Mutual-Structure for Joint Filtering

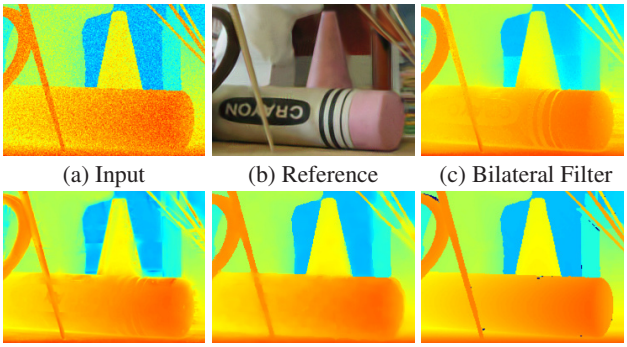
Xiaoyong Shen[†] Chao Zhou[†] Li Xu[‡] Jiaya Jia[†]

[†]The Chinese University of Hong Kong [‡]SenseTime Group Limited

http://www.cse.cuhk.edu.hk/leojia/projects/mutualstructure/

Abstract

Previous joint/guided filters directly transfer the structural information in the reference image to the target one. In this paper, we first analyze its major drawback – that is, there may be completely different edges in the two images. Simply passing all patterns to the target could introduce significant errors. To address this issue, we propose the concept of mutual-structure, which refers to the structural information that is contained in both images and thus can be safely enhanced by joint filtering, and an untraditional objective function that can be efficiently optimized to yield mutual structure. Our method results in necessary and important edge preserving, which greatly benefits depth completion, optical flow estimation, image enhancement, stereo matching, to name a few.



(a) Input (b) Reference (c) Bilateral Filter

(d) Guided Filter (e) Ours (f) Ground Truth

Figure 1. Joint image filtering on a structure-inconsistent image pair. (a) and (b) are the target and reference images respectively. (c) and (d) are the results of bilateral and guided image filtering respectively, which transfer color-image edges to depth. (e) is our result that does not contain the texture patterns and unwanted edges from (b), and (f) is the ground truth.

1. Introduction

Image filters are fundamental tools widely used in image editing [11], denoise [10, 2], optical flow [25, 24], stereo matching [14, 12, 29] and image restoration [18, 27]. Several filters process a single image to either preserve edges [2, 9, 30, 11, 28, 15, 7] or remove texture [31, 26]. Another group of filters, involving joint bilateral filter [2] and guided image filter [11], can take extra images as reference or guidance.

Joint filters are fundamentally helpful in several tasks. For example, in stereo matching, joint bilateral and guided image filters were employed to aggregate the cost volume [29, 12]. For depth refinement and completion, corresponding RGB images were used in joint filtering [17]. The common property is to employ an additional image during the course of target image filtering. The reference image provides structural guidance of how the filter should perform. Thus structural-preserving or removal on the target image can be achieved locally depending on what is contained in the reference image.

Analysis of Joint Filter Joint filter makes a basic assumption on the reference images, i.e., they should contain correct structural information. Otherwise, the guidance could be insufficient, or even wrong locally.

However, many practical tasks with images in RGB/depth [13], flash/no-flash [18], optical flow field/RGB [25], disparity map/RGB [14, 12], RGB/NIR [27], day/night [19] commonly contain inconsistent structures, such as noise, holes, texture, shadow, highlight and multi-spectrum data. They easily cause trouble to the filtering process.

One example is shown in Figure 1, where (a) and (b) are the input and reference images. Because (b) is with extra edges not related to depth and the input image (a) is noisy, joint filter generates unwanted structures as shown in (c) and (d). It is thus necessary in joint filtering to choose correct edges since many of them are not suitable.

Our Mutual-Structure for Joint Filtering We in this paper address the structure inconsistency problem and propose the concept of *mutual-structure* to enhance the capability of joint processing in restoring structure based on common information in target and reference images. The main contribution is the principle not to completely trust the reference

image in affinity definition. Instead, we take possible difference between the reference and target images into account and estimate their mutual structures as a new reference for joint filtering. Our result for Figure 1 is shown in (e), which does not transfer those erroneous reference edges and textures.

This goal is achieved via a new objective function considering the common information between the target and reference images, which we will detail later. This framework is general to handle images with diverse structure or in different spectral configuration. It optimally suppresses information that does not present commonly in input images.

Our method benefits a large group of applications, including depth/RGB image restoration, stereo matching, shadow detection, matching outlier detection, joint segmentation and cross-field image restoration. Our code is publicly available for further employment and evaluation.

2. Background and Motivation

We review joint/guided image filters. The methods can be categorized into local and global ones.

Local Joint Methods Local joint filters are mostly the joint extension of single-image edge-preserving filters. Weighted mean filter includes anisotropic diffusion [5], bilateral filter [2, 8, 15, 3, 28, 30], guided filter [11], and geodesic distance based filter [4, 9]. They define different types of affinities between neighboring pixels considering color difference and spatial distance. The affinity is then set as weights to locally smooth images. Edges can be preserved because large affinities are yielded in low contrast regions while low affinities are set along edges. The joint extension of these weighted mean filters sets affinity weights according to another reference image.

Another line is weighted median [14, 32], which imposes weights for different pixels under an affinity definition when computing medians. A joint weighted median filter can be constructed by computing weights from the reference image. Another general mode filter is presented in [23].

Global Joint Schemes Global methods optimize functions. They include total variation (TV) [20], weighted least squares (WLS) [6], and scale map scheme [27]. These methods restore images by optimizing functions involving all or many pixels and containing regression terms defined in the weighted L_1 or L_2 norm. Similar to local filter, joint global optimization is yielded after calculating the weights based on the reference image.

