ROBUST INTER-SCALE NON-BLIND IMAGE MOTION DEBLURRING

*Chao Wang,LiFeng Sun,ZhuoYuan Chen,ShiQiang Yang**

Computer Science,Tsinghua Univ.,China.
Tsinghua National Laboratory for Information Science and Technology. Key Laboratory of Media and Networking, MOE-Microsoft.
w-c05@mails.tsinghua.edu.cn

JianWei Zhang †

Department Informatics,
Hamburg Univ.,Germany

ABSTRACT

Kernel estimate errors and image noise are major causes of visual artifacts in image motion deblurring. We propose an inter-scale non-blind image motion deblurring approach that significantly reduces those artifacts. We use Gaussian Scale Mixture Field of Experts (GSM FOE) model as image prior. The inter-scale smoothness constraint is adopted to suppress the ringing artifacts. In each scale, image details are recovered by the residual deconvolution and the cross bilateral filter (CBF). We further propose a std-controlled CBF to denoise the result. The experimental results are much better than those of previous methods.

Index Terms— Image deblurring

1. INTRODUCTION

Recovering the latent unblurred image from a motion-blurred photograph is called image deconvolution/deblurring, which is an important and chronic problem for many scientific applications, such as medical imaging and consumer photography. If the motion blur kernel is assumed to be known, this problem is defined as non-blind deconvolution. Non-blind deconvolution is highly illposed since the blur operator is often lowpass and the high frequency image details are cut off. Direct inverse of the blur kernel will excessively amplify the image noise and kernel estimate errors at the lost frequencies, resulting in the notorious ringing artifacts around the sharp edges. Since the magnitude of ringing artifact is usually proportional to the amplitude of image edge [1], the key to suppress the ringing is to accurately locate the edges of the latent image. Traditional methods [2] can not achieve satisfying results for severely blurred images because they use the blurred image to locate the edges which is very difficult. The inter-scale and intra-scale non-blind deconvolution algorithm (ISISD) [3] overcomes this weakness by performing deconvolution in scale space, and utilize the coarser scale to locate the edges of the finer scale. It produces excellent results even for images with large blurs. But its results contains lots of noise because it recovers image details by using the residual deconvolution [1] which mistakes image noise as tiny edges.

The above methods assume the blur kernel has been accurately estimated. However, accurate kernel is not always available and even small kernel errors result in significant artifacts [4]. To suppress the ringing artifacts caused by the kernel errors, Shan et al.[4] propose a smoothness constraint method. Very nice results are obtained when the kernel is small (e.g. 30×30 pixels or fewer). But the ringing will show up again for a large blur because the smoothness constraint also utilizes the blurred edges to locate the latent edges.

In this paper, we propose a novel inter-scale non-blind deconvolution that produces high-quality deblurred result even from many challenging blurred images given noisy kernels. Our algorithm is performed in scale space. First, we adopt the Gaussian Scale Mixture Field of Experts (GSM FOE) model [5] as the latent image prior. GSM FOE model greatly help restore image details while smoothing out noise since it captures most image energies [5]. Second, based on the temporal optimized result, we adopt the residual deconvolution and the cross bilateral filter (CBF) [1] to restore more image details, and propose a std-controlled cross bilateral filter to remove image noise. Finally, we adopt the intra-scale and inter-scale smoothness constraints to suppress the ringing artifacts.

2. OPTIMIZATION IN EACH SCALE

For the input blurred image I and blur kernel K, we build an image pyramid $\{I^s\}_{s=1}^S$ and kernel pyramid $\{K^s\}_{s=1}^S$ using bicubic downsampling and scale factor of $\sqrt{2}$ until the kernel size at the coarsest level reaches 9×9. We progressively restore a pyramid $\{L^s\}_{s=1}^S$ for the latent image L from coarse to fine. At each scale, we assume the following equation holds: $I^s = K^s \otimes L^s + n^s$, where $\otimes$ denotes the con-

*Thanks to the National Basic Research Program (973) of China under Grant No.2006CB303103, the National Natural Science Foundation of China under Grant No. 60833009, the National High-Tech Research and Development Plan (863) of China under Grant No.2006AA01Z118 for funding.

