Learning High-Order Filters for Efficient Blind Deconvolution of Document Photographs

Lei Xiao, Jue Wang, Wolfgang Heidrich, Michael Hirsch

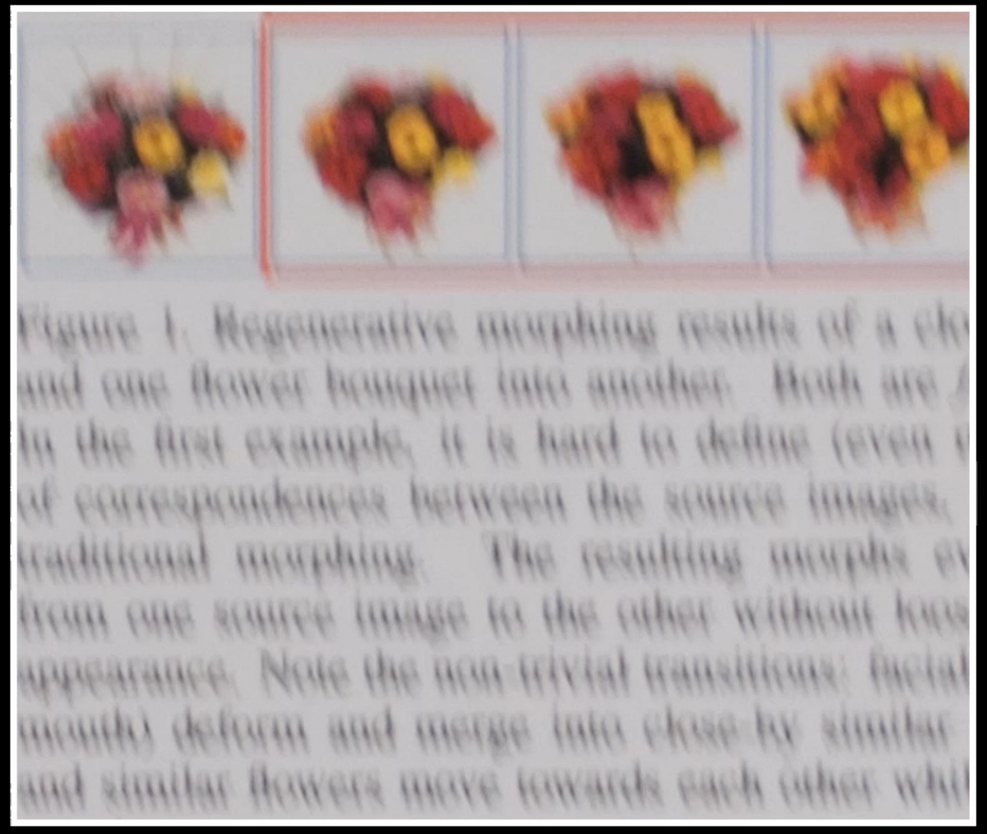
UNIVERSITY OF BRITISH COLUMBIA
ADOBE RESEARCH
KAUST
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Increasingly common to take photos of text documents

popularity of mobile cameras

printed articles, receipts, newspapers, books, etc

But fragile to blur caused by camera shake at hand-held shots



patch from an example input

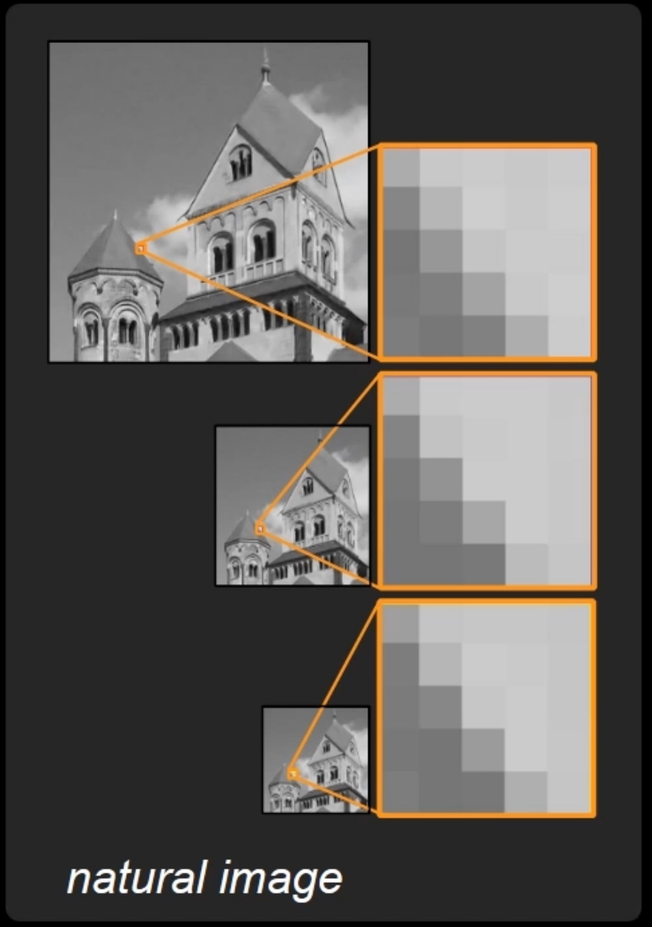
Blind Deconvolution

?

