Scale-Invariant Image Inpainting Using Gradient-Based Image Composition

Mrinmoy Ghorai[(⊠)], Soumitra Samanta, and Bhabatosh Chanda

Indian Statistical Institute, Kolkata, India
`mgre04@gmail.com`

Abstract. In this paper, we propose a novel scale-invariant image inpainting algorithm that combines several inpainted images obtained from multiple pyramids of different coarsest scales. To achieve this, first we build the pyramids and then we run an image inpainting algorithm individually on each of the pyramids to obtain different inpainted images. Finally, we combine those inpainted images by a gradient based approach to obtain the final inpainted image. The motivation of this approach is to solve the problem of appearing artifacts in traditional single pyramid-based approach since the results depend on the starting scale of the pyramid. Here we assume that most of the inpainted images produced by the pyramids are quite good. However, some of them may have artifacts and these artifacts are eliminated by gradient based image composition. We test the proposed algorithm on a large number of natural images and compare the results with some of the existing methods to demonstrate the efficacy and superiority of the proposed method.

Keywords: Inpainting · Scale-invariant · Multiple pyramid · Image composition

1 Introduction

In image inpainting, the problem is to fill a target (also known as missing or unknown) region by some visual information in order to obtain a visually pleasant and undetectable inpainted image. The information is copied either from the source (also called known) region of the input image or from a database of images similar to the input image. Image inpainting becomes a popular area in the field of image processing and computer vision as it is applied for several application like restoration (scratch removal) and image editing (text or object removal). The most difficult task is object removal since the target region is considered as a large blob. Among the various image inpainting methods, exemplar (patch) based method become most popular to handle the problem of object removal.

In the exemplar (patch) based methods, the texture and structure information is propagated from the source region into the inside of the target region. The concept is to copy the best matched patch from the source region and paste to the selected partially known target patch. This idea was proposed mainly for

texture synthesis [4]. To propagate the structure information, Criminisi et al. [2] first proposed an exemplar-based inpainting method which gives equal importance on the structure filling and the texture synthesis. For this, they compute a priority term depending on isophote strength for the ordering of filling of the target patches. The patch with highest priority is selected for patch inpainting. To improve the method proposed in [2] some other inpainting methods are introduced by changing the priority term and the dissimilarity measure [15,17,19]. These methods usually select a single patch from a set of candidate patches to infer the target patch. Recently, some authors [1,8,9,18] propose a linear combination of candidate patches to infer the target patch. Besides these methods, sparse representation is widely used in image inpainting [5,16,20], where the target patch is inferred by the sparse linear combination of a set of pre-defined patches (dictionary) or candidate patches. In [20], structure sparsity and patch sparse representation are introduced to provide a good solution of priority estimation and patch inference. Further, Li et al. [12] incorporate a Curvelet based multi direction feature in the constraint of sparse representation to maintain structure consistency and texture clarity in the unknown region. To incorporate *coherence* property of natural images some authors [1] suggest to inpaint the target region by the weighted average of the candidates of neighbor patches. Komodakis and Tziritas [10] propose MRF-based image inpainting, namely priority belief propagation (p-BP), by favoring the similarity with the overlapping region of the neighboring patches. Liu and Caselles [13] formulate inpainting as a global energy optimization problem using multiscale graph cuts algorithm. Darabi et al. [3] proposed *image melding* to improve PatchMatch [1] by incorporating geometric and photometric transformations.

Some of the previous methods [1,3,10,13] use pyramid-based technique as the basic framework to capture features in different scales. These methods start from the coarsest level and inpaint each sub-level to reach the finest level. So the quality of the resulted image depends on the inpainted image in the coarsest level. If the coarsest level produce unsatisfactory result then this information is propagated in the finest level, resulting a visually unpleasant inpainted image. Also if we do inpainting at each level of the pyramid individually, it is difficult to say which resolution is perfect for inpainting. For different images the resolution in the coarsest level may not be same to obtain the best result.

