

Invited Paper

Detail Recovery for Single-image Defocus Blur^{*1}YU-WING TAI,^{†1} HUIXUAN TANG,^{†2} MICHAEL S. BROWN^{†1}
and STEPHEN LIN^{†3}

We presented an invited talk at the MIRU-IUW workshop on correcting photometric distortions in photographs. In this paper, we describe our work on addressing one form of this distortion, namely defocus blur. Defocus blur can lead to the loss of fine-scale scene detail, and we address the problem of recovering it. Our approach targets a single-image solution that capitalizes on redundant scene information by restoring image patches that have greater defocus blur using similar, more focused patches as exemplars. The major challenge in this approach is to produce a spatially coherent and natural result given the rather limited exemplar data present in a single image. To address this problem, we introduce a novel correction algorithm that maximizes the use of available image information and employs additional prior constraints. Unique to our approach is an exemplar-based deblurring strategy that simultaneously considers candidate patches from both sharper image regions as well as deconvolved patches from blurred regions. This not only allows more of the image to contribute to the recovery process but inherently combines synthesis and deconvolution into a single procedure. In addition, we use a top-down strategy where the pool of in-focus exemplars is progressively expanded as increasing levels of defocus are corrected. After detail recovery, regularization based on sparsity and contour continuity constraints is applied to produce a more plausible and natural result. Our method compares favorably to related techniques such as defocus inpainting and deconvolution with constraints from natural image statistics alone.

1. Introduction

Image blur due to defocus is a common feature in photographs. Although this blur may sometimes be desirable for certain visual effects, the consequent loss of local appearance detail can be detrimental to computer vision applications and

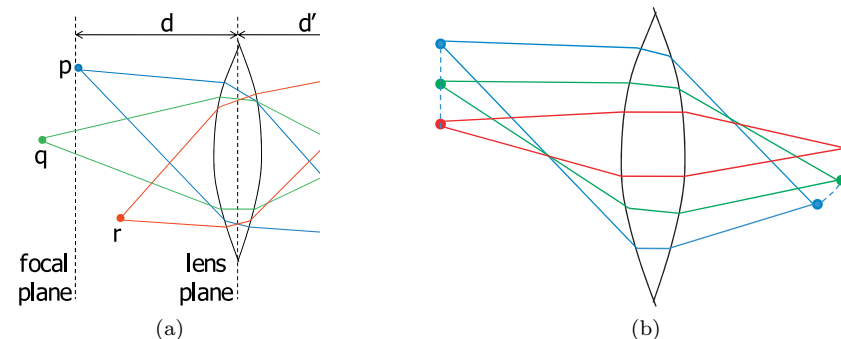


Fig. 1 Optical geometry for different causes of defocus blur. (a) Defocus due to limited depth of field. (b) Defocus due to lens aberrations, shown here for field curvature.

unwanted by viewers. To improve image quality, reduction of defocus blur is often desirable.

Two principal causes of image defocus are the camera's limited depth of field (DOF) and lens aberrations that cause light rays to converge incorrectly onto the imaging sensor. Defocus of the first type is illustrated in **Fig. 1** (a) and is described by the thin lens law²⁾:

$$\frac{1}{d} + \frac{1}{d'} = \frac{1}{f}$$

where f is the focal length of the lens, d is the distance of the focal plane in the scene from the lens plane, and d' is the distance from the lens plane to the sensor plane. For given values of f and d' in an optical system, radiance from scene points p on the corresponding focal plane will be correctly focused onto the sensor. For points that lie in front of or behind the focal plane (such as r and q respectively), their light rays will converge behind or in front of the sensor, and hence will appear blurred in the image. The observed blur can be modeled as a convolution of the focused image with a point spread function, which can be assumed to be Gaussian^{8),14)}.

Lens aberrations arise from physical limitations of a real lens to form exact images of a scene. Several different types of lens aberrations may occur within an optical system. One example is field curvature, which results from the property that a curved lens will focus light onto a curved imaging plane, as shown in

^{†1} National University of Singapore

^{†2} Fudan University

^{†3} Microsoft Research Asia

^{*1} This work was done when Yu-Wing Tai and Huixuan Tang were interns at Microsoft Research Asia

Fig. 1(b). Typically, a camera system is designed such that defocus from this and other lens aberration effects are minimized toward the center of the image and increases radially. The overall effects of defocus-based lens aberrations can also be modeled as a Gaussian blur²⁰⁾ that tends to increase with radial distance from the image center.

For limited DOF and lens aberrations, the defocus blur can thus be formulated as the following convolution:

$$I_{ob} = I \star h + n,$$

where I_{ob} is the observed image, I represents an in-focus image of the scene, h is a spatially-variant Gaussian blur kernel, and n denotes additive noise.

Because of this effect on images, the defocus problem has often been addressed using blind deconvolution approaches that attempt to recover the in-focus image I and the underlying blur kernel h simultaneously. However, even when the blur kernel is known, deconvolution is well known to be ill-posed with numerous possible solutions that yield the same defocused image. Many of these solutions can appear rather unnatural. For example, the result of the theoretically optimal Wiener filter²⁵⁾ often exhibits ringing on the image boundaries and can corrupt existing fine details. Blind deconvolution becomes additionally challenging in the typical case of spatially-varying defocus, for which location-specific blur kernels must be determined in order to restore the image.