†Thanks to the cross-modal interaction in natural and artificial cognitive systems (CINACS) for funding.

volution operator and n^s denotes the additive Gaussian noise. For simplicity, we omit s in the following forms except for the inter-scale smoothness constraint. In Bayes framework, we write the posterior distribution over the unknowns L as:

$$p(L|I) \propto p(I|L)p(L), \tag{1}$$

where $p(I|L)$ denotes the likelihood of noise $n = I - K \otimes L$, and $p(L)$ represents the image prior. In each scale, our goal is to solve for the maximum a-posteriori (MAP) solution, which finds the latent image L that maximizes $p(L|I)$.

For the likelihood of noise, we adopt Shan et al's model[4]:

$$p(I|L) = \prod_{\partial^* \in \Theta} \prod_i N((\partial^* I)(i)|\partial^*(K \otimes L)(i), \sigma_q^2), \tag{2}$$

where ∂^* represents the operator of any partial derivative with $\kappa(\partial^*) = q$ representing its order. $\partial^* n$ follows a Gaussian distribution with standard deviation $\sigma_q = \sqrt{2^q}\sigma_0$, where σ_0 denotes the standard deviation of n. $\Theta = \{\partial^0, \partial_x, \partial_y, \partial_{xx}, \partial_{yy}, \partial_{xy}\}$ represents a set of partial derivative operators [4].

For the image prior, we adopt GSM FOE model [5] as global prior $p_g(L)$ to reduce the ill-posedness, and the intra-scale and inter-scale smoothness constraint as local prior $p_l(L)$ to suppress the ringing artifacts, i.e., $p(L) = p_g(L)p_l(L)$. GSM FOE is defined as: $p_g(L) = \frac{1}{Z}\prod_{i,m} \Psi(\omega_m^T L(i))$, where $\Psi(x) = \sum_{j=1}^{J} \frac{\pi_j}{\sigma_j} \exp(-\frac{x^2}{2\sigma_j^2})$. $\Psi()$ is a Gaussian Scale Mixture function, Z is the partition function and ω_m is the linear filter. i indexes over image pixels, j indexes over GSM elements and m indexes over linear filters. $\omega_m^T L(i)$ is the result of applying the linear filter ω_m to image L at location i. We fix the GSM function $\Psi()$ to the one in [5] and train the linear filters ω_m for different scales of the image pyramid using the patches sampled from the training set [5]. To solve the MAP problem of (1) in frequency domain, we need to take the logarithm of all probability terms, and the log-prior $log(p_g(L)) = log(\frac{1}{Z}\prod_{i,m} \Psi(\omega_m^T L(i)))$ is, however, very challenging to optimize. Similar to Shan et al [4], we overcome this difficulty by fitting the log-GSM function $log(\Psi(x))$ using concatenated three piece-wise convex functions:$\Phi(x) = \begin{cases} -k_1|x| - c_1 & |x| \le l_t^1 \\ -k_2|x| - c_2 & l_t^1 < |x| \le l_t^2 \\ -(ax^2 + b) & |x| > l_t^2. \end{cases}$

$l_t^1 = 0.06$ and $l_t^2 = 0.2$ indexes the positions where the three functions are concatenated. The fitting parameters are $k_1 = 73.995$, $c_1 = 2.154$, $k_2 = 16.041$, $c_2 = 5.0515$, $a = 6.488$ and $b = 8.688$. Then the global prior $p_g(L)$ becomes: $p_g(L) \propto \prod_{i,m} exp(\Phi(\omega_m^T L(i)))$.

We adopt the smoothness constraint [4] to suppress the ringing artifacts at the coarsest scale. The smoothness constraint enforces the blurred image gradients to be similar to the latent image gradients in the smooth area Ω, where pixel value's local standard deviation is smaller than a threshold which is set to 0.02. It is well known that given the gradients, the pixel values can be figured out by solving the linear Poisson function [6]. Therefore the similarity between gradients indicates the similarity between pixel values. It further follows that the similarity between pixel values indicates the similarity between their responses to one same linear filter. So the smoothness constraint can be imposed on the outputs of the linear filters learned by GSM FOE model. The differences between corresponding outputs are defined to follow a Gaussian distribution with zero mean and standard deviation σ_l. Then the smoothness constraint is defined as:

$$p_l(L) = \prod_m \prod_{i \in \Omega} N(\omega_m^T L(i) - \omega_m^T I(i)|0, (\sigma_l)^2). \tag{3}$$