To over come this problem, we propose a multiple pyramids based image inpainting algorithm using gradient based image composition. Our main contributions are as follows: (1) build multiple pyramids of the input image with different starting scales; (2) search for candidate patches in different sub levels of the pyramid and run image inpainting method in each pyramid to produce different inpainted images; (3) combine the inpainted images by gradient-based image composition to produce the final inpainted image. Here, we first build multiple pyramids with different resolution of the input image in the coarsest levels of the pyramids, and then obtain an inpainted image in each of the finest levels of the pyramids using traditional pyramid-based approach. So by this way we get multiple inpainted images, one for each of the pyramids. Then we
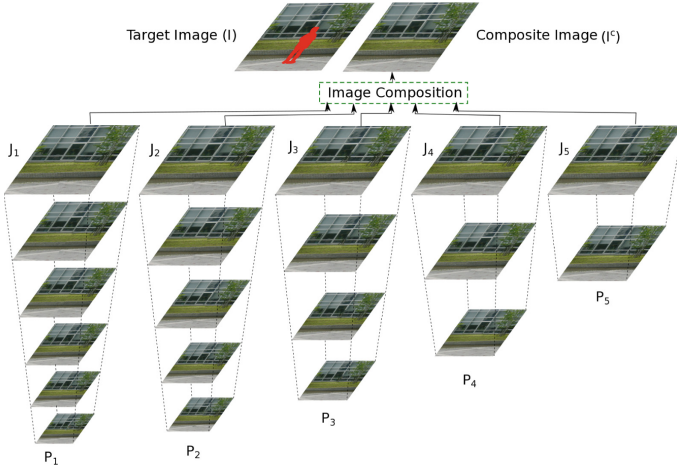
Fig. 1. Overview of the proposed algorithm. $P_1, P_2, \ldots, P_5$ are the pyramids with different resolution of the input image in the coarsest levels. $J_1, J_2, \ldots, J_5$ are the inpainted images obtained from the pyramids $P_1, P_2, \ldots, P_5$ respectively using traditional image inpainting algorithms. I^c is the final inpainted image produced by the image composition of $J_1, J_2, \ldots, J_5$.

combine these inpainted images by gradient-based image composition to produce a final inpainted image. In image composition, our assumption is that most of the inpainted images produced by the pyramids are quite good. However, some of them may have artifacts due to the dependency of inpainting on the resolution of the input image in the coarsest level. Our image composition method based on [23] successfully eliminates these artifacts in the final inpainted image. For image inpainting, we use sparsity-based method [20] in the coarsest level and exemplar-based method [2] in each sub-level of the pyramids. The reason behind the choice of two different algorithms is that the sparsity-based method is able to capture the structure information efficiently compared to the exemplar-based method. But for textural information the exemplar-based method gives better result compared to the sparsity-based methods. Figure 1 shows an overview of the proposed algorithm.

Rest of the paper is organized as follows. The main steps of the proposed method: multiple pyramid-based image inpainting and gradient-based image composition are presented in Sect. 2. In Sect. 3, experimental results and comparisons with some previous methods are explained. The concluding notes and further prospects of the proposed work are given in Sect. 4.

2   Proposed Inpainting Method

The proposed image inpainting method consists of two basic frameworks: multiple pyramid-based image inpainting and gradient-based image composition.

2.1 Multiple Pyramid-Based Inpainting

In this section we describe our proposed image inpainting method based on generating multiple pyramids of the input image. The method consists of several intermediate steps such as multiple pyramids building, searching for candidate patches and sub-level image inpainting. Each of these steps is described elaborately in the following subsections.

Multiple Pyramids Building: Here our goal is to build multiple pyramids for different starting scales of the input image and obtain an inpainted image for each of the pyramids. We choose the blur kernel B to be a Gaussian kernel with a given variance σ_B^2. The value s is an integer scalar factor (e.g., 4 or 5), by which the HR input image I_H is downsampled to the LR image I_L, which to be inpainted first and then upscaled gradually to obtain the HR inpainted image $\hat{I}_H$, i.e. if the input image I_H is of size $X \times Y$, the LR image at the coarsest level will have a size of $\frac{X}{s} \times \frac{Y}{s}$. The LR images at different levels of the pyramid is obtained by

$$I_{-n} = (I_H * B_n) \downarrow_{r^n} \qquad (1)$$

where r is the pyramid cross-level scale factor. The LR image I_{-n} is the downsampled version of the original HR input image I_H with total rescale factor r^n.