In this paper, we present a method that takes advantage of the spatially-varying defocus in an image. Our method seeks local image areas with similar image content but different defocus levels. Among image patches with similar content, those with less defocus contain greater appearance detail and are used as exemplars for deblurring corresponding patches that are more defocused. This approach, which is also used in Ref. 7), is similar to image hallucination except that exemplar information must be gleaned from only the image itself, rather than from an image database^{1),10),18),22),24)}.

The main difficulty in this approach is that a single image contains rather limited exemplar information for deblurring. As a result, an ideal exemplar may not be present in the image for a given defocused patch. Moreover, for a substantially defocused patch, there exists significant ambiguity as to what its ideal exemplar should be. In Ref. 7), the most in-focus patch with the closest

correspondence independent of blur is taken as the exemplar. But while such an exemplar may provide the best solution locally, use of such exemplars may not lead to a globally coherent solution for the deblurred image, resulting in texture seams and blocking effects as shown in Fig. 5(b).

In this paper, we introduce a novel correction algorithm that addresses these practical issues of exemplar-based single-image deblurring. To maximize the use of limited image information, our method employs a flexible and progressive scheme for exemplar identification. In this scheme, patches of various orientations and sizes are considered to facilitate the search for exemplars and to broaden the pool of candidate exemplars for a given defocused patch. Also, the set of possible exemplar patches is progressively expanded by adding patches deblurred by our algorithm as it proceeds. The exemplar set is further expanded by including deconvolved patches from uncorrected blurred regions in addition to in-focus patches from the original image. In this way, exemplar-based synthesis and deconvolution are combined into a common deblurring framework.

Our method also reduces incoherence in deblurred solutions, which is often caused by a lack of suitable exemplars. Instead of utilizing the exemplar that yields the best local solution, we select exemplars using Markov chain based inference that allows some local accuracy to be exchanged for a more globally coherent result. Even with Markov chain based inference, the synthesized solution may nevertheless contain some noticeable artifacts such as jagged image features, as illustrated in Fig. 4(c), because suitable exemplars do not exist in the image. We reduce this problem with a postprocessing step that aims to improve image quality by enforcing two priors. One is the *contour continuity prior*²⁶⁾, which utilizes anisotropic diffusion to increase smoothness along image contours. The other is the *natural image statistics prior*, which has previously been shown to be useful for deblurring^{15),16)}, and can also sharpen the diffusion of the contour continuity prior.

With this technique, we obtain recovery solutions that not only are consistent with both the observed blurred image areas and the sharper contextual information that exists in the image, but also exhibit global coherence even with the limited exemplar data that is available in a single image. This approach is validated in our experiments with comparisons to related techniques.

2. Related Work

Detail recovery is closely related to several areas, including image deconvolution, image hallucination, and texture synthesis.

Defocus blur resulting from a global convolution procedure can be optimally solved by Wiener filtering²⁵⁾. Deconvolution, however, is ill-conditioned as more than one solution is possible. As a result, a regularization term is typically added to constrain the solution. The Total Variation (TV) regularizer¹⁹⁾, which minimizes the magnitude of the gradient image, has been often used. While TV-based methods generally work well on artificial images, they often over-smooth the interiors of regions and produce unnatural edges. Recently, foreground information was proposed as a regularizer to solve for defocused areas in an image⁷⁾. This approach yields more appealing solutions than the TV regularizer, but artifacts such as visible seams between patches are also introduced due to matching problems.

Natural image statistics have also been used as a prior for regularization in blind deconvolution. In Ref. 9), image statistics are fit to a learned prior in removing the effects of camera shake from photographs. An image specific prior was used in Ref. 15) to segment spatially-variant motion blur. The blur is assumed to result from movement of constant velocity, and the prior is learned from derivatives orthogonal to the direction of movement. More recently¹⁶⁾ demonstrated that a sparsity constraint provides plausible solutions in deconvolving depth blur for natural images. Our approach also leverages this sparsity constraint but applies this to a combined synthesized and deconvolved result. In addition, we incorporate a Contour Continuity prior in the regularization procedure.

Another approach for recovering fine-scale details is by image hallucination, based on a reconstruction constraint between low-resolution and high-resolution images, as well as a prior on the high-resolution image. The reconstruction constraint may be learned from training pairs of low resolution and high-resolution images depicting either a specific class of objects^{1),18)} or natural images¹⁰⁾. Additionally, the reconstructed high-resolution patches are often constrained to be similar to their original low-resolution versions after smoothing and downsampling. In Ref. 24), a group of linear transformations between high-resolution and

low-resolution training pairs is learned, and used in conjunction with a generic image prior for exemplar selection. Our work uses an approach similar to image hallucination, with the significant difference that our training data is restricted to a single image. How to deal with this limited information is the principal issue in our method.

A third approach is to replace defocused image areas by stitching together patches taken from sharper image examples. Texture synthesis methods based on patch-based sampling (e.g., Ref. 5), 6), 13), 17)) aim for seamless results that are visually consistent with the sample data. User specified constraints may also be incorporated to guide the synthesis process (e.g., Ref. 21)). These techniques, however, do not utilize image information from within the synthesis area, and therefore will generate image data that generally does not match the actual scene. An exception to this is Ref. 7), which also addresses spatially varying defocus blur. In their work, sharper image patches that closely match defocused regions are found and used as a regularizer to recover fine-scale image details.

