Mutual-Structure for Joint Filtering

Xiaoyong Shen[1] · Chao Zhou[1] · Li Xu[2] · Jiaya Jia[1]

Received: 16 April 2016 / Accepted: 18 May 2017 / Published online: 3 June 2017

Abstract Previous joint/guided filters directly transfer structural information from the reference to the target image. In this paper, we analyze the major drawback—that is, there may be completely different edges in the two images. Simply considering all patterns 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. We also use an untraditional objective function that can be efficiently optimized to yield mutual structure. Our method results in important edge preserving property, which greatly benefits depth completion, optical flow estimation, image enhancement, stereo matching, to name a few.

Keywords Image filter · Mutual structure · Joint estimation · Depth refinement · Stereo matching

Communicated by Rene Vidal, Katsushi Ikeuchi, Josef Sivic, Christoph Schnoerr.

✉ Xiaoyong Shen
xyshen@cse.cuhk.edu.hk

Chao Zhou
zhouc@cse.cuhk.edu.hk

Li Xu
xuli@sensetime.com

Jiaya Jia
leojia@cse.cuhk.edu.hk

[1] The Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong

[2] SenseTime Group Limited, Shatin, China

1 Introduction

Image filters are fundamental tools widely used in image editing He et al. (2010), denoising Gastal and Oliveira (2012); Carlo and Roberto (1998), optical flow Xu et al. (2012a); Xiao et al. (2006), stereo matching Ma et al. (2013); Hosni et al. (2013); Yang (2014) and image restoration Petschnigg et al. (2004); Yan et al. (2013). Several filters process single images to either preserve edges Carlo and Roberto (1998); Gastal and Oliveira (2011); Yang et al. (2009); He et al. (2010); Yang (2012); Paris and Durand (2006); Fattal (2009) or remove texture Zhang et al. (2014a); Xu et al. (2012b). Another group of filters, involving the joint bilateral filter Carlo and Roberto (1998) and guided filter He et al. (2010), can take extra images as reference or guidance.

Joint filters are helpful in several tasks. For example, in stereo matching, joint filter can aggregate the cost volume Yang (2014); Hosni et al. (2013). For depth refinement and completion, corresponding RGB images were used in joint filtering Park et al. (2011). The common property is that the reference image provides structural guidance of how the filter should perform. Thus edge preserving or removal on the target image can be achieved locally.

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

However, many practical tasks with images in RGB/ depth Lu et al. (2014), flash/ no-flash Petschnigg et al. (2004), optical flow field/ RGB Xu et al. (2012a), disparity map/ RGB Ma et al. (2013); Hosni et al. (2013), RGB/ NIR Yan et al. (2013), day/ night Raskar et al. (2004) commonly contain inconsistent structure, such as noise, holes, texture, shadow, highlight and multi-spectrum data. They cause trouble during filtering.

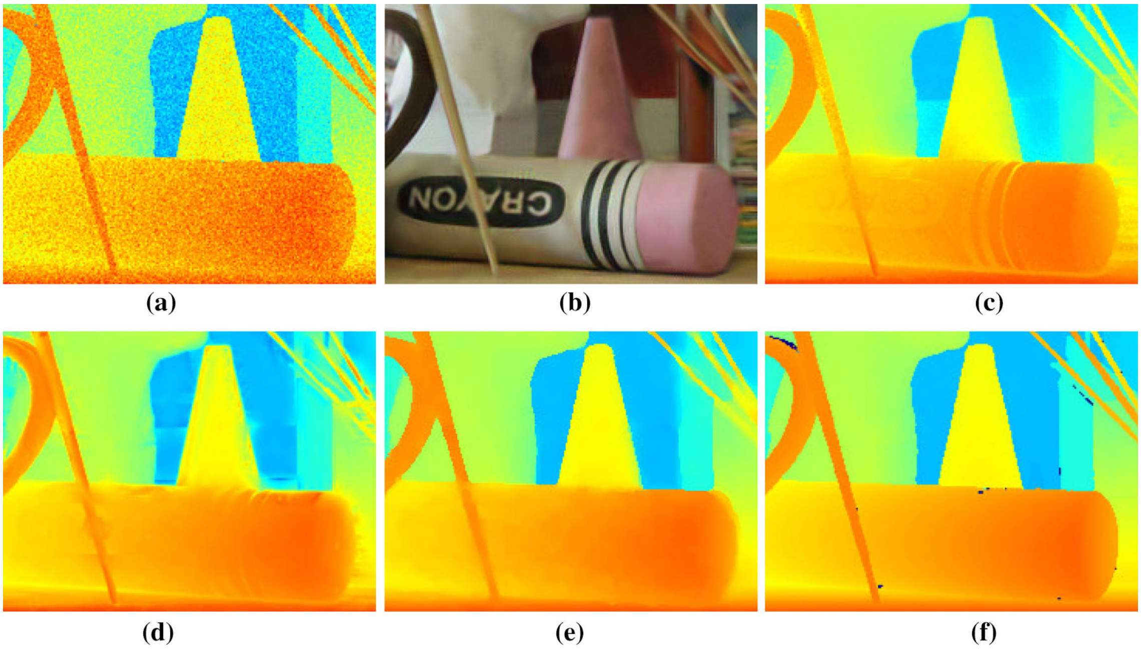


Fig. 1 Examples of joint image filtering on structure-inconsistent image pair. **a**, **b** Are the target and reference images respectively. **c**, **d** Are the results of bilateral and guided image filters respectively, which transfer *color image* structure to depth. **e** Is our result that contains less erroneous patterns from **b**. **f** Is the ground truth (Color figure online)

One example is shown in Fig. 1, where (a) and (b) are the input and reference images. Because (b) includes extra edges not related to depth and the input image (a) is noisy, joint filter generates unwanted structure as shown in (c) and (d).

Our Mutual-Structure for Joint Filtering In this paper, we 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. Instead, we take possible difference into account and estimate mutual structure as a new reference for joint filtering. Our result is shown in Fig.1e, which does not transfer those erroneous reference edges and texture.

This goal is achieved via a new objective function considering the common information between the target and reference images, which will be detailed later. This framework is able to handle images with diverse structure or in different spectral configurations. It optimally suppresses dissimilar information.

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.

The manuscript is an extension of its conference version Shen et al. (2015) published in ICCV'15. The change is fourfold. We give more analysis why the algorithm works in Sect. 5.1. We then propose a more efficient numerical solution in Sect. 5.2. More evaluation on joint depth/RGB restoration and stereo matching is conducted. Finally, we present more applications, including joint shadow detection, in experiment sections.

2 Background and Motivation

We review joint/guided image filters, which are categorized into local and global classes.

Local Joint Methods Local joint filters are mostly the joint extension of single-image edge-preserving filters. Those calculating weighted mean include anisotropic diffusion Farbman et al. (2010), bilateral filter Carlo and Roberto (1998); Frédo and Julie (2002); Paris and Durand (2006); Chen et al. (2007); Yang (2012); Yang et al. (2009), guided filter He et al. (2010), and geodesic distance based filter Criminisi et al. (2010); Gastal and Oliveira (2011). They define various 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 sets affinity weights according to another reference image.

Another line is with weighted median Ma et al. (2013); Zhang et al. (2014b), 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. The general mode filter is presented in van de Weijer and van den Boomgaard (2001).

Global Joint Schemes Global methods optimize functions. They include total variation (TV) Rudin et al. (1992), weighted least squares (WLS) Farbman et al. (2008), and scale map scheme Yan et al. (2013). 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 weights based on the reference image.

To summarize related work, almost all joint image filters identify important structure based on the reference image. These methods work best when the reference data only contains 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 and special-type images. We describe our method in following sections.

3 Mutual-Structure for Joint Filtering

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

- *Mutual structure* As shown in (c), mutual structure can be intuitively understood as common edges in the corresponding two patches. These edges are *not* necessarily with the same magnitude. The gradient direction can also be reversed.
- *Inconsistent structure* Inconsistent structure represents different patterns in two patches. There may be many such structures in an image pair as shown in Fig. 2d. When an edge appears only in one patch 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).