So far we have discussed how to generate a pyramid of LR images from a HR input image. Now we want to generate K pyramids $P_1, P_2, \ldots, P_K$ from the HR image I_H with different rescale factors. For this, we define the scale factor r by

$$r_p = r_{p-1} + z \qquad \text{for} \quad p = 2, 3, \ldots, K \qquad (2)$$

where z is the factor which controls the scale of the image in the coarsest level of the pyramids. The value of r_1 is set to 1.35 in our experiments. Typically, the parameter z takes the value 0.03. The pyramid P_1 is generated using the scale factor r_1, similarly the pyramid P_2 is generated using r_2 and so on.

Searching for Candidate Patches: Several searching algorithms [1,3,10] are proposed in literature using traditional pyramid based framework. These methods search the candidate patches in the same sub-level in which the image is inpainted. In natural images, local image structure, which can be captured by small patch, repeats across different scales of an image. So we can use the input image to be inpainted, with up-sampled or down-sampled images, for searching the candidate patches. Previously, these types of multi sub-level search algorithm were proposed for super-resolution [7,21,22]. They had used either one-step search algorithm or a pyramid of recursively scaled images to search the patches.

Similar to one-step approach, we construct a pair of scaled images with the original image for estimating the search region. The reason behind this motivation is that the most relevant patches similar to the target patch can be found
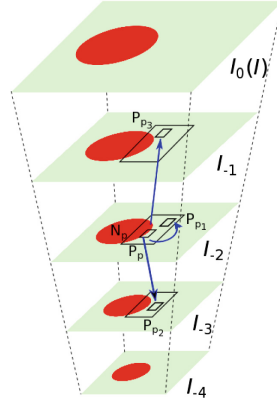
Fig. 2. Overview of the search algorithm. I_{-2} is the current level for image inpainting. For a target patch P_p, we search for the candidate patches in a neighborhood of P_p in I_{-1}, I_{-2} and I_{-3} and select the patches P_{p_3}, P_{p_1} and P_{p_2} respectively.

within the images with small rescaled factor. It is observed that the sufficient amount of good patches could be found by only one rescaling. Let $\mathcal{D}$ denotes an image downscaling operator, s.t. $\mathcal{D}(I) = (I) \downarrow_r$, where r is a suitably chosen small scale factor; let $\mathcal{U}$ denotes an image upscaling operator, s.t. $\mathcal{U}(I) = (I) \uparrow_r$. To inpaint an image I in any sub-level of a pyramid, the image I itself along with $\mathcal{D}(I)$ and $\mathcal{U}(I)$ can be used as sources for searching the candidate patches. However, for the coarsest level, since no downscale image is available, only upscale image is used as source; and on contrary, for the finest level, since no upscale image is available, only downscale image is used as source.

The method, namely, pyramid of recursively scaled images use all the sub-levels as the most relevant source for searching the candidate patches. To avoid the exhaustive search, since it is time consuming, several authors solved it by reducing the search region to a local window [12,13], directional search [6,11], search along user-specified curves [17]. Similar to local search window [12,13], we consider a neighborhood of the target patch in the form of a rectangular region.

In this work, the search region depends on the scale of the pyramid. If in the n-th sub-level the input image is I_{-n} of scale factor r^n, and the search radius is bounded by $\Omega = \min\{r^{-n} \times \kappa, \min(p/2, q/2)\}$ where $p \times q$ is the size of the image in the n-th sub-level, then the radius for one-step search algorithm i.e., for the images I_{-n-1} and I_{-n+1} are Ω/r and $r\Omega$ respectively. Other than this, if we consider Ω for all the three images, then the search range will increase for I_{-n-1} and decrease for I_{-n+1} with respect to the scale of the images. To fix the search range uniformly through out the scale of the images, we have applied varying size of search window with a particular scale factor. Figure 2 shows an overview of the proposed patch searching algorithm.