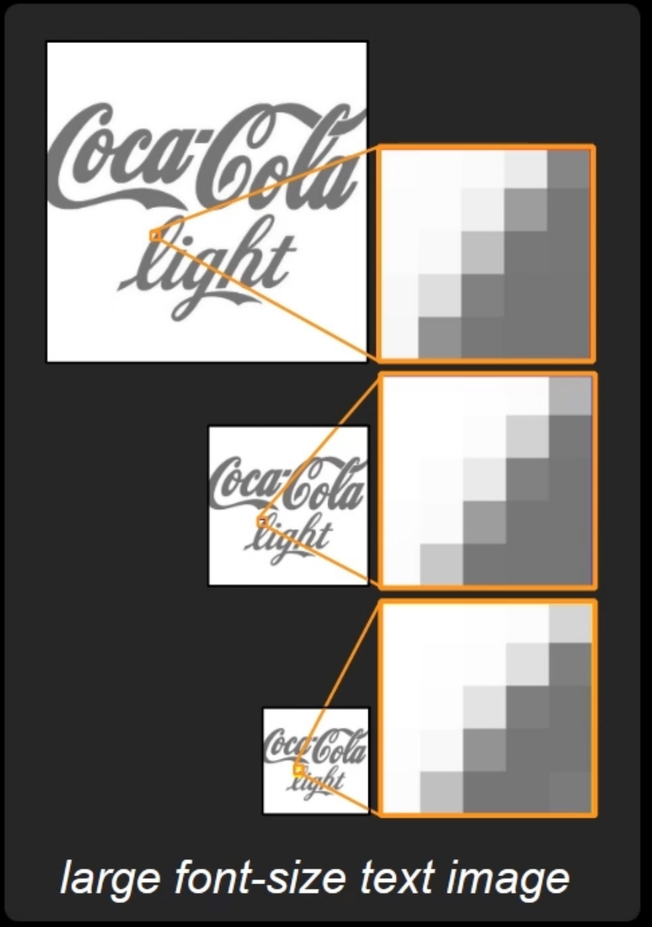
sharp and ***legible*** image

Sparse gradient priors do not work well for common text documents (typically w/ small fonts)

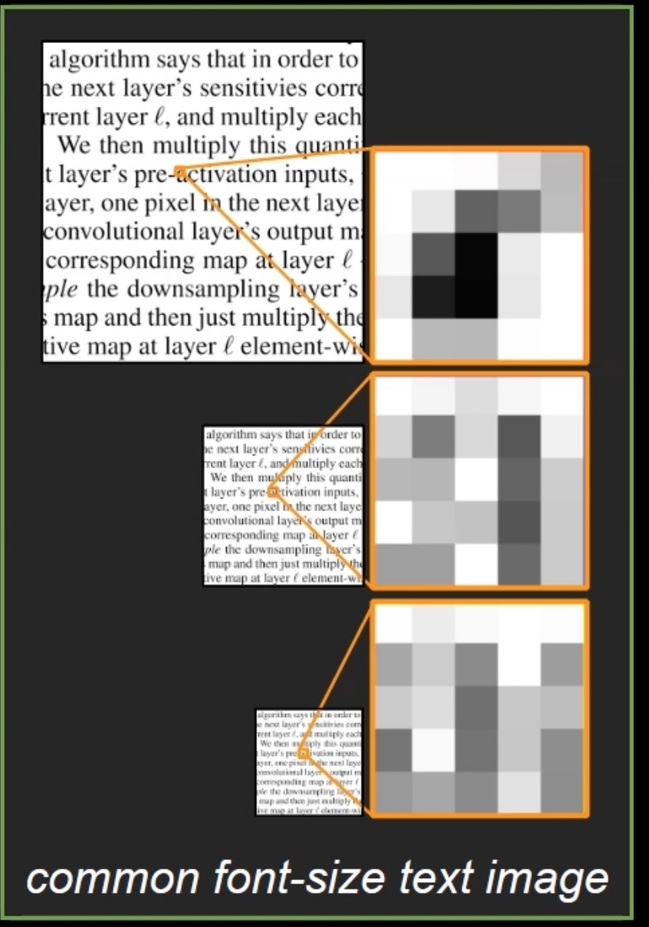
- Patch comparison at varying scales



natural image



large font-size text image

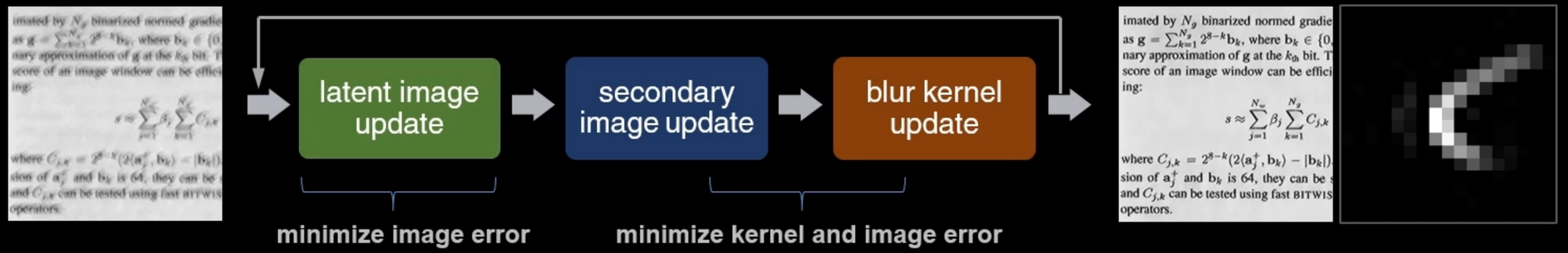


common font-size text image

high-order structures

Our solution: high-order filter priors learned via discriminative learning

- Multi-scale, interleaved shrinkage-fields network:

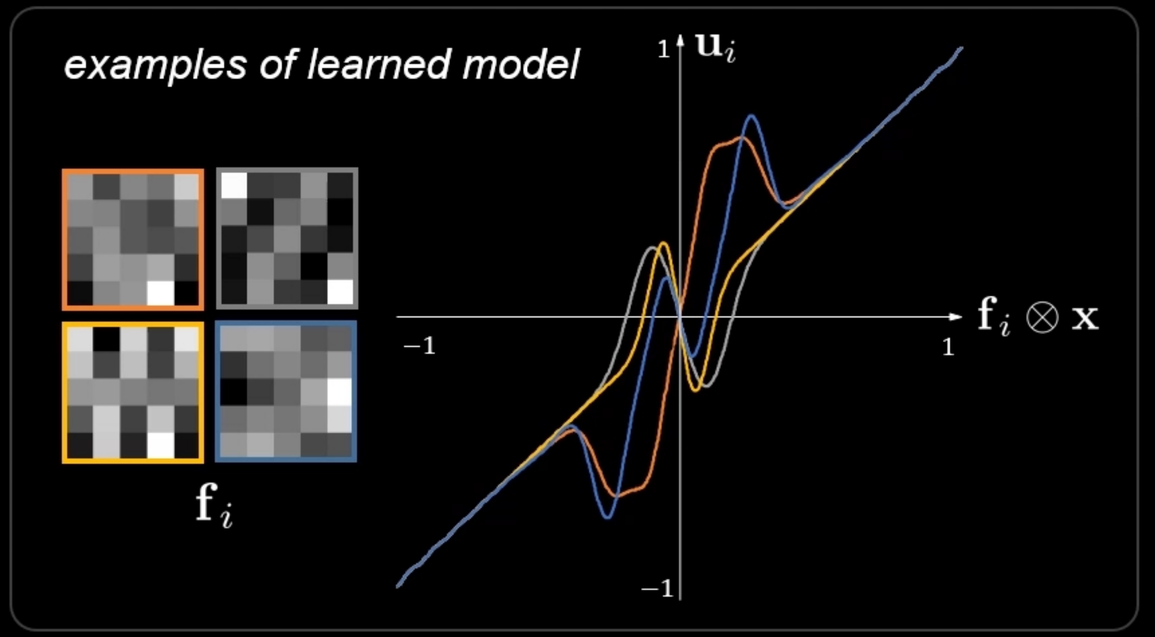


- Image update step:

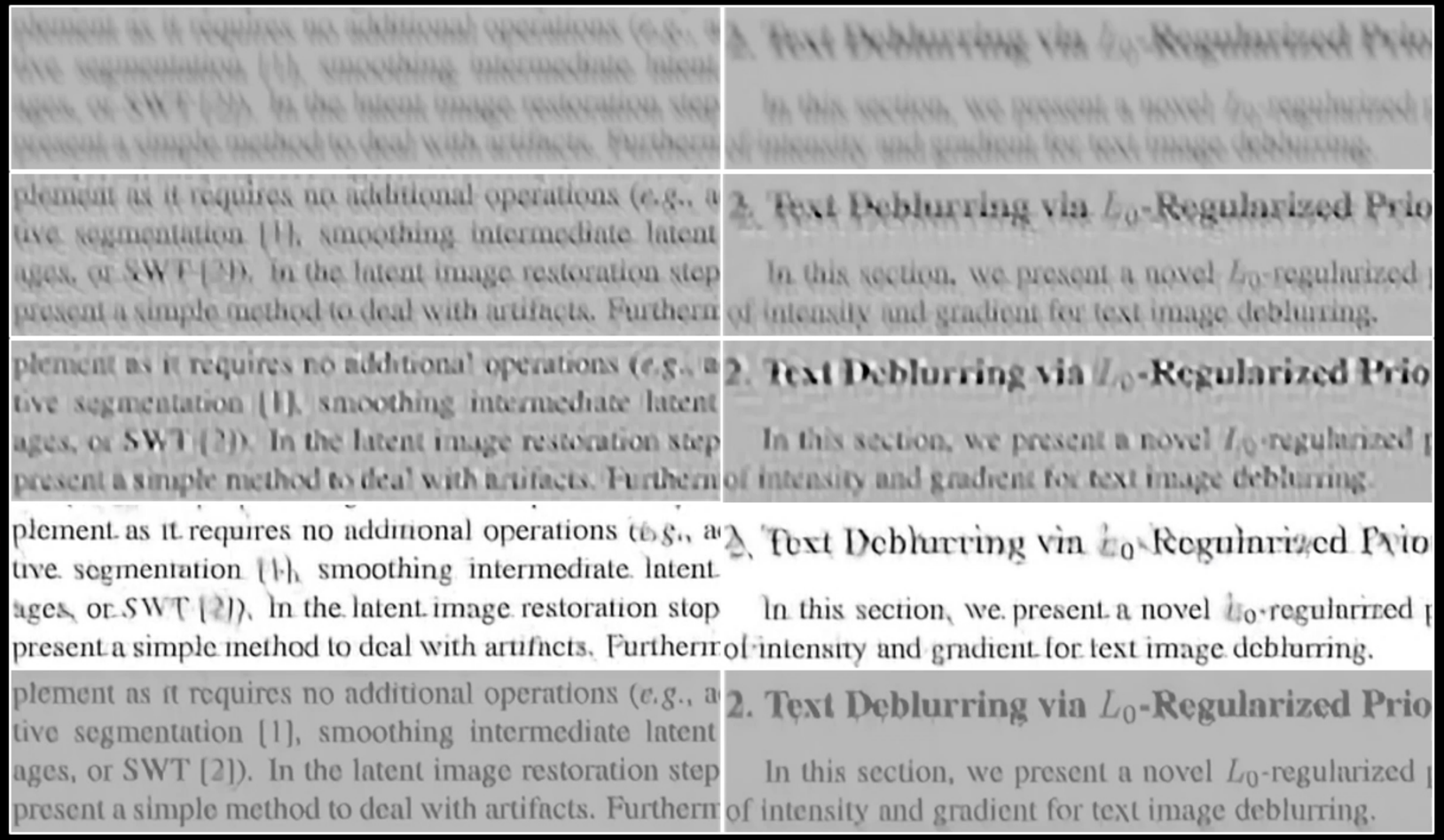
$$\mathbf{x} = \underset{\mathbf{x}}{\operatorname{argmin}} ||\mathbf{b} - \mathbf{k} \otimes \mathbf{x}||_2^2 + \sum_{i=1}^{N} \alpha ||\mathbf{u}_i - \mathbf{f}_i \otimes \mathbf{x}||_2^2$$