For the finer scales, since the up-sampled deblurred image at previous scale is a good approximation to the current scale, we use the smooth area of previous scale for the inter-scale constraint. We set 0.06 to the threshold of local standard deviation to select a large smooth area. The local prior becomes:

$$p_1(L^s) = \prod_m \prod_{i \in \Omega} N(\omega_m^T L^s(i) - \omega_m^T \overrightarrow{L}^s(i)|0, (\sigma_l^s)^2), \tag{4}$$

where $\overrightarrow{L}^s = I^s$ if $s = 1$, and $\overrightarrow{L}^s = L^{s-1}\uparrow$ otherwise. $L^{s-1}\uparrow$ denotes up-sampled L^{s-1}.

The MAP solution of L at each scale can be obtained by minimizing the energy: $E_L = -ln(p(L|I))$. By taking the priors on noise and image into (1), we get

$$\begin{aligned} E_L = &(\sum_{\partial^* \in \Theta} \frac{1}{2\sigma_q^2} \|\partial^* L \otimes K - \partial^* I\|_2^2) + \\ &\|\sum_m \Phi(\omega_m^T L)\|_1 + \sum_m \frac{1}{2\sigma_l^2} (\|\omega_m^T L - \omega_m^T \overrightarrow{L}\|_2^2 \circ M), \end{aligned} \tag{5}$$

where $\circ$ denotes the element-wise multiplication operator and $\|\cdot\|_p$ represents the p-norm operator. M is a $2-D$ binary mask to encode the area for smoothness constraint or inter-scale smoothness constraint. For any element $m_i \in M$, $m_i = 1$ if the corresponding pixel $\overrightarrow{L_i} \in \Omega$, and $m_i = 0$ otherwise. To minimize (5), we adopt the variable substitution scheme as well as the iterative parameter re-weighting technique of Shan et al.[3]. First, we substitute a set of auxiliary variables Ψ_m for $\omega_m^T L$, and add an extra constraint $\Psi_m \approx \omega_m^T L$ for each m. Then we iterate between updating Ψ_m and L while the other is held fixed until convergence [3].

Fig. 1 demonstrates the impacts of our novel image priors by the comparison between Shan et al's method [4] and our algorithm. Shan et al's method adopts the sparse derivative prior and the intra-scale smoothness constraint, while our algorithm uses the GSM FOE prior together with the inter-scale smoothness constraint prior. The results show that our powerful priors reveal more image details while smoothing out more ringing artifacts and image noise.

3. DETAILS RESTORATION

Denoting the temporal result in Fig. 1(d) by L^t, now our aim is to recover the residual image ΔL, so that the final result be-

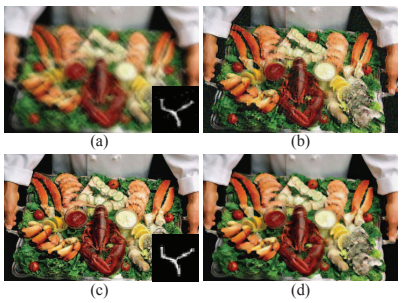

Fig. 1. (a) Synthesized blurred image and the blur kernel estimated by Shan et al.[4]. The standard variance of the added Gaussian noise equals 0.02. (b) Result by Shan et al.[4]. (c) Benchmark image and kernel of 41×41 pixels. (d) Our temporal optimized result.

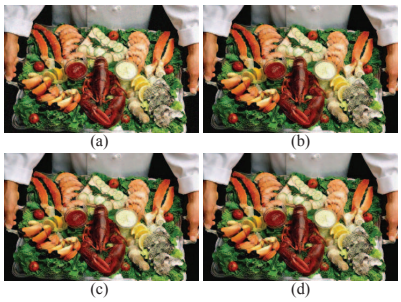

Fig. 2. (a) Result by residual deconvolution. (b) Result by transferring the details from (a) to the image in Fig. 1(d) using the CBF. (c) Denoised result by the CBF. (d) Denoised result by the std-controlled CBF.

comes $L^f = L^t + \Delta L$. To well recover ΔL while constraining the ringing, we adopt the residual deconvolution (RD) and the cross bilateral filter (CBF) [1]. The RD restores more tiny details but still brings into some small ringing artifacts as shown in Fig. 2(a). The CBF transmits the fine details from Fig. 2(a) to L^t. The final result L^f is shown in Fig. 2(b).