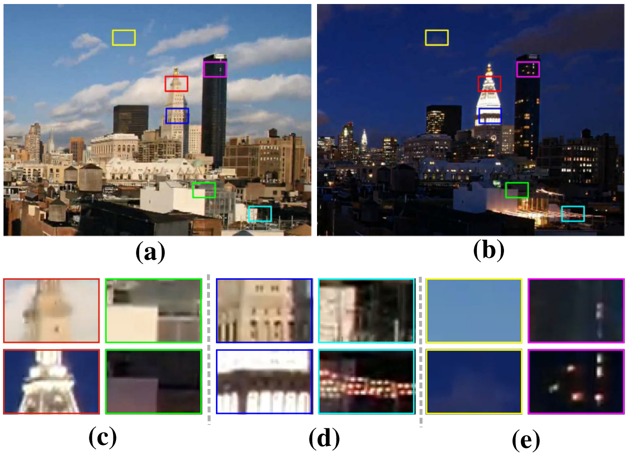


Fig. 2 Examples of image structure correlation in a day/night image pair. **a**, **b** Day and night images respectively. **c** Mutual structure close-up. **d** Inconsistent structure patches. **e** Smooth patches. The images are from the time-lapse video of Shih et al. (2013)

Among these types of joint structure, inconsistent edges generally cause big problems if we transfer erroneous patterns to the target image. In this paper, we aim to find *mutual-structure* and let it guide the joint filtering process. Accordingly, we not only filter the target image, but optimize the reference as well based on a structure similarity measure.

We give 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 structure 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 the set of pixels 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 is 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 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 denoted as

$$cov(I_p, G_p) = \frac{1}{|N|} \sum_{q \in N(p)} (I_q - \bar{I}_p)(G_q - \bar{G}_p). \tag{2}$$

$N(p)$ is the set of pixels in patch p and $|N|$ is the number of pixels in $N(p)$. $\bar{I}_p$ and $\bar{G}_p$ are the mean intensity of patch p in I and G respectively. $\sigma(I_p)$ denotes variance of patch p in I as

$$\sigma(I_p) = \frac{1}{|N|} \sum_{q \in N(p)} (I_q - \bar{I}_p)^2, \tag{3}$$

and $\sigma(G_p)$ is the variance of patch p in G that is similar to $\sigma(I_p)$ in definition.

When two patches share 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 patch structures are similar.

Albeit ideal in measurement, NCC is hard to use directly due to its nonlinearity. We provide the following derivation to establish the relationship between NCC and a 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

$$f(I, G, a_p^1, a_p^0) = \sum_{q \in N(p)} \left(a_p^1 I_q + a_p^0 - G_q\right)^2, \tag{4}$$

where a_p^1 and a_p^0 are the regression coefficients. This function linearly represents one patch in G according to 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

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

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

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

$$e(I_p, G_p) = \sigma(G_p)(1 - \rho(I_p, G_p)^2), \tag{6}$$

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

We refer readers to Appendix for the complete proof. The claim explains when $|\rho(I_p, G_p)| = 1$, which means the two patches only contain mutual structure, $e(I_p, G_p)$ reaches 0. Following the same procedure, we construct

$$e(G_p, I_p)^2 = \min_{b_p^1, b_p^0} \frac{1}{|N|} f\left(G, I, b_p^1, b_p^0\right), \tag{7}$$

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 above two functions as

$$\mathcal{S}(I_p, G_p) = e(I_p, G_p)^2 + e(G_p, I_p)^2. \tag{8}$$

According to Eqs. (5) and (6) and considering $\rho(I_p, G_p) = \rho(G_p, I_p)$, this measure boils down to

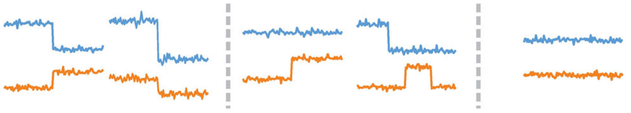

(a) (b) (c)

Fig. 3 1D Example. **a** Mutual structure in two patches. **b** Inconsistent structure. **c** Smooth regions

$$\mathcal{S}(I_p, G_p) = \left(\sigma(I_p)^2 + \sigma(G_p)^2\right)\left(1 - \rho(I_p, G_p)^2\right)^2. \tag{9}$$

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

- *Mutual-Structure Patches* When $|\rho(I_p, G_p)|$ approaches 1, $\mathcal{S}(I_p, G_p)$ moves towards 0 in Eq. (9) indicating the two patches are with common edges as shown in Fig. 3a.
- *Inconsistent Structure Patches* As shown in (b), when NCC $|\rho(I_p, G_p)|$ outputs a small value for patches containing edges [i.e., at least $\sigma(I_p)$ or $\sigma(G_p)$ is large in Eq. (9)], 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.

According to the above analysis, optimizing Eq. (9) 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 similarity term as

$$E_{\mathcal{S}}(I, G, a, b) = \sum_p \left(f(I, G, a_p^1, a_p^0) + f(G, I, b_p^1, b_p^0)\right), \tag{10}$$

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. We consider regression on a single image channel due to the efficiency in optimization. It is also straightforward to make it work in multiple channels.

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 thus incorporate more constraints to avoid it.

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

$$E_d(I, G) = \sum_p \lambda \| G_p - G_{0,p} \| + \beta \| I_p - I_{0,p} \|, \tag{11}$$

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, we reduce patch intensity variance. In Eq. (8), 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

$$E_r(a, b) = \sum_p \left(\varepsilon_1 {a_p^1}^2 + \varepsilon_2 {b_p^1}^2 \right), \tag{12}$$

where ε_1 and ε_2 are very small values, which control smoothness strength on G and I respectively. Note that this term is related to the ridge regression applied by guided image filter He et al. (2010). But our form is different on incorporating two-direction regression errors.

4.3 Final Objective

According to the mutual-structure, our final objective function for jointly estimating I and G combines the above three terms:

$$E(I, G, a, b) = E_{\mathcal{S}}(I, G, a, b) + E_d(I, G) + E_r(a, b). \tag{13}$$

a and b are regression coefficient sets, which are also solved for. 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 alternating optimization based on the derivatives and Jacobi method Yan et al. (2013) 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 a quadratic problem in closed form.

Algorithm 1 Mutual-Structure Estimation

```
Require: $I_0, G_0, N^{\text{iter}}, \lambda, \beta, \varepsilon_1, \varepsilon_2$
Ensure: $I, G$
Estimate ${a^1}^{(0)}$ and ${b^1}^{(0)}$ by Eq. (24).
Compute ${a^0}^{(0)}$ and ${b^0}^{(0)}$ via Eqs (15) and (16).
Update $I^{(0)}$ and $G^{(0)}$ by Eq. (17).
for t:= 0 to $N^{\text{iter}}$ do
    Update $a^{(t)}$ and $b^{(t)}$ according to Eqs. (14), (15) and (16).
    Optimize $G^{(t+1)}$ and $I^{(t+1)}$ by Eq. (17).
end for
$I \leftarrow I^{(N^{\text{iter}})}, G \leftarrow G^{(N^{\text{iter}})}$
```

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

$$\begin{aligned} {a_p^1}^{(t)} &= \frac{cov\left(I_p^{(t)}, G_p^{(t)}\right)}{\sigma(I_p^{(t)}) + \varepsilon_1}, {b_p^1}^{(t)} = \frac{cov\left(G_p^{(t)}, I_p^{(t)}\right)}{\sigma(G_p^{(t)}) + \varepsilon_2}, && (14) \\ {a_p^0}^{(t)} &= \mu(G_p^{(t)}) - {a_p^1}^{(t)} \mu(I_p^{(t)}), && (15) \\ {b_p^0}^{(t)} &= \mu(I_p^{(t)}) - {b_p^1}^{(t)} \mu(G_p^{(t)}), && (16) \end{aligned}$$

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

$$\begin{cases} G_p^{(t+1)} = \frac{1}{M_G^{(t)}} \left(J_G^{(t)} I_p^{(t+1)} + K_G^{(t)} + \lambda G_{0,p} \right), \\ I_p^{(t+1)} = \frac{1}{M_I^{(t)}} \left(J_I^{(t)} G_p^{(t+1)} + K_I^{(t)} + \beta I_{0,p} \right), \end{cases} \tag{17}$$

