Remaining noise is the reconstructions is further damped by exploiting scale-space. The proposed method is robust to noise in the PSF estimate and blurred images.

References

[1] Gibson, Jerry D. and Al Bovik (editors): *Handbook of Image and Video Processing*. Academic Press, Inc., Orlando, FL, USA, 1st edition, 2000.

[2] Levin, Anat, Rob Fergus, Frédo Durand, and William T. Freeman: *Deconvolution using natural image priors*, 2007.

[3] Levin, Anat, Rob Fergus, Frédo Durand, and William T. Freeman: *Image and depth from a conventional camera with a coded aperture*. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[4] Field, David J.: *What is the goal of sensory coding?* Neural Computation, 6(4):559–601, 1994.

[5] Fergus, Rob, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman: *Removing camera shake from a single photograph*. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 787–794, New York, NY, USA, 2006. ACM.

[6] Yuan, Lu, Jian , Long Quan, and Heung Yeung Shum: *Image deblurring with blurred/noisy image pairs*. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[7] Yuan, Lu, Jian Sun, Long Quan, and Heung Yeung Shum: *Progressive inter-scale and intra-scale non-blind image deconvolution*. In *ACM SIGGRAPH 2008 papers*, SIGGRAPH '08, pages 74:1–74:10, New York, NY, USA, 2008.

[8] Meer, Peter: *Emerging Topics in Computer Vision*, chapter Robust Techniques for Computer Vision. Prentice Hall, 2004.

[9] Chartrand, Rick and Wotao Yin: *Iteratively reweighted algorithms for compressive sensing*. In *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

[10] Shewchuk, Jonathan R.: *An introduction to the conjugate gradient method without the agonizing pain*. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 1994.