

Multi-resolution Patch and Window-Based Priority for Digital Image Inpainting Problem

DANG Thanh Trung¹, Chaker LARABI² and Azeddine BEGHDAI¹

¹ Lab. L2TI, Institute Galilee, Université Paris 13, France

e-mail: dang.thanhtrung@univ-paris13.fr, azeddine.beghdadi@univ-paris13.fr

² Lab. XLIM, Department SIC, Université de Poitiers, France

e-mail: chaker.larabi@univ-poitiers.fr

Abstract—Recently, digital image inpainting has attracted strong research interest because of its extensive applications in real life. The terminology "inpainting" refers to automatic restoration of image defects such as scratches or blotches as well as removal of unwanted objects as, for instance, subtitles, logos, etc, such that it is undetectable by viewers without the reference to the original image. Many works on this subject have been published in recent years. This paper introduces a novel unsupervised image completion framework using a modified exemplar-based method in conjunction with a pyramidal representation of an image. A top-down iterative completion is performed gradually with multi-resolution patches and a window-based priority. The proposed approach is verified on different natural images. Also, a comparison with some existing methods coming from literature is carried out and the results show improvement in favor of our approach.

Keywords—inpainting, image completion, restoration, pyramid representation, multi-resolution.

I. INTRODUCTION

Digital image inpainting is a simulation of the activities of experts and professional artists in the museums in restoring the damaged paintings. Their objective is to redraw the damaged regions (such as scratches or blotches) in a visually way based on only their skills and experiences. Currently, it has been extended to digital image processing fields in various applications such as image editing (removing undesired objects, restoring the scratches), film reproduction (deleting the logos, subtitles, etc), or even creating artistic effects (reorganizing objects, smart resizing of images, blending images). The real inpainting problem is extremely difficult because it has to preserve the local smoothness as well as the global content, but without any cue from the original content of the images.

For sake of simplicity, the digital image completion focuses on mathematically modeling and automatically performing the restoration under constraint of local smoothness preservation, but without taking into account the global semantic of image or artistic expertise. In other words, the damaged images are automatically restored such that it still looks natural and undetectable by viewers.

There have been numerous and very different approaches to tackle the inpainting problem. Most of them can be divided into two basic classes [1]: geometry oriented methods and texture oriented methods.

The geometry oriented methods are fundamental. This

approach models the image as a function of smoothness and the restoration is solved by interpolating the geometric information within the adjacent regions into the target region. Several papers have been published in this group such as: partial differential equation (PDE) based [2], total variation [3], diffusion convolution [4], etc. The experimental results indicated that these methods were particularly effective for the synthesis of long, thin regions.

Another class of inpainting algorithms is texture-oriented (or exemplar-based). This approach stems from the texture synthesis techniques where texture is modeled through probability distribution of the pixel brightness values. In [5], the restored regions were completed pixel by pixel based on the matching patch comparison. An extension of this version was presented in [6] where textures were synthesized patch by patch. These approaches are too slow or the input image contains only pure textures. Some sophisticated methods have been introduced such as fragment-based [7], user guidance based [8], probabilistic graphical model based [9][10], geometry-texture hybrid [11][12] etc. However, the computation time of these methods is intolerable and complexity of algorithm is too high. Compared with the first class, the approaches in the second class achieve impressive results in recovering the large damaged regions but the output results may be distorted when the number of patches is insufficient.

Generally, the human visual system is sensitive to high frequency components, so image quality depends highly on these salient features. According to this observation, in [13], authors developed an efficient and simple approach to encourage filling-in from the boundary of the missing region based on the combination of texture synthesis and strength of isophote instead of decomposing the original image as in hybrid methods [11][12] or no-priority texture synthesis [5][6]. This algorithm is a greedy strategy, so the computation time is acceptable but it suffers from the common problems of the greedy algorithms, being filling order (namely priority) is very critical. The isophote-driven priority proposed in [13] generates pleasant outputs in many cases, but it still has many drawbacks. Many modified versions have been developed such as modified priority [14][15][16], non-local mean patch synthesis [17], improvement of patch selection [18], etc. However, the regions with high isophote strength, which may be certain high components such as noise and complex textures, are not

equivalent to the structures e.g edges, contours, etc, therefore, it may lead to undesired results.