$$\mathbf{u}_i = \psi_i(\mathbf{f}_i \otimes \mathbf{x})$$

[Shrinkage-Field, CVPR14]

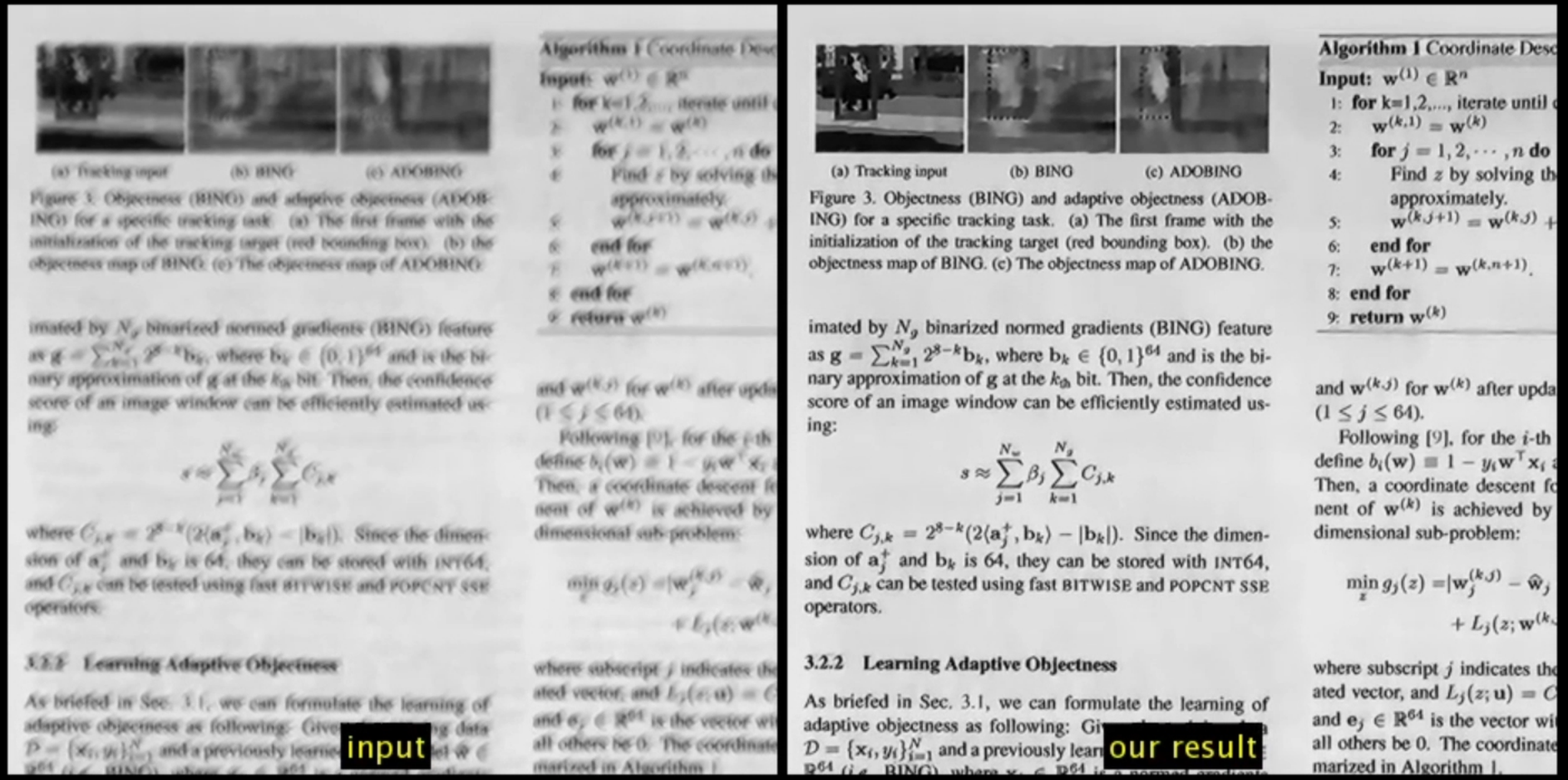


Results on real-world examples



example patches cropped from result images

Results on real-world examples

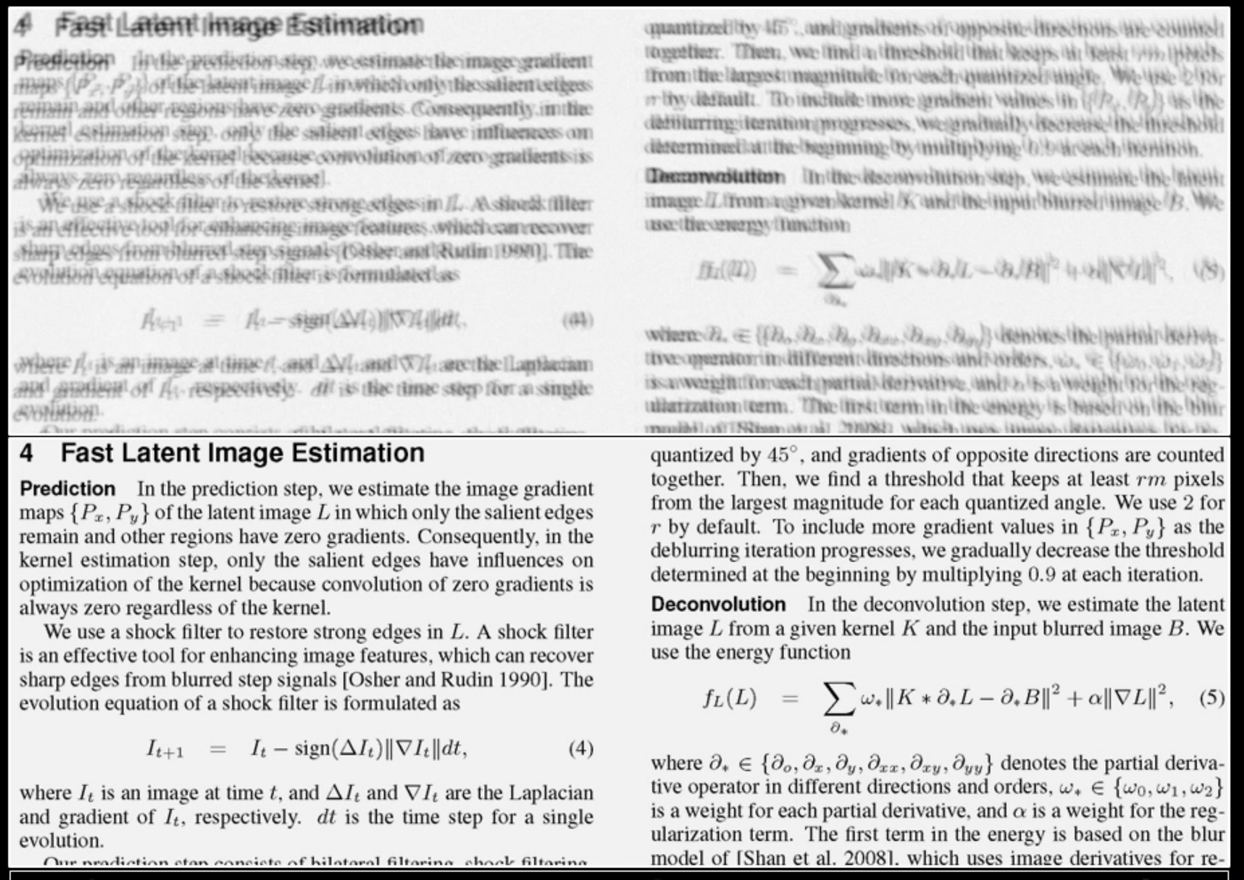


pan and zoom animation

Results on spatially-varying blur

blurry input

our result image



4 Fast Latent Image Estimation

Prediction In the prediction step, we estimate the image gradient maps $\{P_x, P_y\}$ of the latent image L in which only the salient edges remain and other regions have zero gradients. Consequently, in the kernel estimation step, only the salient edges have influences on optimization of the kernel because convolution of zero gradients is always zero regardless of the kernel.

We use a shock filter to restore strong edges in L. A shock filter is an effective tool for enhancing image features, which can recover sharp edges from blurred step signals [Osher and Rudin 1990]. The evolution equation of a shock filter is formulated as

$$I_{t+1} = I_t - \text{sign}(\Delta I_t)\|\nabla I_t\|dt, \quad (4)$$

where I_t is an image at time t, and ΔI_t and ∇I_t are the Laplacian and gradient of I_t, respectively. dt is the time step for a single evolution.