3. Overview

Given an image that contains both focused and defocused regions, our goal is to use the information available in the focused areas to recover the details of the defocused areas in order to produce a sharp and focused image. To make this problem tractable, we assume the focused and defocused areas contain similar content. Let \mathbf{F} denote the focused areas of an image, \mathbf{f} be the simulated defocus image of \mathbf{F} (i.e., $\mathbf{f} = \mathbf{F} \star h$), \mathbf{d} be the defocused areas of an image and \mathbf{D} be the deblurred image of \mathbf{d} . Image patches within \mathbf{F} , \mathbf{f} , \mathbf{D} and \mathbf{d} are denoted by $\tilde{\mathbf{F}}$, $\tilde{\mathbf{f}}$, $\tilde{\mathbf{D}}$ and $\tilde{\mathbf{d}}$ respectively.

The problem then can be formulated into the following Bayesian optimization framework:

$$\mathbf{D}^* = \arg \max_{\mathbf{D}} P(\mathbf{D}|\mathbf{d}, \mathbf{f}, \mathbf{F}) = \arg \max_{\mathbf{D}} P(\mathbf{F}|\mathbf{d}, \mathbf{f}, \mathbf{D})P(\mathbf{D}) \quad (1)$$

where $P(\mathbf{F}|\mathbf{d}, \mathbf{f}, \mathbf{D})$ represents the likelihood probability of a deblurred result, which can be maximized by choosing $\tilde{\mathbf{D}}^* = \tilde{\mathbf{F}}^*$ such that the distance $dist(\tilde{\mathbf{d}}^*, \tilde{\mathbf{f}}^*)$ between two image patches is minimized, and $P(\mathbf{D})$ is the prior probability that encodes prior knowledge about \mathbf{D} . Previous approaches^{1),10),18),22),24)} solve the

above equation by using a large database of primitives from which an optimal \mathbf{D}^* can be found by searching the nearest neighbor in \mathbf{F} with minimum $\text{dist}(\tilde{\mathbf{d}}, \tilde{\mathbf{f}}^*)$. To preserve consistency among neighboring patches,^{10),18)} defined a Markov network with $P(\mathbf{D}) = \prod_{\tilde{\mathbf{D}}_j \in \mathcal{N}(\tilde{\mathbf{D}}_i)} P(\tilde{\mathbf{D}}_i, \tilde{\mathbf{D}}_j)$ as the compatibility matrix of neighboring patches, which measures the distance $\text{dist}(\tilde{\mathbf{D}}_i, \tilde{\mathbf{D}}_j)$ in the overlapping area of $\tilde{\mathbf{D}}_i$ and $\tilde{\mathbf{D}}_j$. The optimal solution \mathbf{D}^* with this neighbor compatibility measure can be solved by using belief propagation²²⁾ or graph cuts¹²⁾. To further refine the recovered details as demonstrated in Ref. 22), back projection¹¹⁾ can be applied after the process to minimize the reconstruction error, with \mathbf{D}^* as the starting point of the back projection algorithm.

Our major challenge is that we do not have a database of in-focus examples but instead must work with the limited information available in the input image itself. When the number of available patches in \mathbf{F} is significantly limited, the optimal solutions \mathbf{D}^* found by using previous approaches will be unsatisfactory. Essentially the solution space provided by the limited \mathbf{F} is too small.

To maximize the use of available exemplar data in \mathbf{F} , we use a flexible matching scheme that considers exemplars of variable size and orientation. We also progressively expand \mathbf{F} using a multiscale strategy that sequentially processes image regions at increasing levels of defocus. Within each scale, \mathbf{D} is defined according to the defocus level, and its recovery results are then added to \mathbf{F} . In this way, \mathbf{F} increases level-by-level by gradually introducing recovery results, beginning from less defocused areas whose results are expected to be more accurate.

Maximizing the use of image information leads to improvement in the recovery result, but there may nevertheless exist some artifacts due to the limited data. To reduce this problem, we regularize the result using the natural image statistics prior and the contour continuity prior. These two priors are enforced after obtaining the solution \mathbf{D}^* to refine the recovered details.

Our overall procedure can then be summarized as follows: 1) determine the defocus scale of image regions; 2) for the current defocus scale, apply the detail synthesis approach using both in-focus and deconvolved image patches; 3) apply regularization using the natural image statistics and the contour continuity priors; 4) proceed to the next scale level (i.e., back to step 2). We describe each of these

algorithm components in the following sections.

4. Defocus Scale Identification

Our input image must first be segmented into different layers according to defocus level. To achieve this purpose, we propose a simple method based on the following observation. If we blur the defocused image using a Gaussian filter of standard deviation σ and then subtract the original image from it, areas less defocused than σ will produce relatively large differences, while areas with greater defocus than σ produce small differences. This is because fine image details should be present only in the less defocused areas. Hence, we can effectively bi-partition the image into areas with defocus scale smaller than σ and those with defocus scale larger than σ . In order to identify pixels that belong to σ , we can apply the bi-partitioning twice at scales $\sigma - \Delta\sigma$ and $\sigma + \Delta\sigma$. The bi-partitioning can be applied several times with different σ in order to estimate multiple defocus layers.