(a)   (b)   (c)   (d)   (e)   (f)
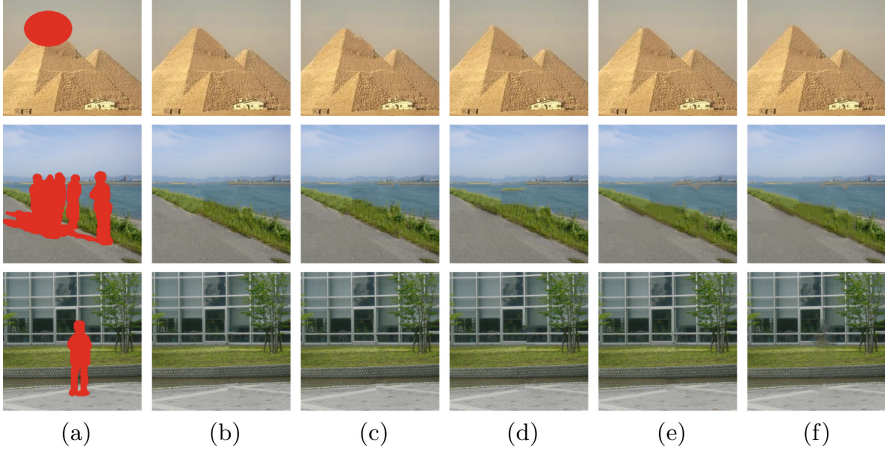
Fig. 3. (a) Target image. (b–f) Inpainted images obtained from 5 pyramids.

Sub-level Image Inpainting: This step proposes a combined approach for image inpainting algorithm applied on the multiple pyramids. Each of the pyramids follow this combined approach to get the inpainted images $J_i = \hat{I}(P_i)$ for $i = 1, 2, \ldots, K$ at the finest level of the pyramids. The updated image at each level of the pyramid is obtained by

$$I_{-n}(T) = \mathcal{U}(\hat{I}_{-n-1}(T)) \tag{3}$$

where $\hat{I}_{-n-1}$ is the LR inpainted image at level $n + 1$ and $I_{-n}(T)$ is the HR image obtained by upscaling (using the upscaling operator $\mathcal{U}$) the LR image $\hat{I}_{-n-1}$. Here T denotes the target region. The LR image at the coarsest level of the pyramid is inpainted using the sparsity-based approach [20].

In practice, we upscaled the LR inpainted image $\hat{I}_{-n-1}$ by Bicubic interpolation and copy the inpainted region (i.e. the inferred target area) of $\mathcal{U}(\hat{I}_{-n-1}(T))$ followed by replacing it to the target region of I_{-n}. In this way we obtain I_{-n} with known target region at the n-th sub-level but this known target region may not be as sharp as the source region of I_{-n} due to the interpolation. To solve this problem we now apply the exemplar-based image inpainting method [2] on I_{-n} considering the source region of I_{-n} as the search region for the candidate patches.

Thus we obtain several inpainted images, one for each of the pyramids, as shown in Fig. 3. The figure shows that artifacts appear in different pyramids such as Fig. 3(c) for the first example near the right structure, Fig. 3(d–f) for the second example in water and grass areas; and Fig. 3(d, f) for the third example in the structure area. It reveals that if we choose those pyramids with a particular starting scale for image inpainting the resulted image is visually unpleasant. This is the main drawback of traditional single pyramid based approach. But, here we also see that most of the inpainted images for a particular example are artifact free and this motivate us to combine them to obtain a better inpainted image.

2.2   Gradient-Based Image Composition

To obtain the final inpainted image, the immediate approaches are either the average or the median operator as given below:

$$\hat{I}^a(x,y) = \frac{1}{K}\sum_{i=1}^{K} J_i(x,y) \tag{4}$$

$$\hat{I}^m(x,y) = \Delta_{i=1}^{K} J_i(x,y) \tag{5}$$

where Δ is a median operator. The main advantage of these approaches is their simplicity with respect to both conceptually and implement wise. Instead of this, they have some drawbacks. The main drawback is the point operation where it only considers the current pixel, and it does not consider the neighboring pixels of the current pixel to update it. Neighborhood processing proves its superiority than point processing to obtain spatially coherent resulted image. Also, the average operator explicitly introduces blur effect in the resulted image, shown in Fig. 4.