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

$$\begin{aligned} J_G^{(t)} &= \mu({b_p^1}^{(t)}) + \mu({a_p^1}^{(t)}), \\ J_I^{(t)} &= \mu({b_p^1}^{(t)}) + \mu({a_p^1}^{(t)}). \end{aligned} \tag{18}$$

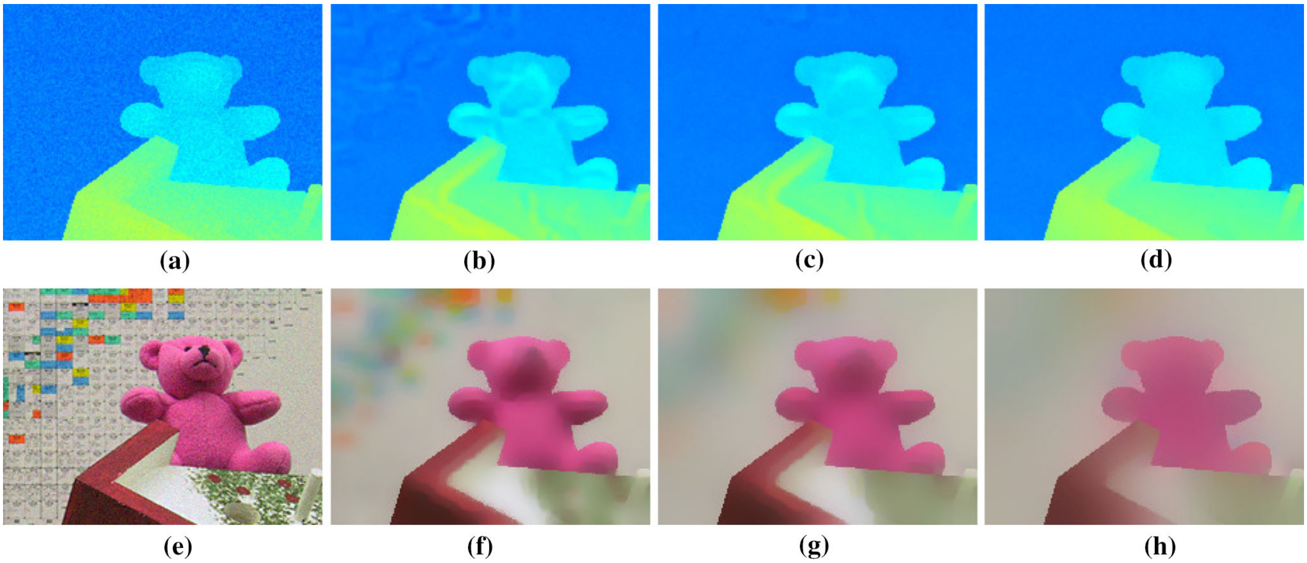
$K_G^{(t)}$ and $K_I^{(t)}$ are the constant denoted as

$$\begin{aligned} K_G^{(t)} &= \mu({a_p^0}^{(t)}) - \mu\left({b_p^1}^{(t)} {b_p^0}^{(t)}\right), \\ K_I^{(t)} &= \mu({b_p^0}^{(t)}) - \mu\left({a_p^1}^{(t)} {a_p^0}^{(t)}\right). \end{aligned} \tag{19}$$

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

$$\begin{aligned} M_G^{(t)} &= \frac{\lambda}{|N|} + \mu\left({b_p^1}^{(t)} {b_p^1}^{(t)}\right) + 1, \\ M_I^{(t)} &= \frac{\beta}{|N|} + \mu\left({a_p^1}^{(t)} {a_p^1}^{(t)}\right) + 1. \end{aligned} \tag{20}$$

The update stages only contain the simple mean operation and multiplication. They can be implemented efficiently


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

(e) (f) (g) (h)

Fig. 4 Result updated in iterations. Given the noisy natural image in **e** and imperfect depth layer in **a**, **b**, **f** show results of I and G in the first iteration. **c**, **g** Are the results after ten iterations. **d**, **h** Are the final results after 20 iterations

employing the box filter. We apply the fast box filter based on the integral image implemented in He et al. (2010).

5.1 Algorithm Analysis

We first update $a^{(t)}$ and $b^{(t)}$ according to Eqs. (14), (15) and (16). Then $G^{(t+1)}$ and $I^{(t+1)}$ are optimized by Eq. (17). ε_1 and ε_2 play important roles for extracting mutual-structures and removing inconsistent ones. We in what follows analyze the effect according to various types of structure correlation.

- *Inconsistent Structure Patches* Since covariance between inconsistent structure patches moves to zero, ε_1 and ε_2 defined in Eq. (14) make $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ close to zeros. According to the optimization form for $G_p^{(t+1)}$ and $I_p^{(t+1)}$ defined in Eq. (17), we achieve

$$G_p^{(t+1)} \approx \frac{|N|}{\lambda + |N|}\left(\mu(\mu(G_p^{(t)})) + \lambda G_{0,p}\right), \qquad (21)$$

$$I_p^{(t+1)} \approx \frac{|N|}{\lambda + |N|}\left(\mu(\mu(I_p^{(t)})) + \beta I_{0,p}\right), \qquad (22)$$

  by omitting the terms related to $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ in Eq. (17). Thus we get the updated $G_p^{(t+1)}$ as the linear combination of the filtered $G_p^{(t)}$ and the original input $G_{0,p}$. This update not only helps removing inconsistent structures but also makes output still similar to the original image.
  Similar analysis applies to the updating process of $I_p^{(t+1)}$. Moreover, larger ε_1 and ε_2 make $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ go closer to zeros, which is desirable as discussed. Relatively large λ or β preserves input image appearance.
- *Mutual-Structure Patches* For $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ defined in Eq. (14), they do not go to zero because of structure covariance $cov(I_p^{(t)})$ and non-zero variance of $\sigma(I_p^t)$ and $\sigma(G_p^t)$. Taking derivatives on $G_p^{(t+1)}$ and $I_p^{(t+1)}$ in Eq. (17), we get $\nabla G_p^{(t+1)} = \eta \nabla I_p^{(t+1)}$ where $\eta = J_G^{(t)}/M_G^{(t)}$. η is not zero because $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ are not in mutual-structure patches. So structure correlation is preserved in iterations because of the gradient relation.
- *Smooth Patches* Similar to inconsistent structure patches, $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ generally are not with small values. Update of $G_p^{(t+1)}$ and $I_p^{(t+1)}$ corresponds to fusion of $G_p^{(t)}$ and $I_p^{(t)}$. This process reduces artifacts and noise in smooth regions.

To demonstrate the iterative updating effect, we show an example in Fig. 4 where the input is a captured depth image with noticeable noise. The reference image is the corresponding color one. Inconsistent edges and texture exist. We show the results of our method in iterations 1 and 10 where inconsistent edges are removed gradually. After convergence in 20 iterations, our results are only with edges existing in both images under proper smoothing to remove noise and inconsistency.

5.2 Initialization

As discussed in Sect. 5.1, $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ are important to remove inconsistent structures and preserve mutual-structure. Good initialization of a_p^1 and b_p^1 is essential for fast convergence and avoiding local minima. Since both $a_p^{1\,(t)}$ and $b_p^{1\,(t)}$ converge to zeros for inconsistent and smooth region patches, initializing them to zeros is a good choice. We roughly find mutual-structure patches between I_0 and G_0 by a generalized normalized cross correlation (NCC) measure, which is defined as

$$\zeta(I_{0,p}, G_{0,p}) = \frac{cov(I_{0,p}, G_{0,p})^2}{\left(\sigma(I_{0,p}) + \varepsilon_1\right)\left(\sigma(G_{0,p}) + \varepsilon_2\right)}, \qquad (23)$$

where ε_1 and ε_2 are very small values to avoid deviation by zero. For simplicity, we define them the same as the parameters in Eq. (12). $\zeta(I_{0,p}, G_{0,p})$ is close to square of NCC since ε_1 and ε_2 are very small values. Thus, $\zeta(I_{0,p}, G_{0,p})$ approaches 1 for mutual-structure patches and is close to