Our prediction step consists of bilateral filtering, shock filtering

quantized by 45°, and gradients of opposite directions are counted together. Then, we find a threshold that keeps at least rm pixels from the largest magnitude for each quantized angle. We use 2 for r by default. To include more gradient values in $\{P_x, P_y\}$ as the deblurring iteration progresses, we gradually decrease the threshold determined at the beginning by multiplying 0.9 at each iteration.

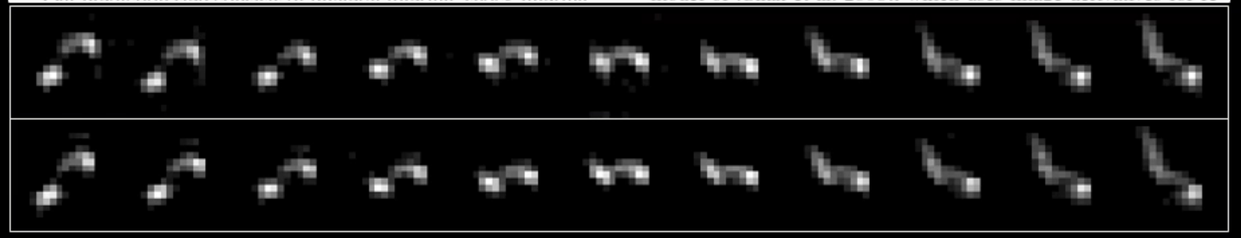
Deconvolution In the deconvolution step, we estimate the latent image L from a given kernel K and the input blurred image B. We use the energy function

$$f_L(L) = \sum_{\partial_*} \omega_* \|K * \partial_* L - \partial_* B\|^2 + \alpha\|\nabla L\|^2, \quad (5)$$

where $\partial_* \in \{\partial_o, \partial_x, \partial_y, \partial_{xx}, \partial_{xy}, \partial_{yy}\}$ denotes the partial derivative operator in different directions and orders, $\omega_* \in \{\omega_0, \omega_1, \omega_2\}$ is a weight for each partial derivative, and α is a weight for the regularization term. The first term in the energy is based on the blur model of [Shan et al. 2008], which uses image derivatives for re-

our blur estimate

true blur



Results on documents containing color figures

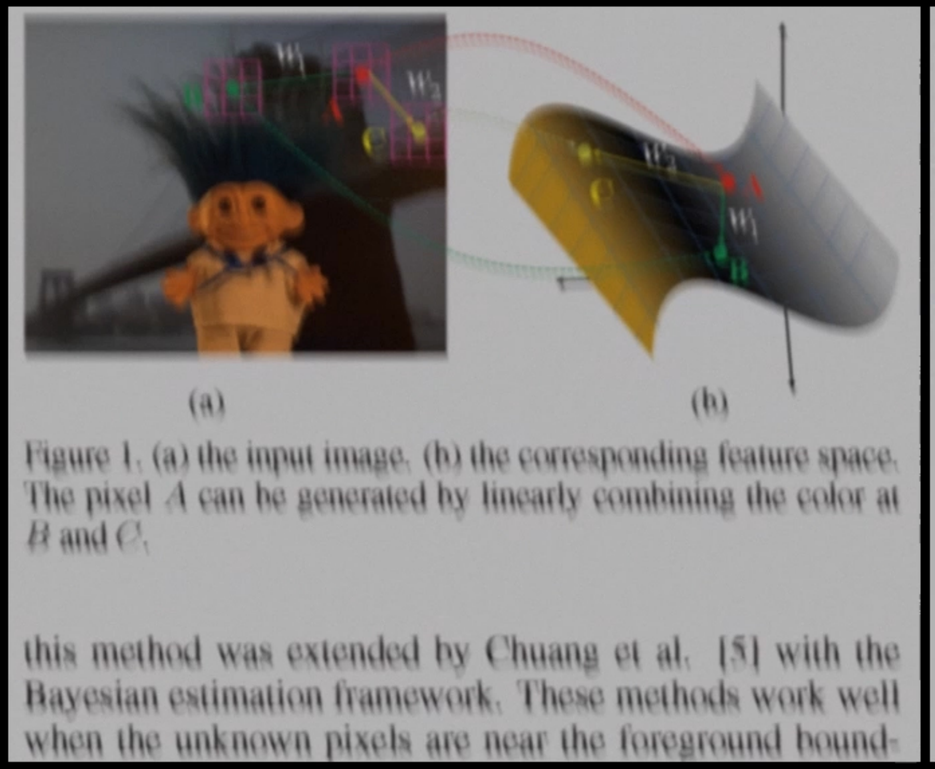


Figure 1. (a) the input image. (b) the corresponding feature space. The pixel A can be generated by linearly combining the color at B and C.