To handle these problems, namely blur and spatial consistency of the final image, we introduce a gradient-based image composition for combining the intermediate inpainted images. Here we assume that at a particular position most of the images have similar gradient and some of them may vary due to the appearance of artifact. We want to remove this artifact and recover the texture quality in the final inpainted image. The proposed work is motivated by multiexposure composition [23].

The gradient direction will change in the inpainted images if some artifacts appear in the target region. To get artifact free inpainted image, we capture local inconsistencies based on the changes of gradient direction within a local window. At each pixel (x,y) of the i-th image, its gradient direction change with respect to the j-th image is computed as

$$D_{ij}(x,y) = \frac{\sum_{l=-\omega}^{\omega} |\theta_i(x+l,y+l) - \theta_j(x+l,y+l)|}{(2\omega+1)^2} \tag{6}$$

where the size of the window is $(2\omega+1) \times (2\omega+1)$ and ω is set to 9 in our experiments. Note that, $D_{ij}(x,y) = D_{ji}(x,y)$ and $D_{ii} = 0$ for all i. Since we assume that local inconsistencies may appear in a small number (one or two) of images, a measure T_i for each image is computed to expose its consistency with respect to the others. The measure T_i can be defined as

$$T_i(x,y) = \sum_{j=1}^{K} \exp\left(\frac{-D_{ij}^2(x,y)}{2\sigma^2}\right) \tag{7}$$

where σ is standard deviation and set to .04 in the experiments. A large value of T implies a small gradient changes, that means the information is frequently occurred in the images. The final weight matrix is obtained by

$$W_i(x,y) = \frac{T_i(x,y)}{\sum_{i=1}^{K} T_i(x,y) + \delta} \tag{8}$$

(a)   (b)   (c)   (d)
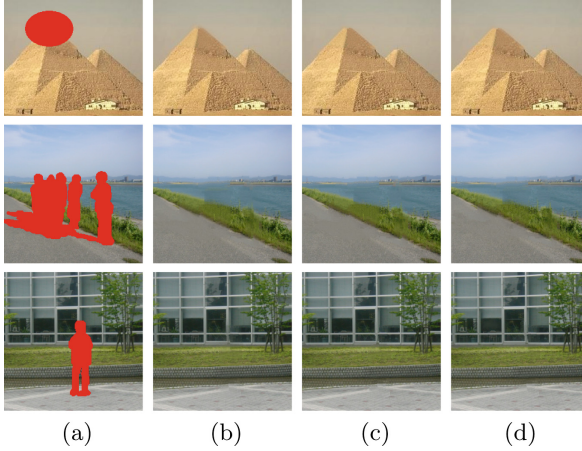
Fig. 4. (a) Target image. (b) Results obtained by averaging all inpainted images from Fig. 3(b–f). (c) Results due to median operator. (d) Results due to image composition.

where δ is a small value such as 10^{-10} to avoid singularity.

So the final inpainted image $\hat{I}^c$ may be expressed as

$$\hat{I}^c(x,y) = \sum_{i=1}^{K} W_i(x,y) J_i(x,y) \qquad (9)$$

For combining all the images seamlessly we apply the multiresolution scheme proposed in [14]. Figure 4 shows some result of image composition applied on 5 different inpainted images from Fig. 3 along with the images obtained by mean and median operators. However, the final inpainted image obtained from the image composition is visually pleasant and almost artifact free.
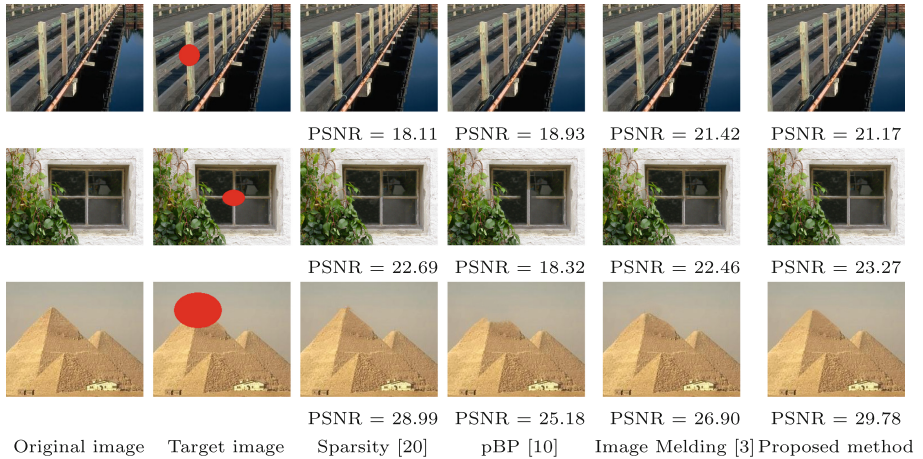
3   Experiments and Results