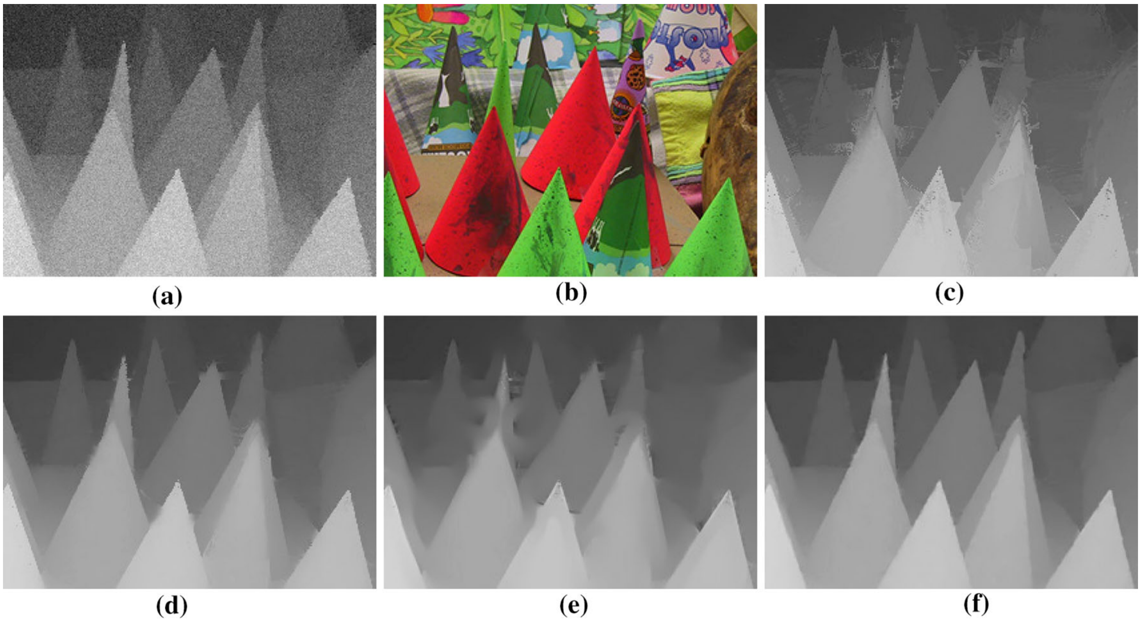


Fig. 5 Comparison with other iterative joint filters. **a**, **b** Show the input and reference images respectively. **c**, **d** Are the results of iterative joint bilateral filter and rolling guidance filter. **e** Is obtained by alternatively applying guided filter using Eq. (25). These three results all have unwanted structure transferred from the *color image* to depth. **f** Is our result (Color figure online)

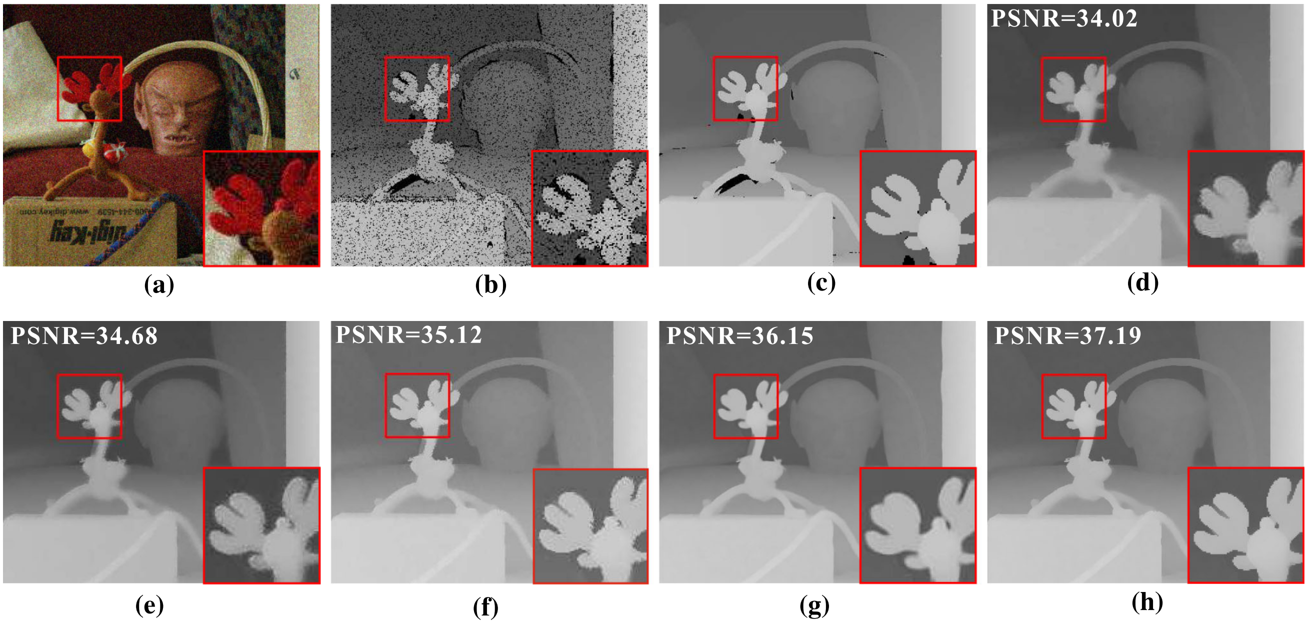


Fig. 6 Noisy RGB/depth image restoration by different methods. **a**, **b** Show the input and reference images respectively. **c** Is the ground truth depth. **d**–**h** Are the results of different methods. Among them, **g** is shown in paper Lu et al. (2014). PSNRs are reported for all results