This proposed method can estimate the defocus scale at each pixel location. However, we expect the defocus scale map to be smooth and defocus scale discontinuities to align with image edges. This problem can be formulated as an energy minimization problem over a Markov Network with the data term and the pairwise energy term as follows¹⁶⁾:

$$\begin{aligned} E(\bar{\sigma}_i) &= \begin{cases} 0 & \bar{\sigma}_i = \sigma_i \\ 1 & \bar{\sigma}_i \neq \sigma_i \end{cases}, \\ E(\bar{\sigma}_i, \bar{\sigma}_j) &= \begin{cases} 0 & \bar{\sigma}_i = \bar{\sigma}_j \\ \exp(-|I_i - I_j|^2 / \phi^2) & \bar{\sigma}_i \neq \bar{\sigma}_j \end{cases} \end{aligned} \quad (2)$$

where $\bar{\sigma}_i$ is the state label we want to find to optimize this energy, the factor ϕ is set to 0.1 in our implementation, and I_i and I_j are the image intensities at locations i and j respectively. We use graph cuts¹²⁾ to find the optimal solution of Eq. (2). **Figure 2** shows a defocus estimation result using our proposed method.

5. Detail Recovery through Exemplar-based Synthesis

The task now is to transfer the details from in-focus areas \mathbf{F} of an image to the

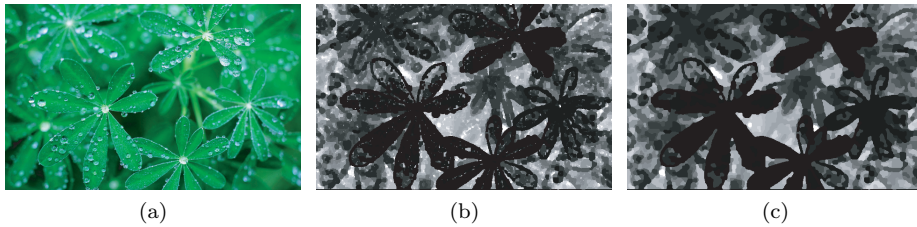


Fig. 2 Defocus scale estimation. (a) Input image. (b) Raw defocus scale map. (c) Refined defocus scale map using graph cuts. Darker areas indicate smaller defocus scales.

defocused regions \mathbf{d} . This detail transfer is performed in a multiscale process. At each scale, we separate pixels into three classes: those that are in focus (\mathbf{F}), those that are defocused at the current scale level (\mathbf{d}) and irrelevant regions. The irrelevant regions include areas defocused at a different scale, and they are ignored at the current level of processing.

In this section, we will describe how to obtain exemplars from \mathbf{F} and \mathbf{d} for synthesis, and then present the Markov chain based inference for optimal patch selection with consideration of neighborhood compatibility.

5.1 Exemplar Search

We generate exemplar patch pairs from \mathbf{F} and \mathbf{d} through convolution of \mathbf{F} and deconvolution of \mathbf{d} using a Gaussian filter that approximates the defocus kernel. Since patches generated from deconvolutions of \mathbf{d} result in a zero match cost, we add an offset match cost (i.e., penalty cost) to these deconvolution patches as described in Section 3. This penalty increases quadratically with blur level. With this offset, our method generally uses more deconvolution patches in deblurring areas with slight defocus, and more in-focus exemplars for areas with greater defocus. The intuition for this penalty cost is that when the defocus scale is small, deconvolution tends to produce better results than synthesis with very few ringing artifacts. When the defocus scale is large, deconvolution produces results with significant ringing, so in-focus patches are preferred as exemplars.

For even greater use of the limited image data, we combine deconvolution with the exemplar-based synthesis by supplementing \mathbf{F} with deconvolved patches from $\tilde{\mathbf{d}}$, denoted as $\hat{\mathbf{d}}$. This means that at each defocus level in the multiscale scheme, we have both the in-focus patches, $\tilde{\mathbf{f}}$, as well as deconvolved imagery $\hat{\mathbf{d}}$. However,

to encourage the use of in-focus exemplars and discourage the use of deconvolved patches from highly defocused areas, we add a penalty $f(\sigma)$ to the distance cost $dist(\mathbf{d}, \hat{\mathbf{d}})$ for deconvolved imagery, where σ is the defocus level described above in previous section. $f(\sigma)$ is designed to increase quadratically with the level of defocus σ . With this penalty, deconvolution patches will be selected only when there exists no suitable in-focus exemplar. For images that contain no useful exemplar data, our method thus becomes equivalent to deconvolution.

To speed up the search process, we cluster the exemplar patches based on its local mean and local contrast. We also facilitate the exemplar search by computing the orientation of each local patch using a set of orientation filter banks and then aligning patch orientations prior to clustering. For a given defocused patch, the K best candidate exemplars are found by first finding the clusters with the closest local mean and local contrast and then performing a comparison between border pixels in the patch and those of its neighbors, as done in patch-based texture synthesis techniques. These K candidates will then be evaluated using Markov chain based inference to determine the final exemplar. Note that the candidate patches can be obtained from both in-focus and deconvolved exemplars.