Here we set the parameters and evaluate the proposed method by qualitative analysis on object removal. For comparison, we take some of the previous state-of-the-art methods based on Xu's sparsity [20], Komodakis's p-BP [10][1], Darabi's image melding [3][2], and Liu's graph cuts [13]. The size of the patch is set to 9 in our experiments. The value of κ (control the patch search region) is set to 40. The number of pyramids K is set to 5. The sparsity-based method [20] and the proposed method are implemented in MATLAB environment. For other methods we have used the auothor's implementation code/results for our comparison. The three examples of Fig. 5 have taken 298, 413 and 613 s respectively to inpaint the target region by our proposed method.

---

[1] http://lafarren.com/image-completer/.
[2] http://www.ece.ucsb.edu/~psen/melding/.

| | | PSNR = 18.11 | PSNR = 18.93 | PSNR = 21.42 | PSNR = 21.17 |
		PSNR = 22.69	PSNR = 18.32	PSNR = 22.46	PSNR = 23.27
		PSNR = 28.99	PSNR = 25.18	PSNR = 26.90	PSNR = 29.78
Original image	Target image	Sparsity [20]	pBP [10]	Image Melding [3]	Proposed method

Fig. 5. Quantitative and qualitative comparison of different inpainting methods for blob removal.

In Fig. 5, we have shown some quantitative comparison for blob type target region. We have considered peak-signal-to-noise ratio (PSNR) as the quantitative measure, computed only over the inpainted region. The PSNR values, computed for individual color channels (R,G,B), are averaged and the average values are given below the corresponding inpainted images. The average PSNR values over all the images are 23.26, 20.81, 23.59, 24.74 respectively due to four said methods. So the quantitative comparison reveals that our method performs better than the other methods.

In Fig. 6, we consider large blob type target region for inpainting. The figure shows the target images and the inpainted images due to sparsity [20], p-BP [10], image melding [3] and the proposed method. The figure clearly shows that the proposed method produces better inpainted images compared to sparsity [20] in the 2nd example (2nd row, 2nd column), image melding in the first example (1st row, 4th column); and the method produces comparable results with p-BP [10]. In the other examples, our method produces either better or comparable results with respect to the images provided by the previous methods.

Figure 7 shows the comparisons of the proposed method with graph cuts [13], sparsity-based [20] and p-BP [10]. The figure shows that the proposed method is surely better than [10, 20], and almost comparable with graph cuts [13]. However, in the first example, the texture clarity of [13] is slightly better than the proposed method.