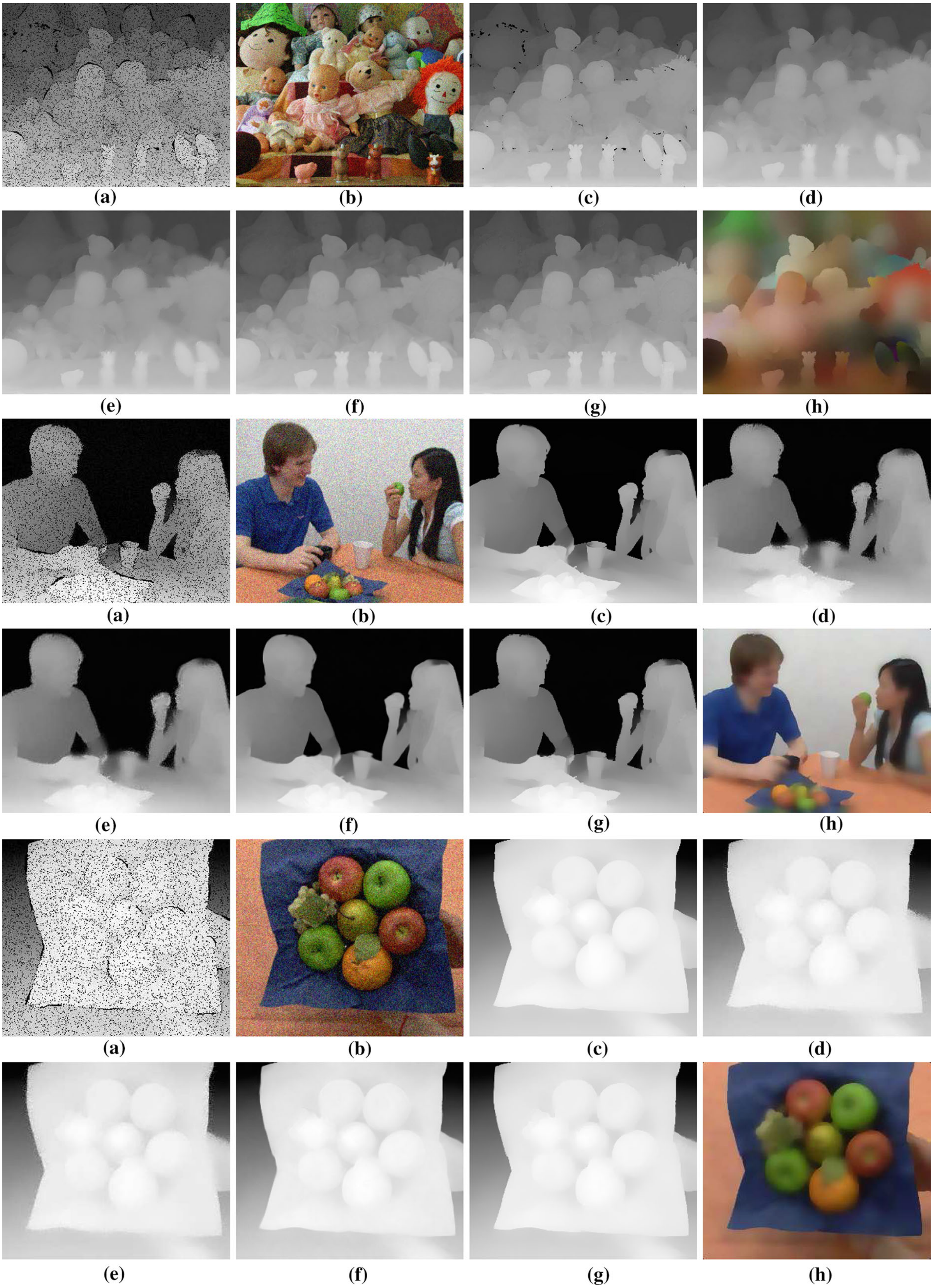


Fig. 7 More Examples of noisy RGB/depth image restoration. **a**, **b** Are the inputs. **c** Shows the ground truth. **d**, **e** Are results of bilateral filter Paris and Durand (2006), guided filter He et al. (2010) and method of Lu et al. (2014). **g** Is our result. **h** Shows our estimated mutual-structure

Table 1 Comparison of different methods for RGB/depth restoration on the dataset of Lu et al. (2014)

Methods	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
Bilateral filter	31.59	29.18	35.36	38.77	39.37	37.35	32.69	34.86	44.51	40.66	39.93	33.61	35.00	39.67	41.39
Guided filter	31.40	28.97	35.24	38.56	39.13	37.16	32.63	34.65	44.89	40.21	39.71	33.55	34.70	39.47	40.87
Weighted median	33.92	31.48	37.10	41.32	40.73	39.84	35.67	37.22	45.86	42.75	41.95	36.90	37.43	41.99	43.49
Lu et al.	35.24	33.10	39.00	42.70	42.66	41.13	38.05	**39.78**	46.23	42.39	42.62	37.72	39.03	42.01	**44.08**
Ours	**35.48**	**33.34**	**39.11**	**43.05**	**42.84**	**41.25**	**38.17**	39.67	**46.46**	**42.79**	**42.89**	**37.85**	**39.44**	**42.19**	44.07
Methods	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Bilateral filter	34.35	34.86	33.96	36.74	36.13	37.56	33.73	39.11	33.66	42.76	44.28	38.44	36.90	36.83	38.14
Guided filter	34.23	34.74	33.61	36.30	35.81	37.43	33.56	38.64	33.24	42.84	44.32	38.06	36.82	35.82	37.07
Weighted median	38.87	39.35	37.69	39.27	39.05	40.23	36.20	**43.78**	36.10	44.61	45.50	39.59	**41.83**	39.12	41.15
Lu et al.	39.13	38.88	38.59	**39.91**	39.12	41.40	**37.16**	42.35	36.15	45.26	46.13	**40.78**	39.97	40.32	41.26
Ours	**39.40**	**39.65**	**38.62**	39.87	**39.60**	**41.72**	37.10	43.49	**36.93**	**45.33**	**46.67**	40.64	41.54	**40.51**	**41.63**

We report PSNRs of joint bilateral filter Paris and Durand (2006), guided filter He et al. (2010), joint weighted median Zhang et al. (2014b) and ours. The best results are highlighted

0 for others. Our initialization for ${a_p^1}^{(t)}$ and ${b_p^1}^{(t)}$ is set to

$$\begin{cases} {a_p^1}^{(0)} = 0, {b_p^1}^{(0)} = 0, & \text{when} \quad \zeta(I_{0,p}, G_{0,p}) < \tau, \\ {a_p^1}^{(0)} = \frac{cov(I_{p,0}, G_{p,0})}{\sigma(I_{p,0}) + \varepsilon_1}, & \\ {b_p^1}^{(0)} = \frac{cov(G_{p,0}, I_{p,0})}{\sigma(G_{p,0}) + \varepsilon_2}, & \text{otherwise}, \end{cases} \tag{24}$$

where τ is the threshold, set to 0.8 in all our experiments. For mutual-structure patches, ${a_p^1}^{(0)}$ and ${b_p^1}^{(0)}$ are directly estimated from input by Eq. (14). Computation of $\zeta(I_{0,p}, G_{0,p})$ can be incorporated into our algorithm since it equals to ${a_p^1}^{(t)}{b_p^1}^{(t)}$ when $I_p^{(t)}$ and $G_p^{(t)}$ are $I_{0,p}$ and $G_{0,p}$ respectively. With ${a_p^1}^{(0)}$ and ${b_p^1}^{(0)}$, ${a_p^0}^{(0)}$ and ${b_p^0}^{(0)}$ are computed via Eqs. (15) and (16) respectively. We estimate $G_p^{(0)}$ and $I_p^{(0)}$ by Eq. (17). The complete Algorithm 1 involves the initialization step.

Compared with our algorithm presented in the conference version Shen et al. (2015) where initialization is performed by rolling guidance filtering Zhang et al. (2014a), our new scheme achieves faster convergence. In our experiments, only 14 iterations are enough on average to produce compelling results while the original scheme needs 20 iterations. The running time is also shortened.

5.3 Relationship with Other Methods

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

We first compare our solution with iterative joint bilateral filter Paris et al. (2009), which iteratively filters the input with the fixed reference image. Although both methods are edge preserving, the iterative joint bilateral filter does not address our aforementioned structure transfer problem. We show an example in Fig. 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 compare our method with rolling guidance filter (RGF) Zhang et al. (2014b). 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 Gastal and Oliveira (2012). As shown in (d), it still cannot get the mutual structure and is hard to avoid incorrect structure transfer.

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

$$\begin{aligned} I^{(t+1)} &= GF(I^{(t)}, G^{(t)}), \\ G^{(t+1)} &= GF(G^{(t)}, I^{(t+1)}), \end{aligned} \tag{25}$$

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 Fig. 5e, which suffers from the same problem.

Our result shown in (f) is better because we take both removal of inconsistent structure and preservation of mutual edges into account.

6 Experiments and Applications

Our method takes aligned target and reference images as input. We employ the dense multi-modal and spectral matching method Shen et al. (2014) to align them if there exists non-rigid displacement between images.