5.2 Markov Chain Based Inference

Markov chain based inference is used to ensure consistency among neighboring patches, especially for high frequency primitives such as contours. Recall that in the previous section, we have selected the K exemplar candidates at each defocused patch location. The size of the candidates is set to be slightly larger than the area to be synthesized, such that there exists overlap between neighboring patches. We define a neighbor compatibility matrix based on a non-parametric comparison within the overlap area. As described in Eq. (1), the optimal label assignment of patches can be computed by solving a Markov Network with belief propagation, where the data cost is the non-parametric distance between $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{f}}$ and the pairwise energy term is the non-parametric distance between $\tilde{\mathbf{D}}_i$ and $\tilde{\mathbf{D}}_j$ within the overlap area. For further details on Markov chain based inference, readers are referred to Ref. 22).

6. Regularization

In this section, we describe the two additional priors used to refine the recovered

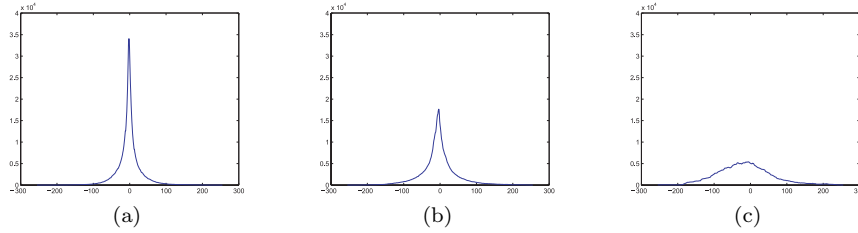


Fig. 3 The gradient distribution of (a) natural images, (b) blurred images, and (c) severely blurred images.

image details.

6.1 Natural Image Statistics Prior

Recent research on natural image statistics has shown that, although real-world scenes vary greatly in content, the distribution of spatial gradients follows a distribution with most of its mass on small values and with long tails, as shown in **Fig. 3** (a). Since this model predicts a natural image to mostly contain small or zero gradients and few large gradients, the natural image statistics prior is sometimes referred to as the *sparsity prior*.

This prior has been shown to be useful in the restoration of blurred/defocused images^{9),15),16)}. For a blurred image, its gradient distribution deviates from that expected from natural image statistics, as shown in Fig. 3 (b). This deviation becomes more pronounced for more severely blurred images, as seen in Fig. 3 (c). By employing the natural image statistics prior, we require the gradient distribution of the solution image to follow the natural image statistics distribution.

We use the Laplacian distribution¹⁶⁾ to approximate the natural image statistics distribution:

$$P(I) \propto \prod \exp(-|\nabla I|^\alpha) \quad (3)$$

where α is an exponential coefficient for which $0 < \alpha < 1$, and ∇I denotes first derivatives of an image I . The sparseness energy defined on I can be written as

$$E_s(I) \propto -w \sum_k \log(P_k(I)) \quad (4)$$

where k is the number of filters used for calculating the first derivative response of I .

The natural image statistics prior, however, is difficult to enforce in the synthesis step, because natural image statistics describes a primarily global property while exemplar selection is a local decision. We thus enforce this prior after synthesis, as done in Ref. 9). Taking the reconstruction error $E(\mathbf{d}, \mathbf{D}) = \exp(-|\mathbf{d} - C_f \mathbf{D}^*|^2)$ into account, the optimal solution that minimizes reconstruction errors with the Laplacian prior can be found by solving a sparse set of linear equations $\mathbf{A}\mathbf{D} = \mathbf{b}$:

$$\mathbf{A} = C_f^T C_f + w \sum_k C_{g_k}^T C_{g_k}, \quad \mathbf{b} = C_f^T \mathbf{d} \quad (5)$$

where C_f denotes the defocus convolution matrix, and C_{g_k} is a set of filters in matrix form used for calculating the first derivative response of \mathbf{D} . Note that \mathbf{d} and the \mathbf{D} are written in vector form. We use iterative re-weighted least squares (IRLS)¹⁶⁾ to obtain the optimal solution of Eq. (5). In the IRLS process, the exemplar-based synthesis result is used as the starting point. Similar approaches were used in Refs. 4), 22), 23), but with back projection applied after synthesis of high frequency details to minimize reconstruction errors. These approaches do not employ regularization on the image gradient distribution. From the in-focus areas \mathbf{F} of an image, the parameter α of the Laplacian distribution can be estimated by fitting the Laplacian distribution to the gradient distribution of \mathbf{F} .

6.2 Contour Continuity Prior

Due to the limited data in \mathbf{F} , contours may not be well recovered, even with the Markov chain based inference algorithm. This is because the ideal exemplars that produce smooth contours do not exist in \mathbf{F} . This problem is significant in our single image approach and we propose to use the contour continuity prior to address it.