In this paper, we propose a novel framework for image completion based on an exemplar-based approach in conjunction with pyramidal representation of the image. The top-down completion is implemented and repeated a fixed number of times. At each level of the pyramid, we use a modified exemplar-based method with multi-resolution patches to restore the candidate regions. Moreover, a novel priority definition, window-based priority, is determined by the local change of patch with neighborhood patches. The experimental results evidence an enhancement in the quality of the output image.

This article is structured as follows: In section 2, more details of our framework are presented. Section 3 is dedicated to the experimental results and comparison with the previously developed methods. The discussion is enclosed to estimate our approach on different natural images. Finally, this paper ends with some conclusions and future works.

II. OUR FRAMEWORK

The exemplar-based methods consist of two main steps: the determination of filling order and the selection of the best matching patch. Our framework is designed to overcome the limitations of the current exemplar-based methods in both two these steps. The proposals are inspired from an observation that the human visual system is sensitive to salient structures which are stable and repetitive at different scales. In other words, one can still see the main structural features when the resolution of the image is reduced to a given factor. Thus, multi-resolution patches are the best way to preserve these features in restoration process. In addition to, a novel priority definition, window-based priority, is introduced to classify more logically image components in multi-scale representation. By applying window-based priority, the decision could be more suitable to maintain the high frequency components in a visually plausible way. By estimating patches at different scales, the selected patch would be more consistent and reasonable to the source patch than one in a single resolution.

For details, some notations that are similar to those in paper [13] are adapted. The whole image domain, I , is composed of two disjoint regions: the inpainting region (or target region), Ω , and the known region, Φ ($\Phi = I - \Omega$). The basic unit of synthesis at pixel p is a patch, Ψ_p , centered at pixel p . Additionally, $\delta\Omega$ represents the set of pixels on the target region boundary (see Fig. 1).

According to above idea, we first use a Gaussian pyramid to generate a set of images G_0, G_1, \dots, G_N with various levels of details, using pyramid operators, i.e., *reduce*, *expand*, etc, where $G_0 = I$ is the input or original image [19]. The inpainting regions are also reduced to the eliminated areas level by level. Then, we apply an exemplar-based method for image completion with some modifications at each level. An iterative process is completed gradually from the lowest

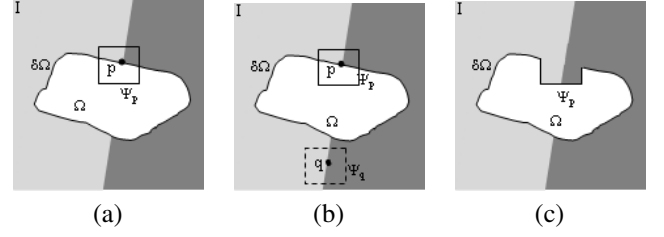


Fig. 1. An exemplar-based method. (a) Estimate the priority; (b) Search the best similar patch and (c) Perform inpainting.

resolution, G_N , to the highest resolution, G_0 . The algorithm is fully described in Table 1. Although different criteria for stopping the process can be used depending on applications, we opted for a fixed number of iterations. All the results shown in this paper were computed with only 2 to 3 iterations.

Table 1. Our framework.

Input : I - input image; N - pyramid levels;
Algorithm :
$G_0 = I$;
$\{G_1, G_2, \dots, G_N\} = \text{buildPyramid}(G_0)$;
Complete G_N with the scheme in [13] using window-based priority in section II-A;
while (stop=false) do
for detail level $i = N - 1$ downto 0 do
Complete G_i with the scheme in [13] using using multi-resolution patch in section II-B and the window-based priority in section II-A;
end for
$\{G_1, G_2, \dots, G_N\} = \text{buildPyramid}(G_0)$;
end while
Output = G_0 ;

A. Window-based priority

Our framework is based a modified version of exemplar-based method, a greedy strategy, so a good priority definition is very important. A confirmed decision based on the priority could not be changed anymore. The error would be accumulated increasingly since there is no opportunity to improve the previous errors.