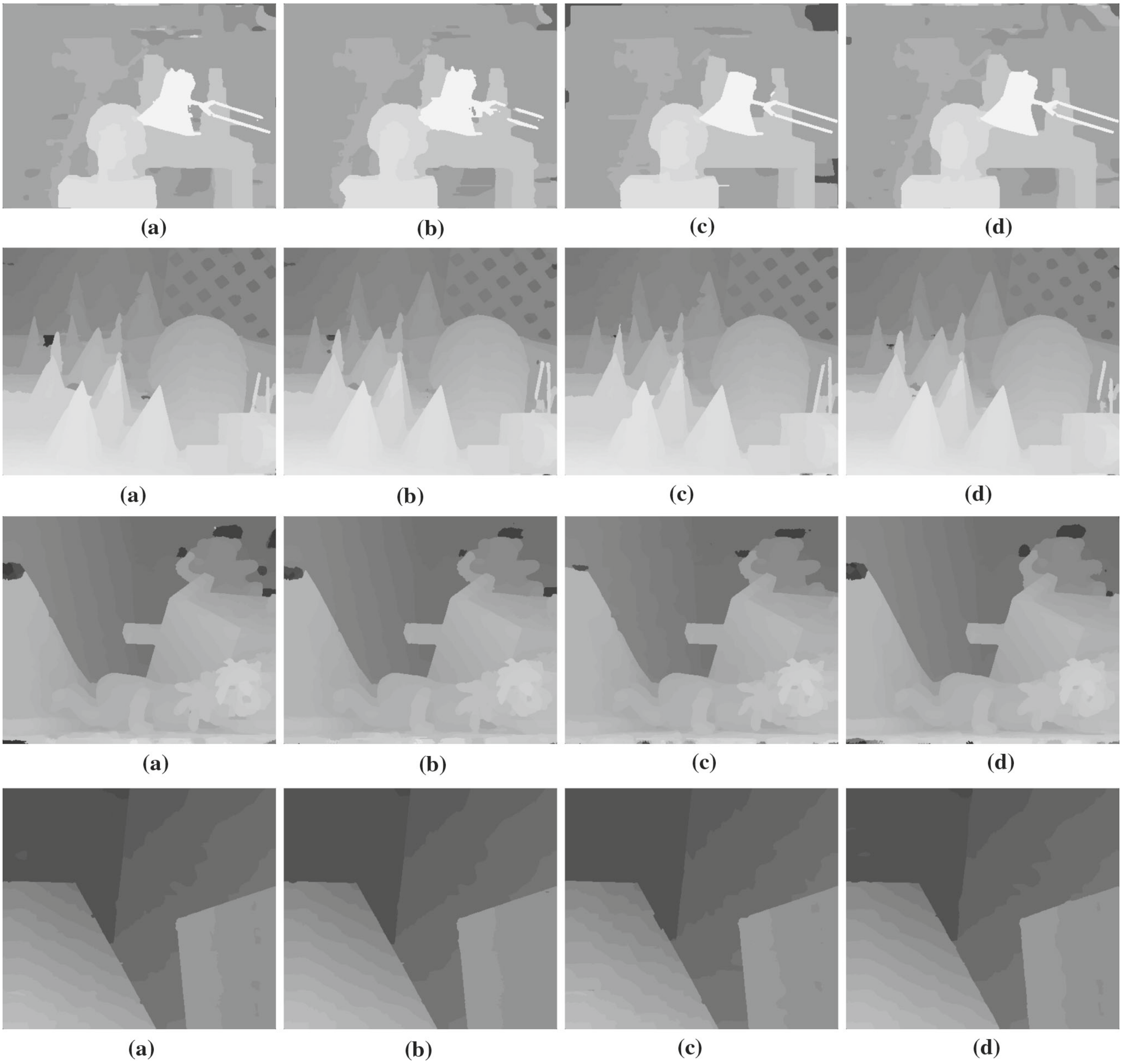


Fig. 8 Comparison on stereo matching. **a**–**c** Show results of bilateral filter, guided filter and tree filter. **d** Is our result. Pixel errors larger than 1 pixel are reported. The inputs are from the Mideleburry stereo matching dataset Scharstein and Szeliski (2002)

We extensively evaluate mutual-structure for joint filtering. Our algorithm is easy to implement and the code is publicly available in our website.[1] 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. We set them around $1E-5$.

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 s with 20 iterations in MATLAB.

6.1 Applications

Our mutual-structure for joint filtering benefits several important applications due to 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 com-

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

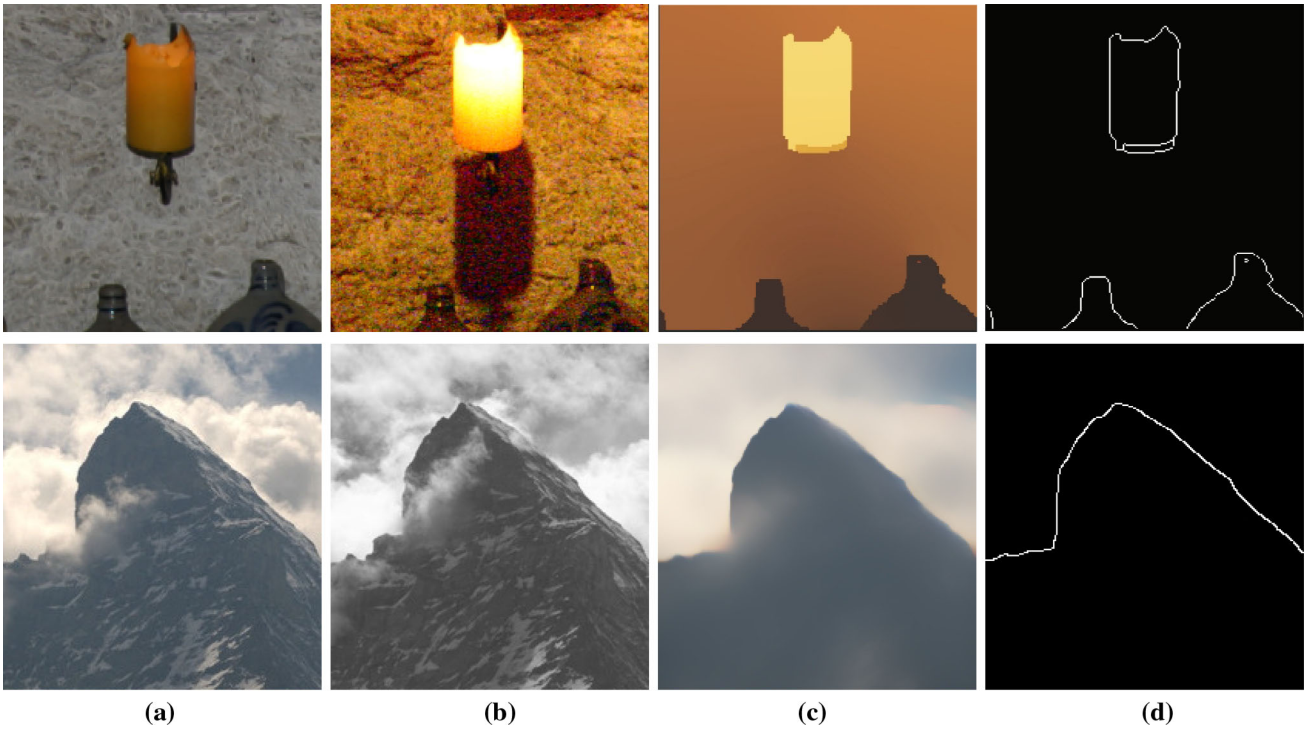


Fig. 9 Joint structure extraction. **a**, **b** Are two inputs. **c** Is the estimated mutual-structure. **d** Shows the common structure of **a**, **b** extracted from the mutual structure (**c**)

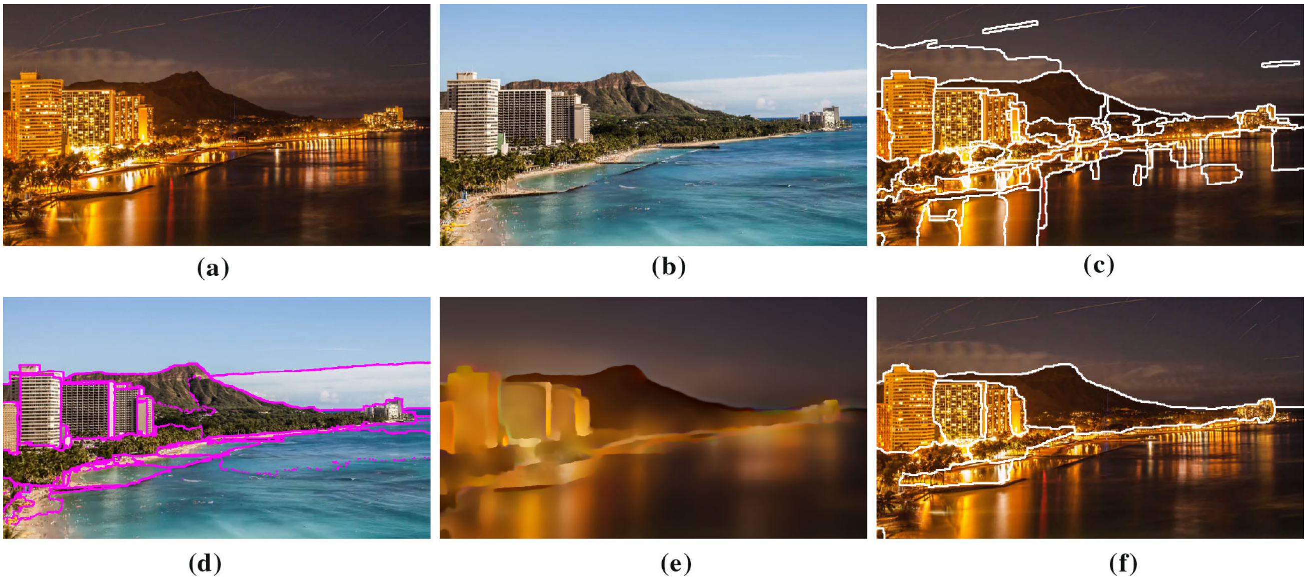


Fig. 10 Example of joint segmentation. **a**, **b** Are the night and day images respectively. **c**, **d** Are the results of MCG Arbeláez et al. (2014) on **a**, **b**. **d** Is our estimated mutual-structure and **f** is the MCG result on our estimated mutual-structure

parable to or outperforms other filtering schemes due to its unique mutual-structure property.