The contour continuity prior was first proposed in Ref. 26) for refinement of optical flows, and was later used in Ref. 3) as a constraint for blur kernel estimation. The contour continuity prior is defined by the anisotropic diffusion tensor:

$$\mathbf{T} = \frac{\nabla I_\perp \nabla I_\perp^T}{|\nabla I|} \quad (6)$$

where ∇I_\perp is the vector perpendicular to the local gradient direction ∇I . The

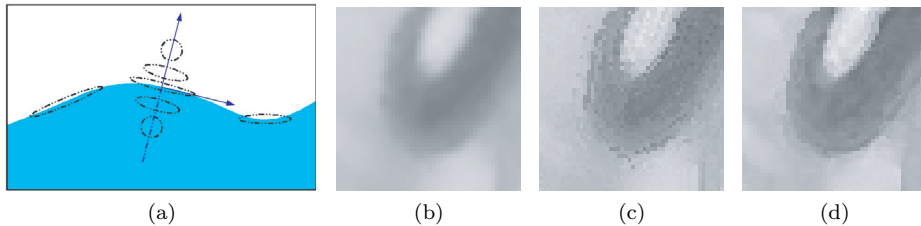


Fig. 4 The contour continuity prior defined by the anisotropic diffusion tensor. (a) The shape and size of Gaussian kernels vary according to the local image structure. (b) The observed image with defocus. (c) Recovered image using exemplar-based synthesis with neighborhood compatibility. (d) Recovered image after applying the contour connectivity prior, which effectively reduces jittering artifacts that arise from insufficient exemplar data.

energy that regularizes contour continuity is thus defined as

$$E_c(k) = \int_{\Omega} \nabla I^T T(\nabla I) \nabla I d\Omega \quad (7)$$

which integrates the per pixel energy over the entire image domain Ω . **Figure 4** (a) illustrates the variation in structure of the anisotropic diffusion tensor. It is an elongated Gaussian along contours and is isotropic in smooth regions.

To preserve discontinuities and edge sharpness, we implement the contour continuity prior in an IRLS process with a local bilateral Gaussian convolution at each iteration. Figure 4 (c) and (d) show a comparison before and after enforcement of this prior.

Note that the contour continuity prior is fundamentally different from the Markov chain based inference described in Section 5.2. The Markov chain based inference finds exemplar patches in a manner that favors contour connectivity, while the contour continuity prior will actually alter the content of \mathbf{D}^* based on local structure to produce smooth contours. Use of the Markov chain based inference is beneficial for this prior, as it yields more coherent synthesis results that are more easily refined using the prior.

In applying this prior, a bilateral anisotropic Gaussian of larger scale would allow more significant refinement of images, but at the same time would lead to greater loss of detail from diffusion. This is a basic tradeoff with this prior, and we use a small scale of $\sigma = 1$ in our implementation.

7. Results

We conducted experiments on our detail recovery method with a variety of inputs. In **Fig. 5**, we compare our approach with the deconvolution method of Ref. 16) and the closest related work, defocus inpainting⁷⁾. Some areas of the input image are severely defocused due to limited depth of field, and in such ar-

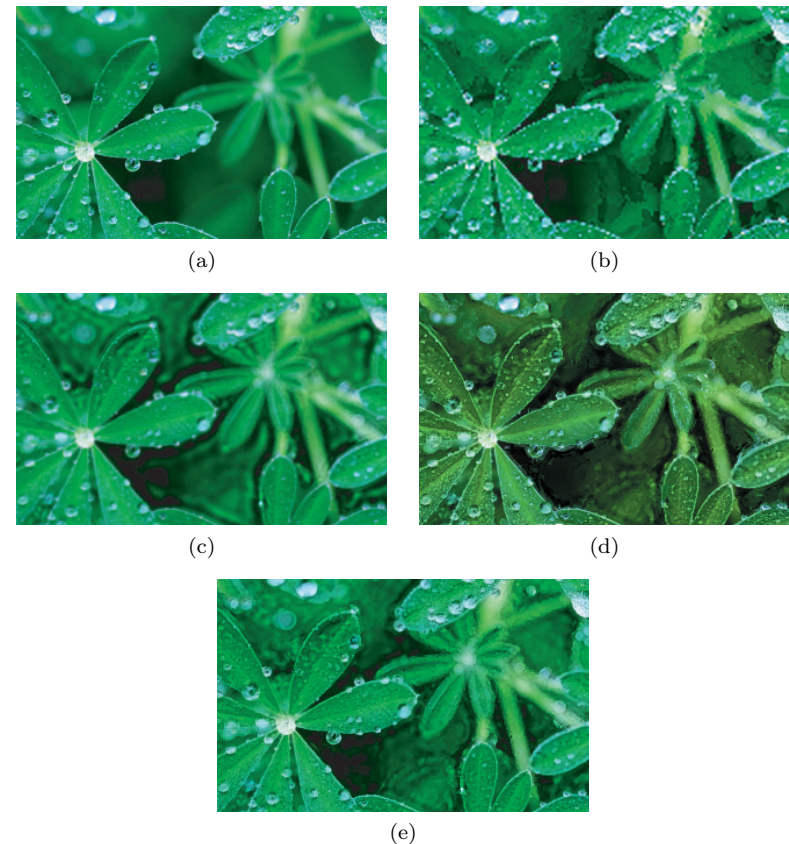


Fig. 5 Comparisons on a leaves image. (a) Input image. (b) Results from defocus inpainting⁷⁾. (c) Results from deconvolution with regularization¹⁶⁾. (d) Our result without regularization. (e) Our result with regularization.