L^f still contains lots of image noise since the RD recovers not only image details but also noise. We observe that L^t contains little noise, benefiting from the powerful image prior and noise model. Then L^t can be used as the edge-preserving term to help remove the noise from L^f. We adopt the CBF [7] for this task. The denoised result L^d at pixel i by the CBF using L^t and L^f is computed by:

$$L_i^d = \frac{1}{Z_i} \sum_{j \in \Omega(i)} g_{\sigma_d}(|j|) g_{\sigma_r}(|L_{i+j}^t - L_i^t|) L_{i+j}^f, \qquad (6)$$

where g_{σ_d} and g_{σ_r} are the spatial and radiometric Gaussian functions with widths controlled by the standard deviation parameters σ_d and σ_r respectively. j is an offset relative to i in the filter kernel $\Omega(i)$, and Z_i is a normalization factor: $Z_i = \sum_{j \in \Omega(i)} g_{\sigma_d}(|j|) g_{\sigma_r}(|L_{i+j}^t - L_i^t|)$. In practice, σ_d is set to the value to cover a pixel neighborhood about 30 pixels [7]. The challenge is to set σ_r so that the noise is smoothed out but details are preserved. However, as shown in Fig. 2(c), even after carefully tuning the parameters, the CBF tends to either over-blur or under-blur the image for some regions because some details have been lost in the edge-preserving image L^t. We observe that in L^t, the pixel values in the smooth regions should be more reliable than those in the highly textured regions since the motion blur can generally be considered a smooth filtering process [4]. Further more, human perception tolerates small noise in highly textured regions. Based on this observation, we adaptively set σ_r according to the standard deviation (std) of pixel values in the local $2\sigma_d \times 2\sigma_d$ window. Large std indicates complex textures, so that we should set small values to σ_r to preserve more details. Otherwise we should assign σ_r with large values to remove more noise for small std. The adaptive σ_r is computed by: $\sigma_{r,i} = \frac{0.04}{1+\upsilon\sigma_i^t}$, where σ_i^t denotes the local std at pixel i in L^t, and υ is the parameter to determine the value domain of σ_r. We experimentally adjust υ and find $\upsilon = 99$ produces the visually best result. υ can also be obtained by experiments on some synthesized examples to produce largest PSNR. The denoised result by the std-controlled CBF, shown in Fig. 2(d), contains less noise and more image details than that by the CBF.

4. EXPERIMENTS

We apply our non-blind deconvolution on various real images. Figure 3(a) shows a image from [4] with a kernel of 23×23 pixels. Fergus et al' method[8] estimate a accurate kernel for this image. Using this kernel, we compare against two leading methods: Levin et al's method [2] and the ISISD method [3]. Levin et al's method produce noticeable ringing artifacts. In contrast, our method and the ISISD method successfully suppress the ringing.

Figure 3(f)[1] is a severely blurred image with a kernel of 55×55. Fergus et al's method [8] fails to estimate the blur kernel. Shan et al's method [4] produces a very noisy kernel as shown in Fig. 3(g). Using this kernel, Shan et al's result contains over-smoothed textures and visible artifacts. The ISISD method [3] recovers more details, but also introduces more ringing. Our result looks much more natural.

Figure 3(k)[8] is a very noisy example. The kernel estimated by Fergus et al.[8] is also noisy. Using this kernel, we tune the parameters of Shan et al's method to smooth out most of noise, but find many details are suppressed inevitable. The ISISD method leaves the image noise almost untouched. In contrast, our approach smoothes out most of noise while successfully preserving image details.

Three more blurred images [1][4] and the kernels estimated by Shan et al.[5] are shown in Fig. 3(p). It is obviously that all our results are of high quality.