To preliminarily summarize related work, almost all joint image filters identify important structures based on the reference image. These methods work best when the reference data contain all useful information. Contrary to these approaches that are based on the perfect-reference-structure assumption, our method considers possibly inconsistent edges, noise, texture, shadow and highlight. These issues are common for natural images and captured multi-spectral data. We explain our method in following sections.

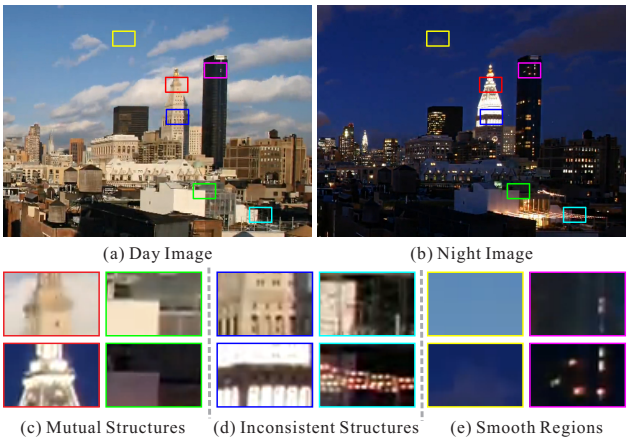


Figure 2. Examples of image structure correlation in a day/night image pair. (a)-(b) Day and night images respectively. (c) Mutual structure patch close-ups. (d) Inconsistent structure patches. (e) Smooth patches. The images are from the time-lapse video of [22].

3. Mutual-Structure for Joint Filtering

Images, even paired and registered, are hardly with the same structures. We roughly categorize the difference into three types using the illustration in Figure 2 where a day/night image pair is presented.

- **Mutual structures** As shown in (c), mutual structures can be intuitively understood as common edges arising in the corresponding two patches. These edges are *not* necessarily with the same magnitude. The gradient direction can also be reversed.

- **Inconsistent structures** Inconsistent structures are different patterns between the two patches. There may be many such structures in an image pair as shown in Figure 2(d). When one edge appears only in one image but not the other, it is regarded as inconsistent.

- **Smooth regions** There are common low-variance smooth patches in images. They are easily influenced by noise and other visual artifacts as shown in (e).

Among these types of joint structures, inconsistent edges generally cause big problems when transforming erroneous patterns to the target image. In this paper, we aim to find the *mutual-structure* in both input images and let it guide the joint filtering process. Accordingly, we not only filter the target image, but as well optimize the reference under our

formulation of mutual-structure based on a structure similarity measure.

We give the definitions that will be used later in this paper. We denote I_0 and G_0 as the target and reference images respectively. The filtering output and updated reference image with mutual structures are denoted as I and G respectively. We also denote by $p = (x, y)^T$ pixel coordinates. $I_{0,p}$, $G_{0,p}$, I_p and G_p are pixel intensities in I_0, G_0, I and G. We process channels separately and use $N(p)$ to denote pixel set in the patch centered at p. The number of pixels in $N(p)$ is $|N|$.

4. Mutual-Structure Formulation

We measure structure similarity between corresponding patches in I and G, and then define corresponding constraints. An objective function to jointly optimize I and G are finally described.

4.1. Structure Similarity

Patch similarity between I and G regarding central pixel p cannot be simply measured by summed gradient difference in the two patches. This problem has been studied for years in many fields. One common and effective measure is the normalized cross correlation (NCC), expressed as

$$\rho(I_p, G_p) = \frac{cov(I_p, G_p)}{\sqrt{\sigma(I_p)\sigma(G_p)}}, \tag{1}$$

where $cov(I_p, G_p)$ is the covariance of patch intensity. $\sigma(I_p)$ and $\sigma(G_p)$ denote the variance. When two patches are with the same edges, even under different magnitudes, $|\rho(I_p, G_p)| = 1$. Otherwise, $|\rho(I_p, G_p)| < 1$. $|\rho(I_p, G_p)|$ is large when the patch structures are similar.

Albeit with ideal properties in measurement, NCC is hard to use directly due to its nonlinearity in our process for structure optimization. To make the problem trackable, we provide the following derivation to establish the relationship between NCC and simple least-square regression.

First, the well known least square regression function $f(I, G, a_p^1, a_p^0)$ of local patches $N(p)$ is expressed as

$$f(I, G, a_p^1, a_p^0) = \sum_{q \in N(p)} (a_p^1 I_q + a_p^0 - G_q)^2, \tag{2}$$

where a_p^1 and a_p^0 are the regression coefficients. This function linearly represent one patch in G by that in I. Then we define $e(I_p, G_p)^2$ as the minimum error with the optimal a_p^1 and a_p^0. It is expressed as

$$e(I_p, G_p)^2 = \min_{a_p^1, a_p^0} \quad \frac{1}{|N|} f(I, G, a_p^1, a_p^0). \tag{3}$$

We prove in the following that $e(I_p, G_p)$ is tightly related to the NCC measure on the same input patches.

Claim 1. *The relation between the mean square error $e(I_p, G_p)$ and NCC measure $\rho(I_p, G_p)$ is*

$$e(I_p, G_p) = \sigma(G_p)(1 - \rho(I_p, G_p)^2), \tag{4}$$

where $\sigma(G_p)$ is the variance of patch centered at p in G.

Proof. In Eq. (3), $e(I_p, G_p)$ reaches the minimum when $a_p^1 = \frac{cov(I_p, G_p)}{\sigma(I_p)}$ and $a_p^0 = \overline{G}_p - a_p^1 \overline{I}_p$, where $\overline{I}_p$ and $\overline{G}_p$ are the mean intensities of patches centered at p on I and G respectively. By simply substituting a_p^1 and a_p^0 into Eq. (3) and arranging it according to Eq. (1), we obtain Eq. (4). □

The claim explains when $|\rho(I_p, G_p)| = 1$, which indicates the two patches only contain mutual structure, $e(I_p, G_p)$ reaches 0. Following the same procedure, we construct

$$e(G_p, I_p)^2 = \min_{b_p^1, b_p^0} \quad \frac{1}{|N|} f(G, I, b_p^1, b_p^0), \tag{5}$$

and also conclude $e(G_p, I_p) = 0$ when $|\rho(I_p, G_p)| = 1$. In this case, we take the I as the guidance image and G is the target, which is unconventional in filter design.