Fig. 6 Chess scene. (a) Input image. (b) Our results.

as deconvolution introduces significant ringing artifacts. In defocus inpainting, limited exemplar data is used, and neighborhood compatibility and contour continuity are not considered. As a result, a relatively small proportion of the image is deblurred, and some broken edges and blocking effects are generated. With the use of expanded exemplar data (but without regularization by priors), our method provides more comprehensive processing of the image and recovers more detail as shown in Fig. 5 (d). By furthermore including regularization by priors as described in Section 6, our algorithm obtains greater spatial coherence as shown in Fig. 5 (e), by suppressing artifacts caused by misalignment of features. The estimated defocus scale map of Fig. 5 is shown in Fig. 2.

In **Fig. 6**, we have a chess scene, which is more challenging than the leaf example because of the disparate image content and the consequently smaller amount of good exemplar data for each type of object. Moreover, the colors of the different objects are similar. Figure 6 (b) shows the detail recovery result by our method. That our approach can successfully recover details in this difficult scenario can partly be attributed to the use of the natural image statistics prior and the contour continuity prior, which help to refine recovered details and to maintain both sharpness and smoothness of edges.

Figure 7 shows an example of flowers. This example demonstrates the effec-



Fig. 7 Flowers example. (a) Input image. (b) Our results.

tiveness of our approach in transferring sharp details to severely blurred edge boundaries. Our method successfully transfers this edge information without introducing ringing artifacts.

8. Discussion

In this work, we proposed a technique for recovering image details that are lost due to defocus blur. The spatial variations in defocus that are commonly present in images are exploited by using more-focused image patches as exemplars in restoring less-focused patches with similar image content. The key issue of this approach is in synthesizing a coherent result from the limited exemplar data in a single image. For this, we have presented algorithm components that take greater advantage of the image data and image priors. With this more comprehensive use of image information, our method obtains more visually plausible results in comparison to related techniques.

Although the proposed method seeks to maximize the use of contextual information, its ability to recover image details is still limited by the available data in the image. Textures or other repeated image content provide a richer context from which to find exemplars. For cases with sparse context, such as Fig. 6, there is a greater reliance on exemplars produced by deconvolution. A consequence of

this is that deconvolution artifacts such as ringing may appear more frequently in the recovery results. Though deconvolution serves as a lower bound on the performance of our method, there often exists enough useful information in a scene to bring appreciable improvements to photographs with defocus blur.

The use of image context may also be limited for local image regions that bear substantial levels of defocus. For such regions, significant ambiguity can exist in the deblurring solution due to a vast space of possible corresponding exemplars. With this weakened constraint on exemplars, an accurate exemplar may not be identifiable from the image context. Our method nevertheless uses spatial coherence to select an exemplar that yields a visually plausible result. In cases where the level of defocus is underestimated, a selected exemplar may contain some amount of defocus blur. With a defocused exemplar, image sharpening can still be obtained but only up to the level of the exemplar. On the other hand, overestimation of defocus can result in an inability to find suitable exemplars. This problem could potentially be mitigated by considering increasingly lower levels of defocus until exemplars are found.

In future work, we plan to investigate efficient methods for accommodating more general geometric transformations in our search for exemplar patches. Because of perspective projection, textures on a surface may exhibit various foreshortening effects which make them more difficult to match to other patches in an image. Since foreshortening appears as affine transformations of patches, we intend to extend our rotation-invariant exemplar search scheme to directly handle affine changes. In addition to expanding the flexibility of the matching algorithm, we also will examine other applications of our technique. For example, we believe that our general framework could potentially be applied to problems such as packet loss in image transmission, in which the degree of image degradation varies spatially over an image. Also, elements of this approach such as the use of the natural image statistics prior and the contour continuity prior may have some utility in superresolution.