Many equations have been built for priority, $P(p)$, the most common one is the product of two components: confidence term, $C(p)$, and data term, $D(p)$, as in [13]:

$$P(p) = C(p)D(p) \quad (1)$$

In paper [16], the authors discovered that the confidence term in [13] decreased exponentially and proposed an additive form of priority which used weights to maintain a balance between the confidence term and the data term.

$$RP(p) = \alpha((1 - \omega)C(p) + \omega) + \beta D(p) \quad (2)$$

where $\omega \in [0, 1]$ is the regularizing factor for controlling the curve smoothness (in experiment, ω is often set to 0.7), the value range of $C(p)$ is regularized to $[\omega, 1]$, and α, β are positive parameters and $\alpha + \beta = 1$.

In these methods, the data term, $D(p)$, depends proportionally on the isophote direction or gradient of the known region. Thus, if the gradient at pixel p is large, the priority will be

high. In other words, when the gradient values of the texture components are greater than those of structure components or the regions are affected by noise components, the isophote-driven priority method may violate the requirement of a good priority and distort the results. The data term based on cross-isophote direction in [14] also suffers the same drawbacks or other one using color distribution analysis in [15] takes an expensive calculation.

Reasonably, the data term should be estimated from all neighborhood pixels in an exemplar centered the current pixel, because the human vision is good at seeing the group of adjacent pixels rather than single pixel. Our first improvement focuses on a better definition of data term based on the local changes of pixel intensities in each window, W_p , centered at pixel p with sifted windows in different directions. The local change of intensity at each pixel $p(x, y)$ is characterized by the following matrix [20]:

$$M(p) = \sum_{W_p} G_{W_p}(x, y) \begin{pmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \left(\frac{\partial I}{\partial y}\right)^2 \end{pmatrix} \quad (3)$$

where G_{W_p} is a Gaussian window function computing a weight sum. This is a 2×2 symmetric and semi-positive matrix which captures the intensity structure of the local neighborhood. There exist two positive eigenvalues, $\lambda_1 \geq \lambda_2 \geq 0$. Two corresponding eigenvectors represent two orthogonal directions directed along the local maximum and minimum variation of image intensities and the eigenvalues measure the effective variations (strength of contours) of image intensities along these vectors. Therefore, our data term is defined as the follows:

$$D(p) = \frac{\lambda_1}{\lambda_2 + \epsilon} \quad (4)$$

where ϵ is a very small positive value which ensures that the denominator term always is non-zero (in our experiment, $\epsilon = 10^{-10}$). This data term implies not only the geometry features such as contours or edges but also the texture features. There are three cases to be considered for each window:

- If the data term is much greater than one i.e. one eigenvalue is high (λ_1 is high) and the other is low (λ_2 is low), the local shifts in one direction cause little change and significant change in the orthogonal direction; the window contains strong edges;
- If the data term is close to one, i.e λ_1 is approximate to λ_2 , there are two possible cases:
 - If both eigenvalues are high, the shift in any direction will result in a significant change, this indicates a texture or complex structures.
 - If both eigenvalues are small, the shift in any direction will cause a little change, the patch is of approximately constant intensity (flat region).

Because both confidence and data terms are evaluated by contribution of all pixels in a window, our priority is called a window-based priority. With this priority, the patches would be classified in a more robust way and the computation

time is acceptable. Table 2 illustrates some numerical values corresponding to subimages in Fig. 2. As our opinion, the contours should be preserved, the priority of pixels on the contour should be higher than others. Obviously, the values in Table 2 evidence that our priority is more reasonable than one in [13][14] for image completion.

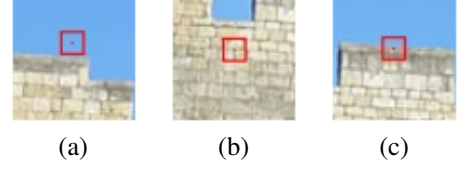


Fig. 2. The window-based priority of different regions. (a) Flat region. (b) Texture region. (c) Edge or contour.

Table 2. Priority of the different regions.