Our Patch Similarity Measure We define our final patch similarity measure as the sum of the two above functions defined symmetrically as

$$\mathcal{S}(I_p, G_p) = e(I_p, G_p)^2 + e(G_p, I_p)^2. \tag{6}$$

According to Eqs. (3) and (4) and considering $\rho(I_p, G_p) = \rho(G_p, I_p)$, this measure boils down to

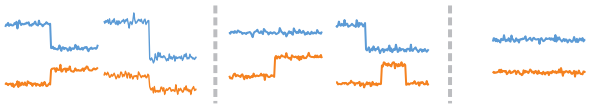
$$\mathcal{S}(I_p, G_p) = (\sigma(I_p)^2 + \sigma(G_p)^2)(1 - \rho(I_p, G_p)^2)^2. \tag{7}$$

We analyze its property in what follows based on the 1D signal example illustrated in Figure 3.

- **Mutual-Structure Patches** When $|\rho(I_p, G_p)|$ approaches 1, $\mathcal{S}(I_p, G_p)$ moves towards 0 in Eq. (7) indicating the two patches are with common edges as shown in Figure 3(a).

- **Inconsistent Structure Patches** As shown in (b), when NCC measure $|\rho(I_p, G_p)|$ outputs a small value for patches containing edges (i.e., at least $\sigma(I_p)$ or $\sigma(G_p)$ is large in Eq. (7)), these edges must be inconsistent. In this case, $\mathcal{S}(I_p, G_p)$ outputs a large value.

- **Smooth Patches** When the patches do not contain significant edges, as shown in (c), $\sigma(I_p)$ and $\sigma(G_p)$ are both small. $\mathcal{S}(I_p, G_p)$ therefore outputs a small value. This special case can also be treated as the mutual-structure patches since they are similarly smooth.



(a) Mutual Structures (b) Inconsistent Structures (c) Smooth Regions

Figure 3. 1D Example. (a) Mutual structure in two patches. (b) Inconsistent structure. (c) Smooth regions.

According to the above analysis, optimizing Eq. (7) to minimize $\mathcal{S}(I_p, G_p)$ can almost achieve our goal in the patch level. We propose image-level optimization to globally search mutual structure.

Final Image Structure Measure Based on the patch-level analysis, we propose the essential image structure similarity term as

$$E_{\mathcal{S}}(I,G,a,b)=\sum_p \big(f(I,G,a_p^1,a_p^0)+f(G,I,b_p^1,b_p^0)\big), \quad (8)$$

which is the sum of patch-level information. a and b are the coefficient sets of $\{a_p^1, a_p^0\}$ and $\{b_p^1, b_p^0\}$ respectively. This term only contains simple least square regression functions, which can be efficiently optimized.

4.2. Other Terms in Global Optimization

We note optimizing only the mutual structure function $E_{\mathcal{S}}(I,G,a,b)$ on I and G may not produce expected results. It is because it can produce the trivial solution where the resulting corresponding patches or the whole images of I and G contain no edge at all. This trivial result is naturally the global optimum of $E_{\mathcal{S}}(I,G,a,b)$. We introduce more constraints to avoid the trivial solution and produce reasonable smoothing effect to remove noise.

The trivial solution can be circumvented by requiring I and G not to wildly deviated from I_0 and G_0 respectively. It thus leads to our image similarity prior function

$$E_d(I,G)=\sum_p \lambda\|G_p-G_{0,p}\|+\beta\|I_p-I_{0,p}\|, \quad (9)$$

where λ and β are two parameters. We apply the l_2-norm distance on intensity due to its fast computation.

Further to introduce reasonable ability to smooth the target image by removing noise and other visual artifacts, we reduce patch intensity variance. In Eq. (6), the two patches in I and G are linearly regressed by each other. Zero variance is yielded when $a_p^1 = 0$ and $b_p^1 = 0$. So the last smoothing term is written as

$$E_r(a,b)=\sum_p(\varepsilon_1 {a_p^1}^2+\varepsilon_2 {b_p^1}^2), \quad (10)$$

where ε_1 and ε_2 control smoothness strength on G and I respectively. Note that this term is related to the ridge regression applied by guided image filter [11]. But our form is different on incorporating two-direction regression errors. It is just one component in our global optimization.

Algorithm 1 Mutual-Structure Estimation

Require: I_0, G_0, N^{iter}, λ, β, ε_1, ε_2
Ensure: I, G
1: Initialize $I^{(0)}$ and $G^{(0)}$
2: **for** t:= 0 **to** N^{iter} **do**
3: Update $a^{(t)}$, $b^{(t)}$ according to Eqs. (12) and (13).
4: Optimize $G^{(t+1)}$ and $I^{(t+1)}$ by Eq. (14).
5: **end for**
6: $I \leftarrow I^{(N^{\text{iter}})}$, $G \leftarrow G^{(N^{\text{iter}})}$

4.3. Final Objective

According to the mutual-structure properties, our final objective function for jointly estimating I and G is the combination of the above three functions:

$$E(I,G,a,b)=E_{\mathcal{S}}(I,G,a,b)+E_d(I,G)+E_r(a,b). \quad (11)$$

a and b are regression coefficient sets, which also need to be optimized. The optimization is a process to get filtering output I and mutual-structure G from I_0 and G_0 after reasonable smoothing.