this method was extended by Chuang et al. [5] with the Bayesian estimation framework. These methods work well when the unknown pixels are near the foreground bound-

blurry input

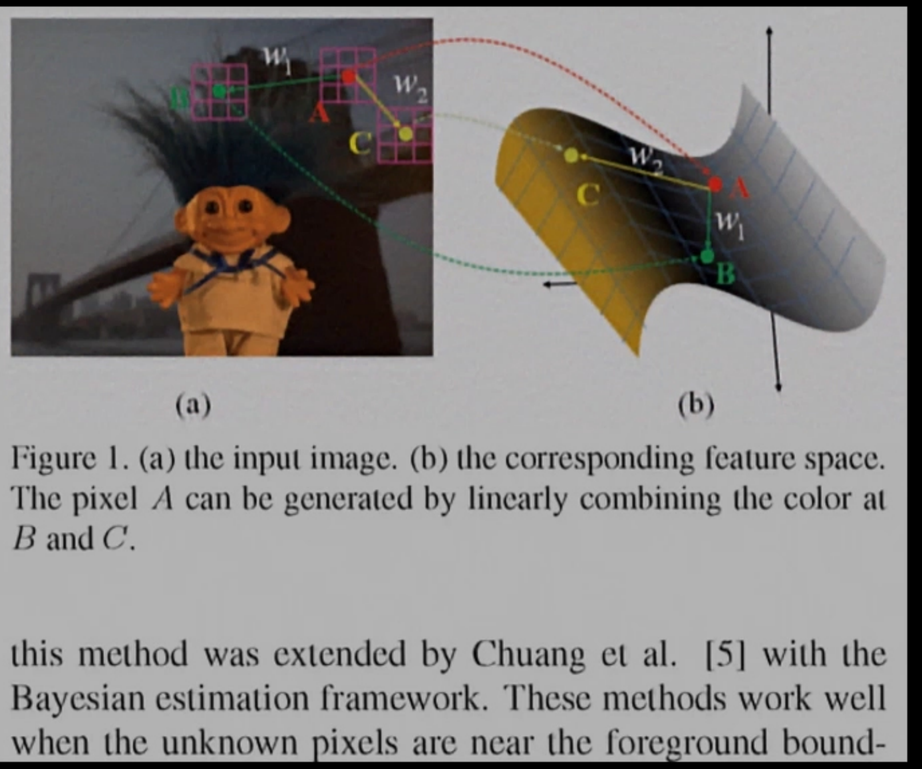


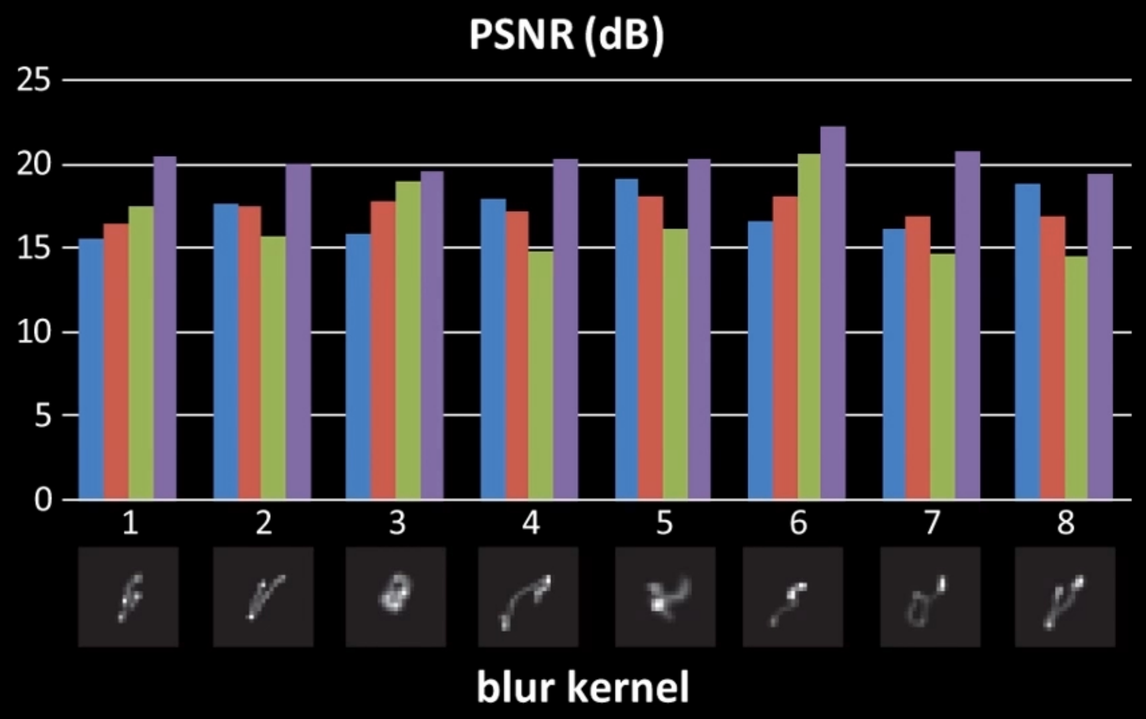
Figure 1. (a) the input image. (b) the corresponding feature space. The pixel A can be generated by linearly combining the color at B and C.