Patch	Priority in [13]	Priority in [14]	Proposed Priority (λ_1, λ_2)
Flat region (Fig. 2a)	0.001	3.269	0 (0, 0)
Texture region (Fig. 2b)	0.032	3.916	1.374 (0.067, 0.049)
Edge or contour (Fig. 2c)	0.024	3.130	4.059 (0.025, 0.006)

B. Multi-resolution patch

The second advantage of our method is the use of a multi-resolution patch-based scheme on the image pyramid [19]. A multi-resolution patch, Ψ_p , consists of two scales: one at the current level l , Ψ_p^l , and one at its parent level, Ψ_p^{l+1} . Figure 3 illustrates an example of multi-resolution patch Ψ_p , that contains various single resolution patches with different sizes. The sizes of patches are global parameters and denoted as $\{(w_1 \times w_1); (w_2 \times w_2)\}$, where the former is for current level and the later is for the higher level and w_1 is often greater than w_2 because of the image resolution. The similarity between two multi-resolution patches is measured by the sum of the squared distance of all known pixels:

$$d(\Psi_p, \Psi_q) = d_1(\Psi_p^l, \Psi_q^l) + d_2(\Psi_p^{l+1}, \Psi_q^{l+1}) \quad (5)$$

where d_1, d_2 are similarity of the known pixels in the current and higher levels, respectively.

A good similarity measure needs to agree perceptually with observers. The similarity measurement based on only color channels is insufficient to propagate accurate linear structures into the target region and leads to garbage growing. This comes from the observation that the human eye is sensitive to not only color intensity but also color intensity variation (color gradient). To maintain this variation, we add to the similarity measurement a new term G representing image gradient as an additional weight. The similarity function now depends on the difference between the patches according to two criteria, the difference in color channels and in gradients.

$$d_t(\Psi_p, \Psi_q) = \sum_i ((I_p^i - I_q^i)^2 + (G_p^i - G_q^i)^2) \quad (6)$$

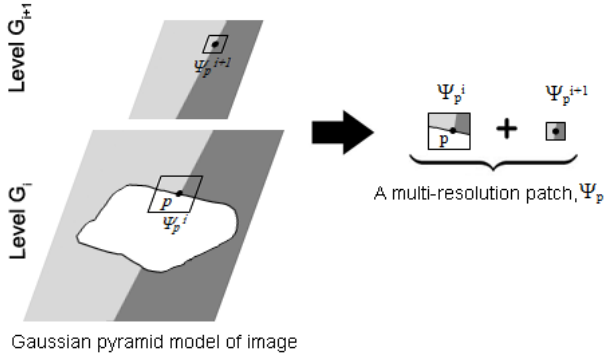


Fig. 3. A multi-resolution patch composed of pixels on different scales

where I_p, I_q are the corresponding RGB vectors in patches Ψ_p, Ψ_q ; G_p, G_q represent the image gradient vectors; $t \in \{1, 2\}$ corresponds to the current and higher levels in the pyramid. The patch with minimal distance to the source patch, Ψ_p , is the chosen one and described as follows:

$$\Psi_{\hat{p}} = \underset{\Psi_q \in \Phi}{\operatorname{argmin}} \{d(\Psi_p, \Psi_q)\} \quad (7)$$

The missing pixels are copied from the corresponding pixels in the selected patch. The image is produced by taking into account the patches structures from the low frequency to the high frequency.

C. Patch selection

Most of the published methods do not deal with the case when there are more than two patches with the same degree of similarity. The obtained results depend highly on the chosen patch. In [18], the authors proposed an approach based on comparison between the variance of known pixel values in the patches. But this approach is inefficient in many cases. Indeed, the local variance measure is not a very discriminative measure (Fig. 4). Here, we propose a solution based on the

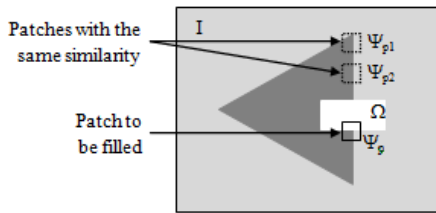


Fig. 4. Patches with the same similarity may generate undesired results.