We use the efficient alternating optimization based on the derivatives and Jacobi method [27] to solve it. We detail our numerical solution below.

5. Numerical Solution

Our alternative updating scheme is sketched in Algorithm 1. The major steps are the following two.

- Given $G^{(t)}$ and $I^{(t)}$, update $a^{(t)}$ and $b^{(t)}$.
- Fix $a^{(t)}$ and $b^{(t)}$, optimize $G^{(t+1)}$ and $I^{(t+1)}$.

t indexes the number of iterations. By decomposing the problem into two sub-ones, each update only needs to solve the quadratic problem in a closed form.

Update $a^{(t)}$ & $b^{(t)}$ Given $I^{(t)}$ and $G^{(t)}$, we update $a^{(t)}$ and $b^{(t)}$ by setting their derivatives to zeros, yielding

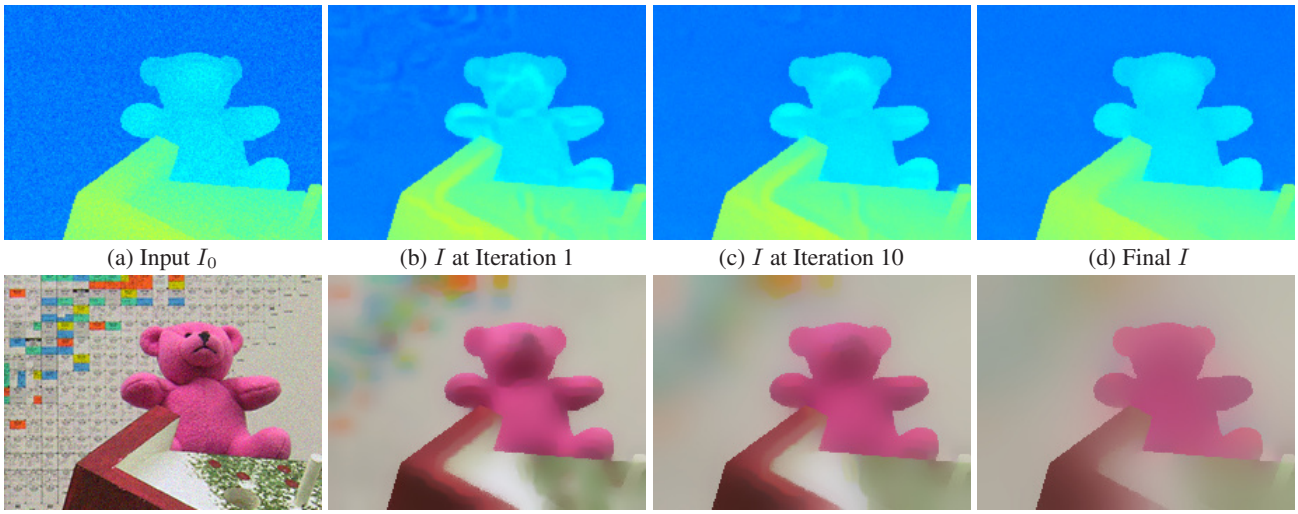
$$a_p^{1(t)}=\frac{cov(I_p^{(t)},G_p^{(t)})}{\sigma(I_p^t)+\varepsilon_1}, a_p^{0(t)}=\mu(G_p^{(t)})-a_p^{1(t)}\mu(I_p^{(t)}), \quad (12)$$

$$b_p^{1(t)}=\frac{cov(G_p^{(t)},I_p^{(t)})}{\sigma(G_p^t)+\varepsilon_2}, b_p^{0(t)}=\mu(I_p^{(t)})-b_p^{1(t)}\mu(G_p^{(t)}), \quad (13)$$

where $\mu(I_p^{(t)})$ and $\mu(G_p^{(t)})$ are the mean intensity of $I^{(t)}$ and $G^{(t)}$ in $N(p)$.

Optimize $G^{(t+1)}$ & $I^{(t+1)}$ With $a^{(t)}$ and $b^{(t)}$, we update $G^{(t+1)}$ and $I^{(t+1)}$ similarly. It yields the linear system as

$$\begin{cases} G_p^{(t+1)}=\frac{1}{M_G^{(t)}}(J_G^{(t)}I_p^{(t+1)}+K_G^{(t)}+\lambda G_{0,p}), \\ I_p^{(t+1)}=\frac{1}{M_I^{(t)}}(J_I^{(t)}G_p^{(t+1)}+K_I^{(t)}+\beta I_{0,p}), \end{cases} \quad (14)$$



(a) Input I_0 (b) I at Iteration 1 (c) I at Iteration 10 (d) Final I

(e) Reference G_0 (f) G at Iteration 1 (g) G at Iteration 10 (h) Final G

Figure 4. Result updated in iterations. Given the noisy natural image in (e) and imperfect depth layer in (a), (b) and (f) show the results of I and G in the first iteration. (c) and (g) are the results after ten iterations. (d) and (h) are the final results after 20 iterations for convergence.

where $M_G^{(t)}$, $J_G^{(t)}$, $K_G^{(t)}$, $M_I^{(t)}$, $J_I^{(t)}$ and $K_I^{(t)}$ are coefficients computed from $a^{(t)}$ and $b^{(t)}$. Among them, $J_G^{(t)}$ and $J_I^{(t)}$ are the coefficients expressed as

$$J_G^{(t)} = \mu(b_p^{1(t)}) + \mu(a_p^{1(t)}), J_I^{(t)} = \mu(b_p^{1(t)}) + \mu(a_p^{1(t)}). \quad (15)$$

$K_G^{(t)}$ and $K_I^{(t)}$ are the constant denoted as

$$\begin{aligned} K_G^{(t)} &= \mu(a_p^{0(t)}) - \mu(b_p^{1(t)} b_p^{0(t)}), \\ K_I^{(t)} &= \mu(b_p^{0(t)}) - \mu(a_p^{1(t)} a_p^{0(t)}). \end{aligned} \quad (16)$$

$M_G^{(t)}$ and $M_I^{(t)}$ are the normalization terms written as

$$\begin{aligned} M_G^{(t)} &= \frac{\lambda}{|N|} + \mu(b_p^{1(t)} b_p^{1(t)}) + 1, \\ M_I^{(t)} &= \frac{\beta}{|N|} + \mu(a_p^{1(t)} a_p^{1(t)}) + 1. \end{aligned} \quad (17)$$

The update stages only contain the simple mean operation and multiplication. They can be implemented efficiently using box filter. We apply the fast box filter based on the integral image implemented in [11].