Figure 8 shows comparison with multi-direction feature (MDF) based method [12] and sparsity-based [20]. Here in both examples our method successfully recover the structure information whereas other methods fail to recover it perfectly. For example, in the first case, the propagation of curve structure by our method is almost free of error. On the other hand, the previous methods produce similar results with wrong propagation of structure.
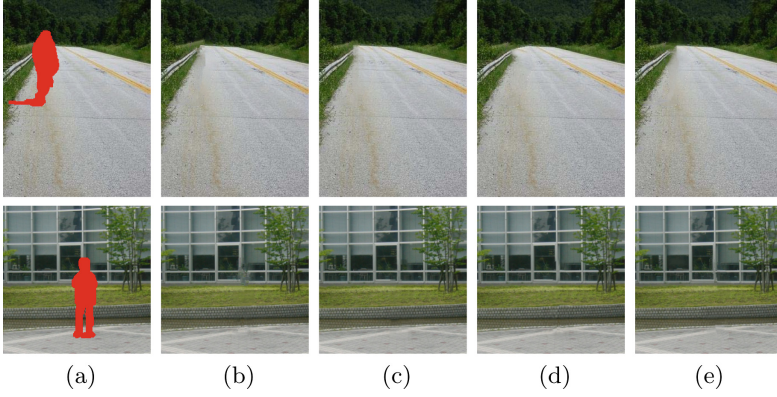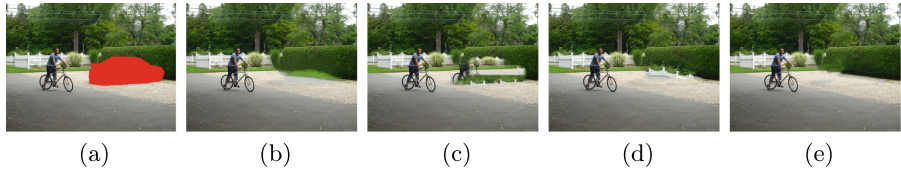
Fig. 6. (a) Target image. (b) Sparsity-based [20]. (c) p-BP [10]. (d) Image melding [3]. (e) Proposed method.
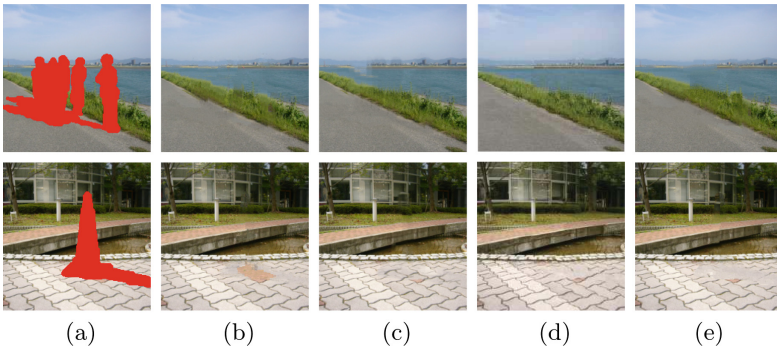
Fig. 7. (a) Target image. (b) Sparsity-based [20]. (c) p-BP [10]. (d) Graph cuts [13]. (e) Proposed method.
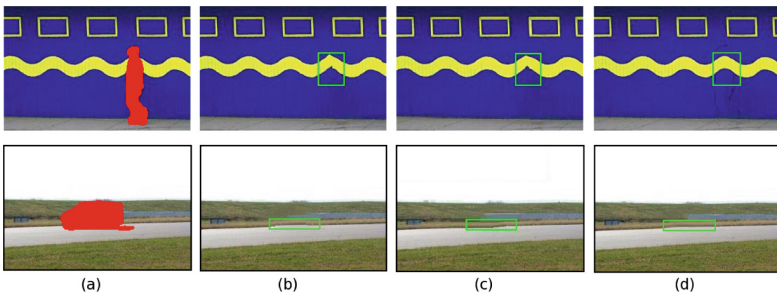
Fig. 8. (a) Target image. (b) Sparsity-based [20]. (c) MDF [12]. (d) Proposed method.

Fig. 9. Failure case. (a) Target image. (b) Sparsity-based [20]. (c) p-BP [10]. (d) Image melding [3]. (e) Proposed method.

Figure 9 shows an example where the proposed method along with competitive methods fail to recover the target region with proper pattern. However, even in this case, the proposed method produces relatively better inpainted image compared to the other approaches. Our result shows that, for random texture, a blurring effect may appear in the texture part of the target region.

4   Conclusion

In this paper we have proposed scale-invariant image inpainting method using gradient-based image composition. The proposed method try to solve the problem of traditional single pyramid based approaches in image inpainting. Technical contribution and experimental results clearly indicate that the proposed multiple pyramid based approach can produce visually satisfactory results compared to the other competitive methods. In future, we will try to incorporate a robust image feature in the proposed image inpainting for recovering more difficult examples of object removal.