References

- 1) Baker, S. and Kanade, T.: Limits on Super-Resolution and How to Break Them, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.9, pp.1167–1183 (2002).
- 2) Born, M. and Wolf, E.: *Principles of Optics*, Cambridge Univ. Press, seventh edition (1999).
- 3) Chen, J. and Tang, C.K.: Robust Dual Motion Deblurring, *CVPR* (2008).
- 4) Dai, S., Han, M., Xu, W., Wu, Y. and Gong, Y.: Soft Edge Smoothness Prior for Alpha Channel Super Resolution, *CVPR* (2007).
- 5) Efros, A.A. and Freeman, W.T.: Image Quilting for Texture Synthesis and Transfer, *Proc. ACM SIGGRAPH*, pp.341–346 (2001).
- 6) Efros, A.A. and Leung, T.K.: Texture Synthesis by Non-Parametric Sampling, *Proc. International Conference on Computer Vision*, pp.1033–1038 (1999).
- 7) Favaro, P. and Grisan, E.: Defocus Inpainting, *Proc. European Conference on Computer Vision*, pp.349–359 (2006).
- 8) Favaro, P. and Soatto, S.: A Geometric Approach to Shape from Defocus, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.27, No.3, pp.406–417 (2005).
- 9) Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T. and Freeman, W.T.: Removing camera shake from a single photograph, *ACM Trans. Graphics*, Vol.25, No.3, pp.787–794 (2006).
- 10) Freeman, W.T., Pasztor, E.C. and Carmichael, O.T.: Learning Low-Level Vision, *International Journal of Computer Vision*, Vol.40, pp.25–47 (2000).
- 11) Irani, M. and Peleg, S.: Improving Resolution by Image Registration, *CVGIP*, Vol.53, No.3, pp.231–239 (1991).
- 12) Kolmogorov, V. and Zabih, R.: What Energy Functions Can Be Minimized via Graph Cuts?, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.26, No.2, pp.147–159 (2004).
- 13) Komodakis, N. and Tziritas, G.: Image Completion Using Global Optimization, *Proc. Computer Vision and Pattern Recognition*, pp.442–452 (2006).
- 14) Kubota, A. and Aizawa, K.: Reconstructing Arbitrarily Focused Images From Two Differently Focused Images Using Linear Filters, *IEEE Trans. Image Processing*, Vol.14, No.11, pp.1848–1859 (2005).
- 15) Levin, A.: Blind Motion Deblurring Using Image Statistics, *NIPS* (2006).
- 16) Levin, A., Fergus, R., Durand, F. and Freeman, W.T.: Image and Depth from a Conventional Camera with a Coded Aperture, *ACM Trans. Graphics* (2007).
- 17) Liang, L., Liu, C., Xu, Y.Q., Guo, B. and Shum, H.Y.: Real-time texture synthesis by patch-based sampling, *ACM Trans. Graphics*, Vol.20, No.3, pp.127–150 (2001).
- 18) Liu, C., Shum, H.Y. and Zhang, C.S.: Two-step approach to hallucinating faces: global parametric model and local nonparametric model, *Proc. Computer Vision and Pattern Recognition*, pp.192–198 (2001).
- 19) Rudin, L., Osher, S. and Fatemi, E.: Nonlinear total variation based noise removal algorithms, *Physica D*, Vol.60, pp.259–268 (1992).
- 20) Sakamoto, T.: Model for spherical aberration in a single radial gradient-rod lens, *Applied Optics*, Vol.23, No.11, pp.1707–1710 (1984).

- 21) Sun, J., Yuan, L., Jia, J. and Shum, H.Y.: Image completion with structure propagation, *ACM Trans. Graphics*, Vol.24, No.3, pp.861–868 (2005).
- 22) Sun, J., Zheng, N.N., Tao, H. and Shum, H.Y.: Generic Image Hallucination with Primal Sketch Prior, *Proc. Computer Vision and Pattern Recognition*, pp.729–736 (2003).
- 23) Tai, Y.W., Tong, W.S. and Tang, C.K.: Perceptually-Inspired and Edge-Directed Color Image Super-Resolution, *CVPR* (2006).
- 24) Tappen, M.F., Russell, B.C. and Freeman, W.T.: Exploiting the Sparse Derivative Prior for Super-Resolution and Image Demosaicing, *Third International Workshop on Statistical and Computational Theories of Vision* (2003).
- 25) Wiener, N.: *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, New York (1949).
- 26) Xiao, J., Cheng, H., Sawhney, H., Rao, C. and Isnardi, M.: Bilateral Filtering-based Optical Flow Estimation with Occlusion Detection, *ECCV* (2006).

(Received November 12, 2008)

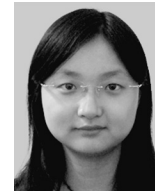
(Accepted January 21, 2009)

(Released March 31, 2009)

(Communicated by Akihiro Sugimoto)



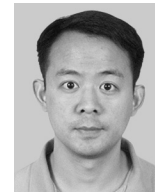
Yu-Wing Tai is a Ph.D. candidate in the Department of Computer Science at the National University of Singapore (NUS). From September 2007 to June 2008, he worked as a full-time student intern at Microsoft Research Asia (MSRA). He was awarded the Microsoft Research Asia Fellowship in 2007. He received a M.Phil and B.Eng (First Class Honors) degree in Computer Science from the Hong Kong University of Science and Technology (HKUST) in 2005 and 2003 respectively. His research interests include computer vision, and image/video processing.



Huixuan Tang received a B.A. degree in 2005 and a M.S. degree in 2008, both from Fudan University, China. She is currently a MSc. student at University of Toronto. Her recent research interest resides in computer vision, especially in computational photography.



Michael S. Brown obtained his B.S. and Ph.D. in Computer Science from the University of Kentucky in 1995 and 2001 respectively. He was a visiting Ph.D. student at the University of North Carolina at Chapel Hill from 1998-2000. He is currently the Sung Kah Kay Assistant Professor in the School of Computing at the National University of Singapore. His research interests include Computer Vision, Image Processing and Computer Graphics. Dr. Brown regularly serves on the program committees of the major Computer Vision conferences (ICCV, CVPR, and ECCV) and has served as an Area Chair for CVPR'09.



Stephen Lin is currently a Lead Researcher in the Internet Graphics Group of Microsoft Research Asia. He obtained a B.S.E. from Princeton University and a Ph.D. from the University of Michigan. His research interests include computer vision and computer graphics. Dr. Lin has served as a Program Chair for the Pacific-Rim Symposium on Image and Video Technology 2009, a General Chair for the IEEE Workshop on Color and Photometric Methods in Computer Vision 2003, and as Area Chairs of the IEEE International Conference on Computer Vision 2007 and 2009.