Algorithm Analysis We first update $a^{(t)}$ and $b^{(t)}$. According to Eqs. (12) and (13), ε_1 and ε_2 make a_p^1 and b_p^1 close to zero for small covariance patches. It introduces the smoothing effect. Update of $G^{(t+1)}$ and $I^{(t+1)}$ by Eq. (14) is under the structure similarity constraint. Similar structures are preserved to minimize the cost.

To demonstrate the iterative updating effects of our algorithm, we show an example in Figure 4 where the input is a captured depth image with much noise and the reference image is the corresponding color image. Inconsistent edges and texture exist. We show the results of our methods in iterations 1 and 10 where inconsistent edges are removed gradually. After convergence in 20 iterations, our results are only with the edges that exist in two images under proper smoothing to remove noise and inconsistency.

Relation with Other Methods Our method is different from other existing filters and from naively applying joint filter in two directions to update the reference and target images in iterations.

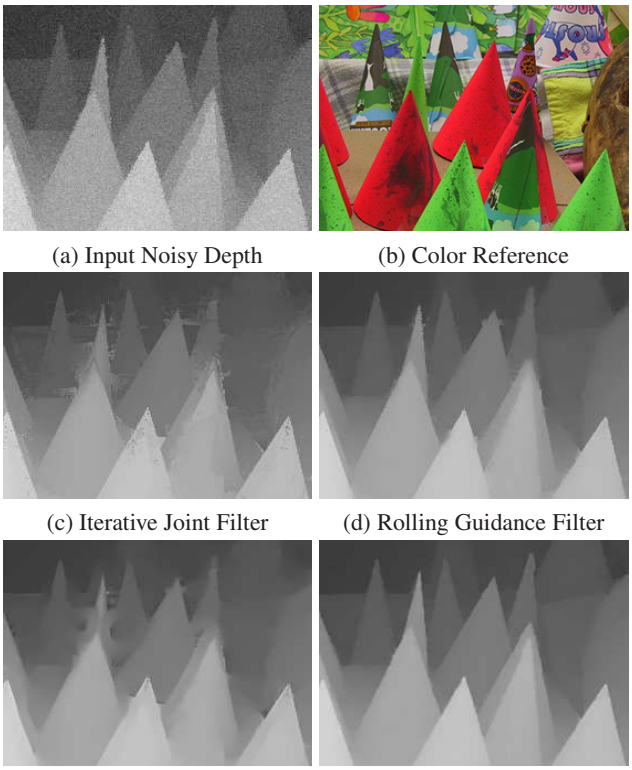
We first compare our solution with iterative joint bilateral filter [16], which iteratively filters the input with the fixed reference image. Although both methods are strong-edge preserving, the iterative joint bilateral filter does not address our aforementioned structure transferring problem from reference to target. We show an example in Figure 5 where (a) and (b) are the input noisy depth and corresponding color image with inconsistent structure. We show the result of iterative joint bilateral filter in (c). Note that other joint filters share similar properties.

We also compare our method with rolling guidance filter (RGF) [32]. We make RGF a joint form on two images by merging channels of the two images into one and employing the high dimensional bilateral filter [10]. As shown in (d), it still cannot get the complete mutual structure.

Another iterative filter to compare is alternatively changing the role of reference and target images and iteratively applying guided image filter. The stages are denoted as

$$I^{(t+1)} = GF(I^{(t)}, G^{(t)}), \quad G^{(t+1)} = GF(G^{(t)}, I^{(t+1)}), \quad (18)$$

where $GF(I^{(t)}, G^{(t)})$ is the guided image filter with input $I^{(t)}$ and guidance image $G^{(t)}$. We set the initialization $I^{(0)}$ and $G^{(0)}$ as I_0 and G_0 respectively. The result is shown in Figure 5(e), which similarly suffers from structure transfer.



(a) Input Noisy Depth (b) Color Reference

(c) Iterative Joint Filter (d) Rolling Guidance Filter

(e) Method of Eq. (18) (f) Ours

Figure 5. Comparison with other iterative joint filters. (a) and (b) show the input and reference images respectively. (c)-(d) are the results of iterative joint bilateral filter and rolling guidance filter. (e) is obtained by alternatively applying guided filter as Eq. (18). These three results all have unwanted structure transferred from the color image to depth. (f) is our result.

Our result shown in (f) does not have this problem because we take both the removal of inconsistent structure and preserving mutual edges into account.

6. Experiments and Applications

Our method takes aligned target and reference images as the input. We employ the dense multi-modal and spectral matching method [21] to align them if there are a level of non-rigid displacement between images.

We extensively evaluate mutual-structure for joint filtering. Our algorithm is easy to implement and the code is publicly available online. The method has parameters λ, β, ε_1, and ε_2. We set λ and β in range $30-300$, which control the deviation to G_0 and I_0 respectively. ε_1 and ε_2 control the smoothness of G and I respectively. We set them around $1E-5$.