References

1. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patch-Match: a randomized correspondence algorithm for structural image editing. ACM Trans. Graph. **28**(3), 1–11 (2009)
2. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based inpainting. IEEE Trans. Image Process. **13**(9), 1200–1212 (2004)
3. Darabi, S., Shechtman, E., Barnes, C., Goldman, D.B., Sen, P.: Image melding: combining inconsistent images using patch-based synthesis. ACM Trans. Graph. (TOG) **31**(4), 82:1–82:10 (2012)
4. Efros, A., Leung, T.: Texture synthesis by non-parametric sampling. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2, pp. 1033–1038 (1999)
5. Fadili, M.J., Starck, J.L., Murtagh, F.: Inpainting and zooming using sparse representations. Comput. J. **52**(1), 64–79 (2009)
6. Fang, C.-W., Lien, J.-J.J.: Rapid image completion system using multiresolution patch-based directional and nondirectional approaches. IEEE Trans. Image Process. **18**(12), 2769–2779 (2009)
7. Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. ACM Trans. Graph. (TOG) **30**(2), 12 (2011)

8. Ghorai, M., Chanda, B.: An image inpainting algorithm using higher order singular value decomposition. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 2867–2872. IEEE (2014)

9. Ghorai, M., Chanda, B.: An image inpainting method using pLSA-based search space estimation. Mach. Vis. Appl. **26**(1), 69–87 (2015)

10. Komodakis, N., Tziritas, G.: Image completion using efficient belief propagation via priority scheduling and dynamic pruning. IEEE Trans. Image Process. **16**(11), 2649–2661 (2007)

11. Le Meur, O., Gautier, J., Guillemot, C.: Examplar-based inpainting based on local geometry. In: 2011 18th IEEE International Conference on Image Processing, pp. 3401–3404. IEEE (2011)

12. Li, Z., He, H., Tai, H.-M., Yin, Z., Chen, F.: Color-direction patch-sparsity-based image inpainting using multidirection features. IEEE Trans. Image Process. **24**(3), 1138–1152 (2015)

13. Liu, Y., Caselles, V.: Exemplar-based image inpainting using multiscale graph cuts. IEEE Trans. Image Process. **22**(5), 1699–1711 (2013)

14. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: a simple and practical alternative to high dynamic range photography. In: Computer Graphics Forum, vol. 28, pp. 161–171. Wiley Online Library (2009)

15. Qin, Y., Wang, F.: A curvature constraint exemplar-based image inpainting. In: International Conference on Image Processing and Signal Processing, pp. 263–267 (2010)

16. Shen, B., Hu, W., Zhang, Y., Zhang, Y.: Image inpainting via sparse representation. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 697–700 (2009)

17. Sun, J., Yuan, L., Jia, J., Shum, H.: Image completion with structure propagation. In: Proceedings of the SIGGRAPH2005, pp. 861–868 (2005)

18. Wohlberg, B.: Inpainting by joint optimization of linear combinations of exemplars. IEEE Signal Process. Lett. **18**(1), 75–78 (2011)

19. Wu, J., Ruan, Q.: Object removal by cross isophotes exemplar-based inpainting. In: Proceedings of the 18th International Conference on Pattern Recognition, vol. 3, pp. 810–813 (2006)

20. Xu, Z., Sun, J.: Image inpainting by patch propagation using patch sparsity. IEEE Trans. Image Process. **19**(5), 1153–1165 (2010)

21. Yang, M.-C., Wang, C.-H., Hu, T.-Y., Wang, Y.-C.F.: Learning context-aware sparse representation for single image super-resolution. In: 2011 18th IEEE International Conference on Image Processing, pp. 1349–1352. IEEE (2011)

22. Zhang, K., Gao, X., Tao, D., Li, X.: Single image super-resolution with multiscale similarity learning. IEEE Trans. Neural Netw. Learn. Syst. **24**(10), 1648–1659 (2013)

23. Zhang, W., Cham, W.-K.: Gradient-directed multiexposure composition. IEEE Trans. Image Process. **21**(4), 2318–2323 (2012)