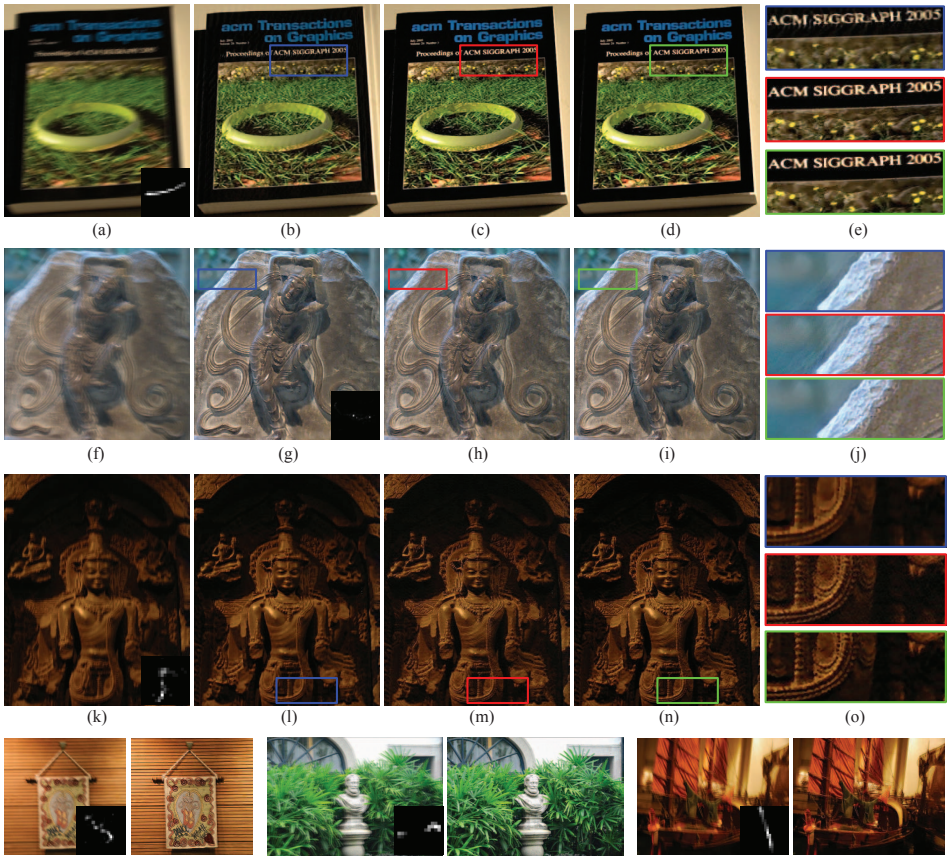

Fig. 3. (a) Real blurred image with small noise from Shan et al.[4] and the small kernel estimated by Fergus et al.[8]. (b) Result by Levin et al.[2]. (c) Result by the ISISD method [3]. (d) Our result. (e) Close-up views of (b)-(d). (f) Real blurred image with large blur kernel from Yuan et al.[1]. (g) Recovered image and kernel by Shan et al.[4]. (h) Result by the ISISD method [3] using the kernel in (g). (i) Our result using the kernel in (g). (j) Close-up views of (g)-(i). (k) Real blurred image with large noise and the estimated kernel from Fergus et al.[8]. (l) Result by Shan et al.[4]. (m) Result by the ISISD method [3]. (n) Our result. (o) Close-up views of (l)-(n). (p) More blurred images [1][4] and our results.

5. CONCLUSION

In this paper, we have presented an inter scale non-blind deconvolution method to remove the motion blur. The high-quality results demonstrate that our approach is very robust to kernel estimate errors and image noise.

6. REFERENCES

[1] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," in *SIGGRAPH*. ACM, 2007.

[2] A. Levin, R. Fergus, F. Durand, and B. Freeman, "Image and depth from a conventional camera with a coded aperture," in *SIGGRAPH*. ACM, 2007.

[3] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Progressive inter-scale and intra-scale non-blind image deconvolution," in *SIGGRAPH*. ACM.

[4] Q. Shan, J.-Y. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," in *SIGGRAPH*. ACM, 2008.

[5] Y. Weiss and W. T. Freeman, "what makes a good model of natural images?," in *CVPR*. IEEE, 2007.

[6] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," in *SIGGRAPH*. ACM, 2003.

[7] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," in *SIGGRAPH*. ACM, 2004.

[8] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *SIGGRAPH*. ACM, 2006.