6.1.1 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 joint bilateral filter Carlo and Roberto (1998), guided image filter He et al. (2010), weighted median filter Zhang et al. (2014b), the method of Lu et al. (2014) and our mutual-structure. (d–

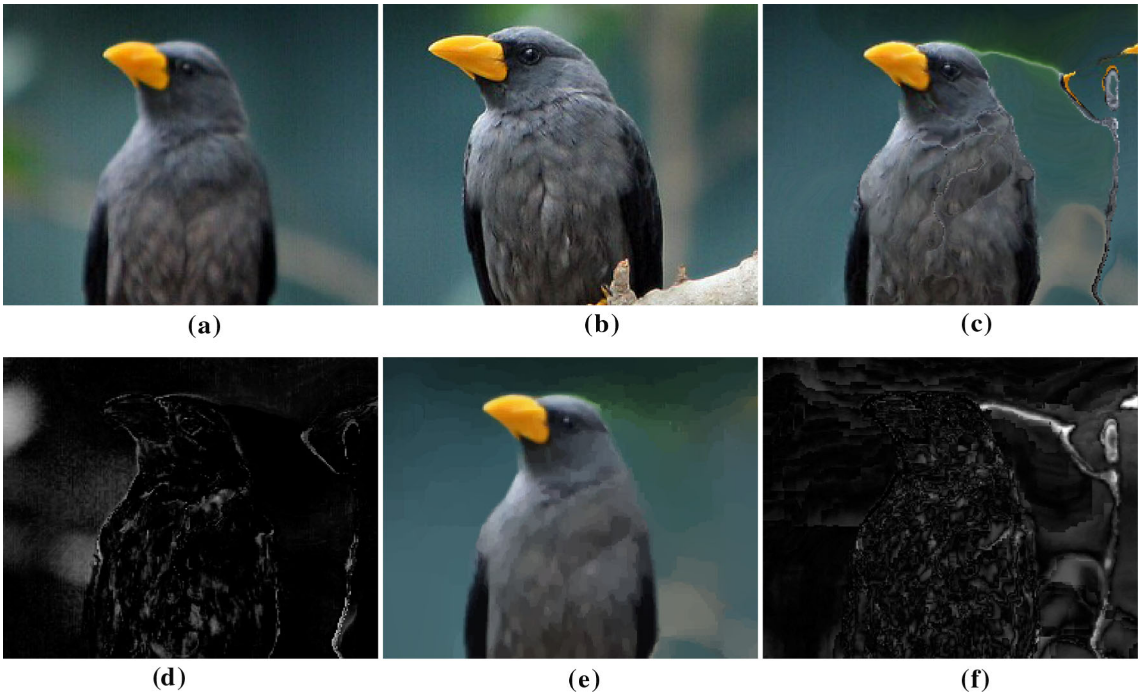


Fig. 11 Image matching outlier detection. **a**, **b** Are the reference and target images respectively. **c** Is the matching result of Xu et al. (2012a). **d** Is the outlier by naively comparing **a**, **c**. **f** Shows our detected matching outlier by comparison of **c** and mutual-structure shown in **e**

f) are produced by joint bilateral filter, guided image filter, and weighted median filter without mutual-structure computation. The result of Lu et al. (2014) in (g) is a bit blurry because of the patch-based scheme. Our method achieves decent results without transferring erroneous structure as shown in (h). PSNR is calculated for each method. More results are shown in Fig. 7.

We also evaluate our mutual-structure method for RGB/depth restoration on the dataset of Lu et al. (2014), which includes 30 pairs of RGB/depth images with ground truth synthesized from Middlebury dataset Scharstein and Szeliski (2002). Our method achieves 0.2% higher PSNR compared with state-of-the-art solution on average as reported in Table 1. The speed is 55+ times faster because we only need a few quick iterations.

6.1.2 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. Hosni et al. (2013). 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 Carlo and Roberto (1998); Yang (2014), guided image filter He et al. (2010); Hosni et al. (2013), and tree filter Yang (2014) in the cost aggregation step. According to the evaluations on Middlebury stereo matching dataset Scharstein and Szeliski (2002) shown in Fig. 8, our method achieves premier performance.

6.1.3 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 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 Fig. 9 where (a) and (b) are input with inconsistent edge structures. (c) shows our estimated mutual-structure, which only includes common edges detected by the traditional Canny edge detector as shown in (d).

Our mutual-structure also benefits joint segmentation for complex scenes as shown in Fig. 10 where (a) and (b) are the night and day images respectively. (c) and (d) are the MCG Arbeláez et al. (2014) segmentation results. (f) is the result with MCG applied to our mutual-structure result (e), which intriguingly is with better segmented objects common in both images thanks to removal of complex and inconsistent patterns. This is exactly what we aim to accomplish.

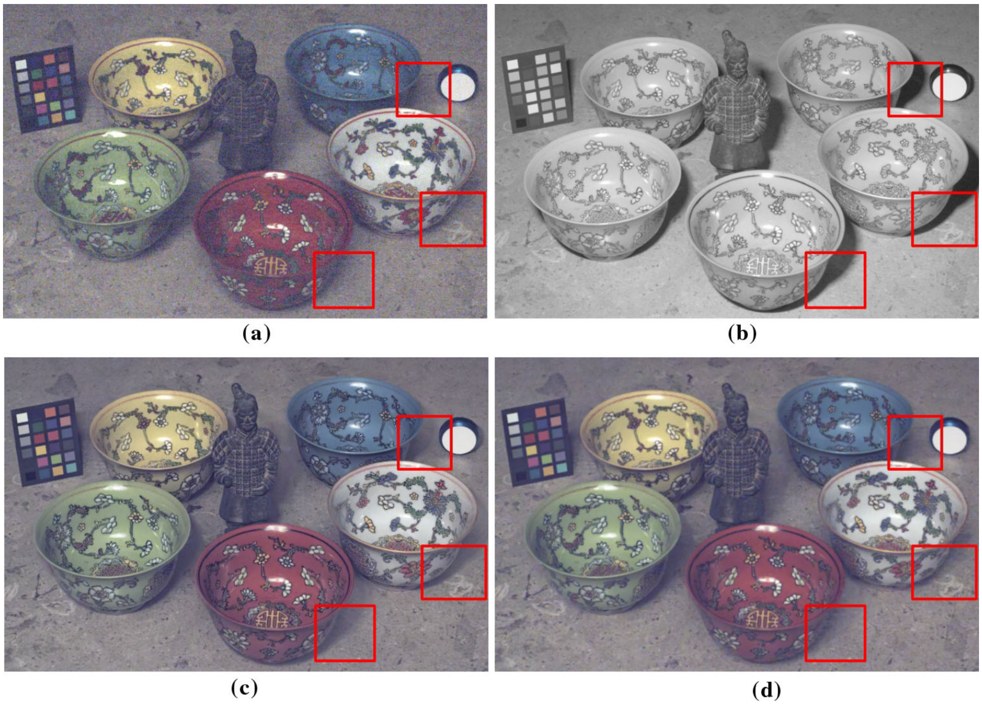


Fig. 12 Example of RGB/NIR image restoration. **a** Is the noisy RGB image and **b** is the clean NIR image with shadow. **c** Is the result of Yan et al. (2013), which transfers the shadow structure to the output. **d** Is our result without this problem. The inputs are from Krishnan and Fergus (2009)