In Algorithm 1, we tested different initialization methods for $I^{(0)}$ and $G^{(0)}$ and found that setting $I^{(0)}$ and $G^{(0)}$ as the smoothed version of I_0 and G_0 can achieve fast convergence. Our initial smoothness is by rolling guidance filter for its ability to remove strong noise pattern and textures. Our method takes about 20 iterations to get the final results.

All our experiments are performed on a PC with an Intel Core i7 3.4GHz CPU (one thread used) and 8GB memory. For an image with size 800×600, the running time is 5 seconds with 20 iterations in MATLAB.

6.1. Applications

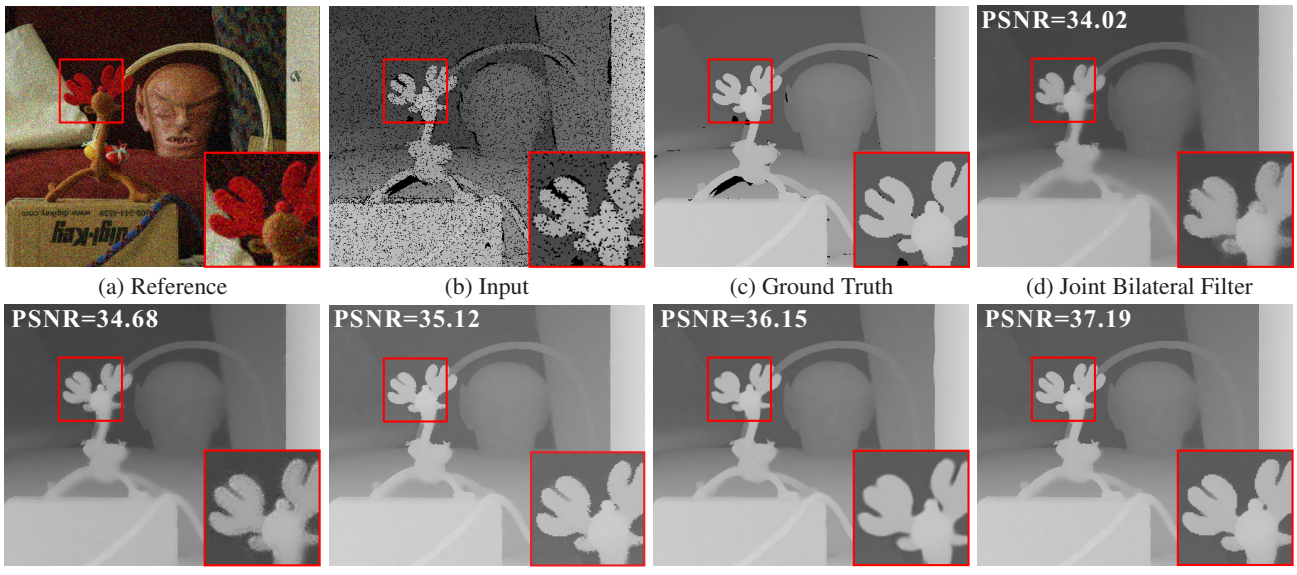
Our mutual-structure for joint filtering benefits several important applications due to the inconsistent structure handling and the high performance. We apply it to RGB/Depth image restoration, stereo matching, RGB/NIR image restoration, joint structure extraction and segmentation, and image matching outlier detection. Our method is generally comparable to or outperforms other filtering schemes due to its unique mutual-structure property.

RGB/depth Restoration Our mutual-structure is suitable for RGB/depth image restoration. While RGB/depth images are captured by depth cameras (e.g. Microsoft Kinect), they always contain inconsistent structures and respective artifacts. Specifically, the RGB image is often with rich details while the depth image is noisy and with holes. Figure 6 shows the comparison of using joint bilateral filter [2], guided image filter [11], weighted median filter [32], the method of [13] and our mutual-structure for joint filtering on these kinds of data. (d-f) are produced by joint bilateral filter, guided image filter, and weighted median filter without mutual-structure computation. The result of [13] in (g) is a bit blurry because of the patch-based scheme. Our method achieves decent results without transferring erroneous structures from the reference as shown in (h). PSNR is calculated for each method.

Moreover, we evaluate our mutual-structure method for RGB/depth restoration on the dataset of [13]. Our method achieves 0.2% higher PSNRs compared with the state-of-the-art solution on average as reported in Figure 7. Moreover, the running time is 50+ times faster because we only need a few quick iterations.

Stereo Matching Considering structure inconsistency between the cost volume and color image, our mutual structure for joint filtering is applicable to stereo matching. We conduct experiments based on the local stereo matching framework provided by Hosni et al. [12]. The framework mainly includes cost volume computation, cost aggregation, disparity computation (winner-take-all) and post processing. Joint image filtering is employed for cost aggregation.

We compare our mutual-structure for joint filtering with other commonly employed filters, such as bilateral filter [2, 29], guided image filter [11, 12], and tree filtering [29] in the cost aggregation step. According to the results shown in Figure 8, our method outperforms other aggregation joint filters.



(a) Reference (b) Input (c) Ground Truth (d) Joint Bilateral Filter

(e) Guided Image Filter (f) Weighted Median Filter (g) Result of [13] (h) Ours

Figure 6. Noisy RGB/Depth image restoration by different methods. (a) and (b) show the input and reference image respectively. (c) is the ground truth depth. (d-h) are the results of different methods. Among them, (g) is released by the authors [13]. PSNRs are reported for all results.