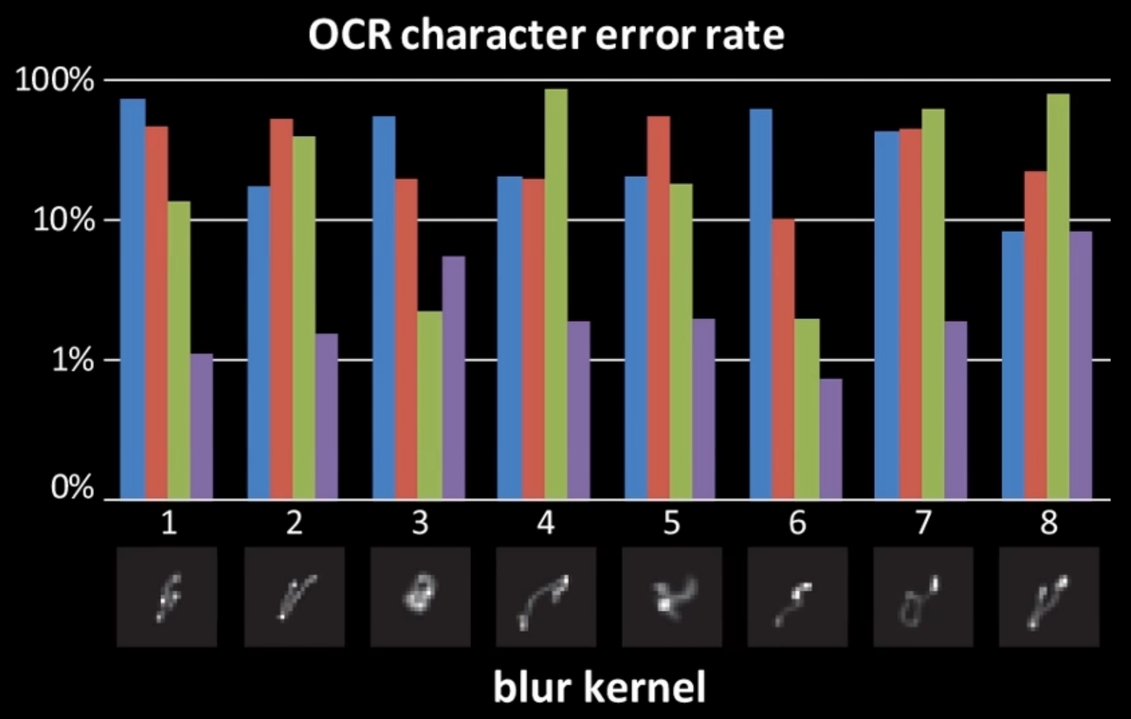
this method was extended by Chuang et al. [5] with the Bayesian estimation framework. These methods work well when the unknown pixels are near the foreground bound-

our result

Approach: predict blur kernel of figure region from surrounding text regions.

Quantitative evaluations





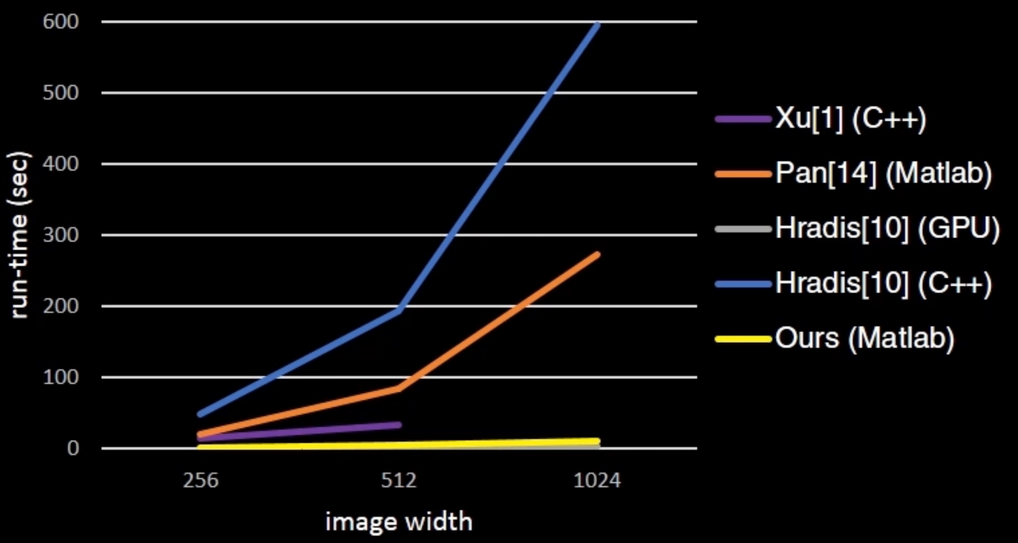
Xu Pan Hradis Ours

Note the logarithmic scale markings in OCR results

High computational efficiency

Table 1: Run-time comparison (in seconds)

Image size	256^2	512^2	1024^2
Xu[1] (C++)	14.8	33.4	-
Pan[14] (Matlab)	19.6	84.3	271.9
Hradiš[10] (C++)	48.5	193.7	594.9
Hradiš[10] (GPU)	0.3	1.0	3.1
Ours (Matlab)	2.0	3.9	11.4



Required operations: 2D FFT, 2D convolution, 1D Look-Up-Table.

Learning High-Order Filters for Efficient Blind Deconvolution of Document Photographs

Lei Xiao, Jue Wang, Wolfgang Heidrich, Michael Hirsch

UNIVERSITY OF BRITISH COLUMBIA
ADOBE RESEARCH
KAUST
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

Poster Session S-2B-08