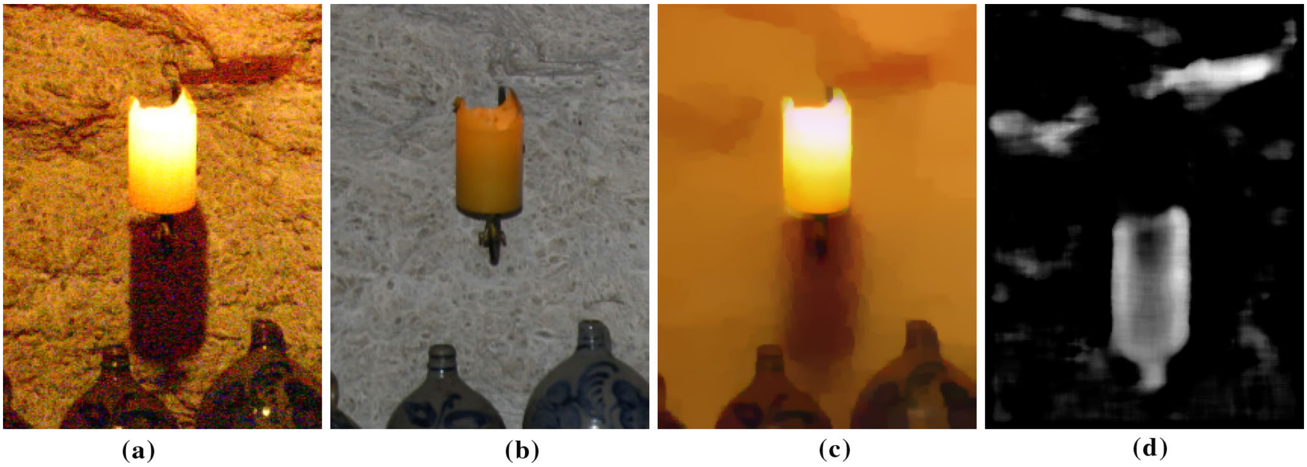


Fig. 13 Example of joint shadow detection. **a**, **b** Are noisy no-flash and flash images respectively. **c** Is our estimated mutual structure of **a**, **b**. **d** Shows our shadow detection result. The inputs are from Petschnigg et al. (2004)

6.1.4 Matching Outlier Detection

One very challenging problem in image matching is on how to detect matching outliers. We handle it by finding common structures, so that the residual between warped images can stand out in contrast to estimated mutual structure. This forms an optimal matching-outlier map. We show an example in Fig. 11 to illustrate the effectiveness to find mismatch. (a) and (b) are the reference and target images respectively. (c) is the matching result estimated by optical flow method Xu et al. (2012a). (d) is the outlier by naively comparing (a) and (c), which cannot reveal the matching outlier because of structure discrepancy between (a) and (b). (f) shows our detected matching outlier by comparison of (c) and mutual-structure shown in (e). Our outlier map shows the matching errors. Since we only consider mismatched structure, textureless regions cannot be dealt with (Fig. 12).

6.1.5 More Applications

Our joint filtering method can also deal with structure transfer in RGB/NIR image restoration. Compared with state-of-the-art method of Yan et al. (2013), our mutual-structure for joint filtering produces comparable results. The running time is 20 times less because of the efficient iteration steps.

Our extracted mutual structure can also be applied to joint shadow detection as shown in Fig. 13 where (a) is the input with shadow while (b) is not. The mutual structure shown in (c) contains common edges between (a) and (b). For this example, the shadow in (a) can be directly obtained by finding the difference between (a) and (c). Our coarse shadow detection result is shown in (d), which manifests that this could be a promising direction for further pursuit (13)

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 structure from the reference to target images. We have discussed that this type of structure discrepancy commonly exists 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 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 413113.

Appendix: Proof of Claim 4.1

Proof In Eq. (5), $e(I_p, G_p)$ reaches the minimum when setting the derivatives $\frac{\partial e(I_p,G_p)^2}{\partial a_p^1}$ and $\frac{\partial e(I_p,G_p)^2}{\partial a_p^0}$ to zeros, which yields

$$a_p^1 = \frac{cov(I_p, G_p)}{\sigma(I_p)}, \quad a_p^0 = \overline{G}_p - a_p^1 \overline{I}_p, \tag{26}$$

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. (5) and arranging it according to Eq. (1), we obtain

$$e(I_p, G_p) = \sigma(G_p)(1 - \rho(I_p, G_p)^2), \tag{27}$$

as shown in Eq. (6). □

References

Arbeláez, P., Pont-Tuset, J., Barron, J. T., Marques, F., & Malik, J. (2014). Multiscale combinatorial grouping. In *CVPR*.

Carlo, T., & Roberto, M. (1998). Bilateral filtering for gray and color images. In *ICCV*.

Chen, J., Paris, S., & Durand, F. (2007). Real-time edge-aware image processing with the bilateral grid. *ACM Transactions on Graphics*, *26*(3), 103.

Criminisi, A., Sharp, T., Rother, C., & Perez, P. (2010). Geodesic image and video editing. *ACM Transactions on Graphics*, *29*(5), 134.

Farbman, Z., Fattal, R., Lischinski, D., & Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics*, *27*(3), 67:1–67:10.

Farbman, Z., Fattal, R., & Lischinski, D. (2010). Diffusion maps for edge-aware image editing. *ACM Transactions on Graphics*, *29*(6), 145.

Fattal, R. (2009). Edge-avoiding wavelets and their applications. *ACM Transactions on Graphics*, *28*(3), 22:1–22:10.

Frédo, D., & Julie, D. (2002). Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, *21*(3), 257–266.

Gastal, E. S., & Oliveira, M. M. (2011). Domain transform for edge-aware image and video processing. *ACM Transactions on Graphics*, *30*(4), 69.

Gastal, E. S., & Oliveira, M. M. (2012). Adaptive manifolds for real-time high-dimensional filtering. *ACM Transactions on Graphics*, *31*(4), 33.

He, K., Sun, J., & Tang, X. (2010). Guided image filtering. In *ECCV*.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C., & Gelautz, M. (2013). Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(2), 504–511.

Krishnan, D., & Fergus, R. (2009). Dark flash photography. *ACM Transactions on Graphics*, *28*(3).

Lu, S., Ren, X., & Liu, F. (2014). Depth enhancement via low-rank matrix completion. In *CVPR*.

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

Paris, S., & Durand, F. (2006). A fast approximation of the bilateral filter using a signal processing approach. In *ECCV*.

Paris, S., Kornprobst, P., Tumblin, J., & Durand, F. (2009). Bilateral filtering: Theory and applications. *Foundations and Trends in Computer Graphics and Vision*, *4*(1), 1–74.

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

Petschnigg, G., Szeliski, R., Agrawala, M., Cohen, M., Hoppe, H., & Toyama, K. (2004). Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, *23*(3), 664–672.

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

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

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal on Computer Vision*, *47*(1–3), 7–42.

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

Shen, X., Zhou, C., Xu, L., & Jia, J. (2015). Mutual-structure for joint filtering. In *ICCV*, pp 3406–3414.

Shih, Y., Paris, S., Durand, F., & Freeman, W. T. (2013). Data-driven hallucination of different times of day from a single outdoor photo. *ACM Transactions on Graphics*, *32*(6), 200:1–200:11.

van de Weijer, J., & van den Boomgaard, R. (2001). Local mode filtering. In *CVPR*.

Xiao, J., Cheng, H., Sawhney, H. S., Rao, C., & Isnardi, M. A. (2006). Bilateral filtering-based optical flow estimation with occlusion detection. In *ECCV*.

Xu, L., Jia, J., & Matsushita, Y. (2012a). Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *34*(9), 1744–1757.

Xu, L., Yan, Q., Xia, Y., & Jia, J. (2012b). Structure extraction from texture via relative total variation. *ACM Transactions on Graphics*, *31*(6), 139.

Yan, Q., Shen, X., Xu, L., Zhuo, S., Zhang, X., Shen, L., & Jia, J. (2013). Cross-field joint image restoration via scale map. In *ICCV*.

Yang, Q. (2012). Recursive bilateral filtering. In *ECCV*.

Yang, Q. (2014). Stereo matching using tree filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(4), 834–846.

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

Zhang, Q., Shen, X., Xu, L., & Jia, J. (2014a). Rolling guidance filter. In *ECCV*.

Zhang, Q., Xu, L., & Jia, J. (2014b). 100+ times faster weighted median filter. In *CVPR*.