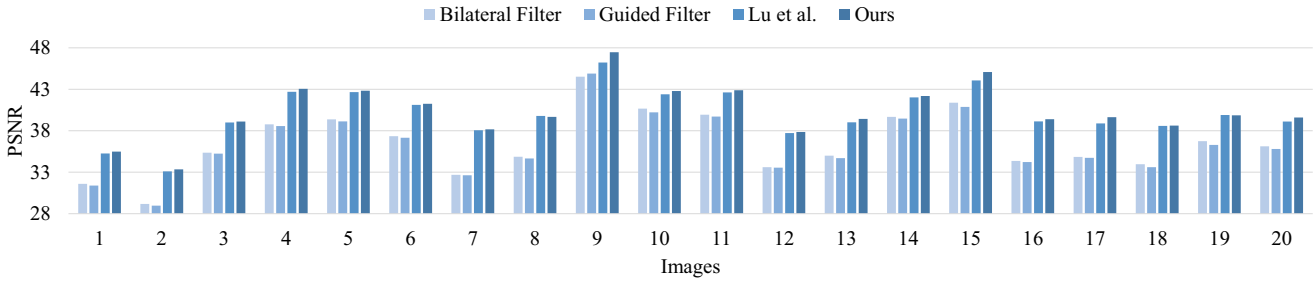


Figure 7. Comparison of different methods for RGB/Depth restoration on the dataset of [13]. We report the PSNRs of the first 20 noisy depth/RGB pairs of the Middlebury dataset using different methods.

Joint Structure Extraction and Segmentation The mutual-structure in our algorithm is actually a solution when the goal is to extract common structures in two images that are from two distinct domains. We conduct experiments on multi-spectral image pairs, which are often with structure inconsistency because of shadow, highlight and moving objects. Two examples are shown in Figure 9.

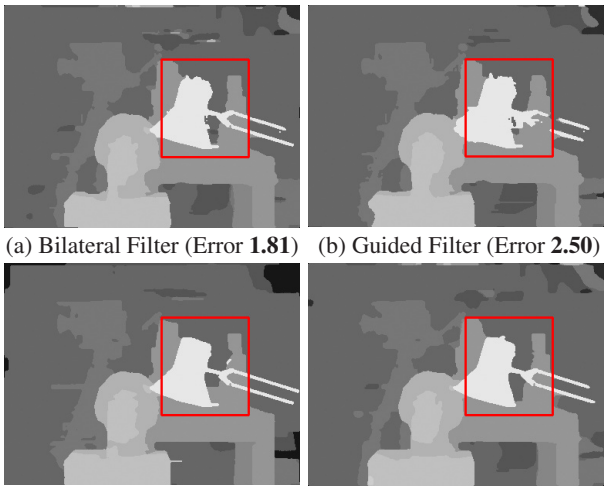
Our mutual-structure also benefits joint segmentation for very complex scenes as shown in Figure 10 where (a) and (b) are the night and day images respectively. (c) is the MCG [1] segmentation result directly on the night image. It does not highlight main components because of the complex content. (d) is the MCG result applied on our mutual-structure, which segments the common objects out in both the day and night images.

Matching Outlier Detection One very challenging problem in image matching is on how to detect matching outliers. We handle this problem by finding common structures, so that the residual between the warped image and the mutual structure forms a good-quality matching-outlier-map. We show an example in Figure 11 to illustrate the effectiveness to find mismatches.

Other Applications Our joint filtering method can also deal with structure transferring in RGB/NIR image restoration. Compared with state-of-the-art method [27], our mutual-structure for joint filtering produces comparable results. The running time is 20 times shorter because of the very efficient iteration steps. More applications, such as joint shadow detection and image enhancement, are provided in our project website.

7. Conclusion and Future Work

We have presented a new scheme for jointly processing images while addressing the common structure inconsistency problem when applying two-image smoothing. It provides new insight on how to avoid transferring unwanted



(a) Bilateral Filter (Error **1.81**) (b) Guided Filter (Error **2.50**)

(c) Tree Filter (Error **1.58**) (d) Ours (Error **1.42**)

Figure 8. Comparison on stereo matching. (a-c) show the results of bilateral filter, guided filter and tree filter. (d) is our result. Pixel errors larger than 1 pixel are reported.

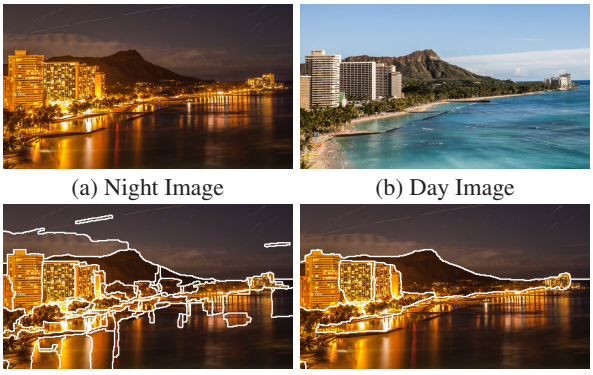
(a) (b) (c) (d)

Figure 9. Joint structure extraction. (a) and (b) are two inputs. (c) is the estimated mutual-structure. (d) shows the common structure of (a) and (b) extracted from the mutual structure (c).

structure from the reference to target images. We have discussed that this type of structure discrepancy commonly arises in almost all image pairs for finding useful information. Our solution stems from maximizing mutual-structure similarity. It leads to an algorithm-level scheme to optimize the mutual-structure. Our future work will be to extend this framework to more tasks in other disciplines where the reference data can be obtained from different sources.

Acknowledgements

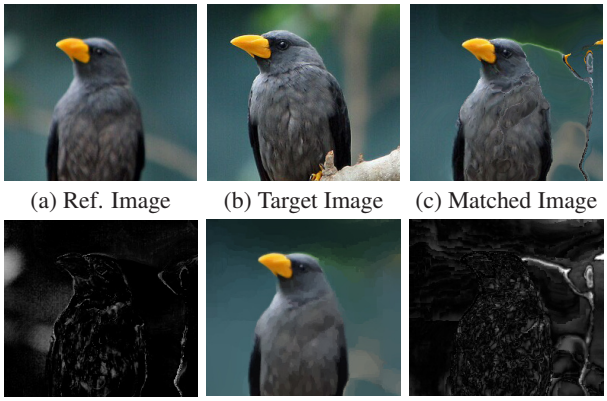
This research is supported by the Research Grant Council of the Hong Kong Special Administrative Region under grant number 412911.



(a) Night Image (b) Day Image

(c) MCG on Night Image (d) MCG on Ours

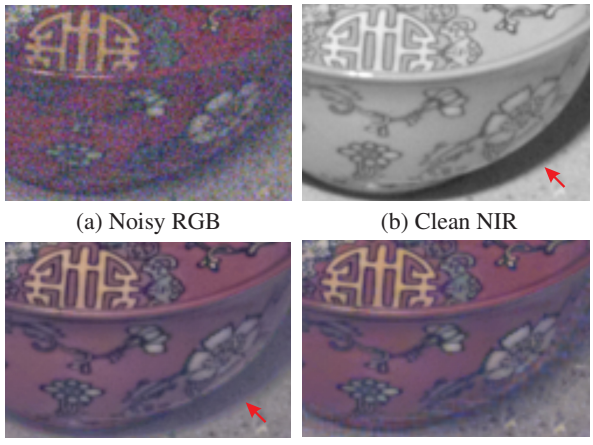
Figure 10. Example of joint segmentation. (a) and (b) are the night and day images respectively. (c) is the result by MCG [1] on (a). (d) is MCG result on our estimated mutual-structure.



(a) Ref. Image (b) Target Image (c) Matched Image

(d) Naive Outlier (e) Mutual Structure (f) Our Outlier

Figure 11. Image matching outlier detection. (a) and (b) are the reference and target images respectively. (c) is the matching result of [25]. (d) is the outlier by naively comparing (a) and (c). (f) shows our detected matching outlier by comparison of (c) and mutual-structure shown in (e).



(a) Noisy RGB (b) Clean NIR

(c) Result of [27] (d) Ours

Figure 12. Example of RGB/NIR image restoration. (a) and (b) are the noisy RGB and clean NIR image respectively. (c) is the result of [27] and (d) is our result.

References

[1] P. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *CVPR*, 2014.

[2] T. Carlo and M. Roberto. Bilateral filtering for gray and color images. In *ICCV*, 1998.

[3] J. Chen, S. Paris, and F. Durand. Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graph.*, 26(3):103, 2007.

[4] A. Criminisi, T. Sharp, C. Rother, and P. Perez. Geodesic image and video editing. *ACM Trans. Graph.*, 29(5):134, 2010.

[5] Z. Farbman, R. Fattal, and D. Lischinski. Diffusion maps for edge-aware image editing. *ACM Trans. Graph.*, 29(6):145, 2010.

[6] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.*, 27(3), 2008.

[7] R. Fattal. Edge-avoiding wavelets and their applications. *ACM Trans. Graph.*, 28(3), 2009.

[8] D. Frédo and D. Julie. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.*, 21(3):257–266, 2002.

[9] E. S. Gastal and M. M. Oliveira. Domain transform for edge-aware image and video processing. *ACM Trans. Graph.*, 30(4):69, 2011.

[10] E. S. Gastal and M. M. Oliveira. Adaptive manifolds for real-time high-dimensional filtering. *ACM Trans. Graph.*, 31(4):33, 2012.

[11] K. He, J. Sun, and X. Tang. Guided image filtering. In *ECCV*, 2010.

[12] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(2):504–511, 2013.

[13] S. Lu, X. Ren, and F. Liu. Depth enhancement via low-rank matrix completion. In *CVPR*, 2014.

[14] Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu. Constant time weighted median filtering for stereo matching and beyond. In *ICCV*, 2013.

[15] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV*, 2006.

[16] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand. Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, 4(1), 2009.

[17] J. Park, H. Kim, Y. Tai, M. S. Brown, and I. Kweon. High quality depth map upsampling for 3d-tof cameras. In *ICCV*, 2011.

[18] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23(3):664–672, 2004.

[19] R. Raskar, A. Ilie, and J. Yu. Image fusion for context enhancement and video surrealism. In *NPAR*, 2004.

[20] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[21] X. Shen, L. Xu, Q. Zhang, and J. Jia. Multi-modal and multi-spectral registration for natural images. In *ECCV*, 2014.

[22] Y. Shih, S. Paris, F. Durand, and W. T. Freeman. Data-driven hallucination of different times of day from a single outdoor photo. *ACM Trans. Graph.*, 32(6), 2013.

[23] J. van de Weijer and R. van den Boomgaard. Local mode filtering. In *CVPR*, 2001.

[24] J. Xiao, H. Cheng, H. S. Sawhney, C. Rao, and M. A. Isnardi. Bilateral filtering-based optical flow estimation with occlusion detection. In *ECCV*, 2006.

[25] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(9):1744–1757, 2012.

[26] L. Xu, Q. Yan, Y. Xia, and J. Jia. Structure extraction from texture via relative total variation. *ACM Trans. Graph.*, 31(6):139, 2012.

[27] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, and J. Jia. Cross-field joint image restoration via scale map. In *ICCV*, 2013.

[28] Q. Yang. Recursive bilateral filtering. In *ECCV*, 2012.

[29] Q. Yang. Stereo matching using tree filtering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2014.

[30] Q. Yang, K.-H. Tan, and N. Ahuja. Real-time o(1) bilateral filtering. In *CVPR*, 2009.

[31] Q. Zhang, X. Shen, L. Xu, and J. Jia. Rolling guidance filter. In *ECCV*, 2014.

[32] Q. Zhang, L. Xu, and J. Jia. 100+ times faster weighted median filter. In *CVPR*, 2014.