standard deviation of variances of neighboring source patches. For each patch Ψ_p , it can be formulated as follows:

$$V(\Psi_p) = \sum_{E \in \{R, G, B\}} \sqrt{\frac{\sum (E(\Psi_p) - \bar{E}(\Psi_{N(p)}))^2}{|N(p)|}} \quad (8)$$

where $N(p)$ is a set of neighborhood centered at p , $E(\Psi_p)$ is the variance of pixel values at the neighboring patch Ψ_p in R, G, B channels and $\bar{E}(\Psi_p)$ is mean variances of $|N(p)|$

neighboring patches in R, G, B channels. The size of $N(p)$ is a global parameter, it should be chosen larger than size of patch. The chosen patch is satisfied the following equation:

$$\Psi_{\hat{p}} = \underset{\Psi_q \in \Phi}{\operatorname{argmin}} \{|V(\Psi_p) - V(\Psi_q)|\} \quad (9)$$

III. EXPERIMENTAL RESULTS

In this section, the performance of proposed image completion framework was evaluated on series of different natural images. Also, a comparison with some existing exemplar-based methods is performed. In most experiments, the size of window priority is set to (5×5) . The patch sizes are set to $\{(9 \times 9); (5 \times 5)\}$ for our approach and (9×9) for other exemplar-based methods. The missing regions are filled by the green color (R=0, G=255, B=0). We selected the methods



Fig. 5. Comparison with existing exemplar-based methods. The first row shows the inpainting images. The remaining rows show the results of methods in [13], [14], [15], [16] and our proposal, respectively.

in [13], [14], [15], [16] for comparison. In the figure 5, the first row shows the inpainting images where the damaged areas occupy 10.3% (in the left image), 9.29% (in the center image) and 14.8% (in the right image) of the total image. The corresponding results of [13], [14], [15], [16] and our proposal are listed in the remaining rows, respectively. These tests are suitable for applications such as removal of large objects or blotches. As we can observe from these figures, our method produces a better visual quality than the other methods. We should mention that for performance evaluation, we use only subjective criteria, i.e., visual appearance of the output because there is no efficient numerical estimation for outputs of inpainting methods. Figure 6 illustrates an example

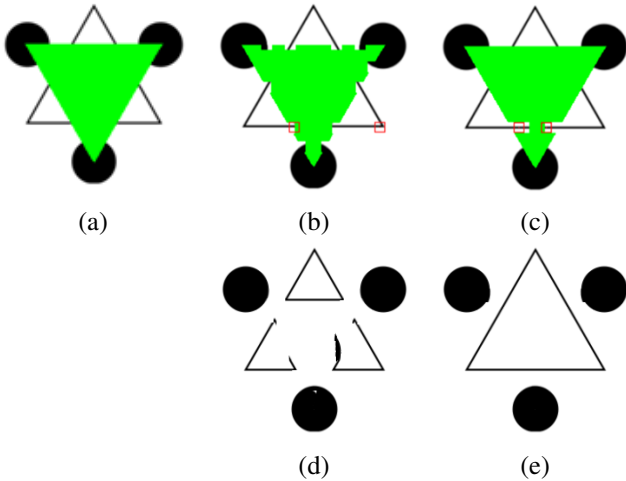


Fig. 6. A restoration of Kanizsa triangle. (a) inpainting image; (b) a patch selection by [18]; (c) our proposed selection; (d) the result of approach in [18]; (e) the result of our proposal

for patch selection, a kanizsa triangle. The first row represents the inpainting image (Fig. 6a) and a selection step in inpainting process of approach in [18] (Fig. 6b) and our approach (Fig. 6c), respectively. An unsuitable selection may generate unexpected results. The corresponding results are shown on the second row: Fig. 6d for approach in [18] and Fig. 6e for our approach. By using standard deviations of the neighboring patches, the selection in Fig. 6c is more consistent than one in Fig. 6b. In fact, there is no evidence to confirm which result is more accurate in this case but these images show a clearly visible effect of our proposal in patch selection.

IV. CONCLUSIONS

In this paper, a novel framework of image completion is introduced using multi-resolution representation of image. The multi-resolution patch ensures the texture and geometric structure with the fixed size patches instead of using different patches or dynamic size patches. The chosen patches are more consistent and adaptive. A common problem to most exemplar-based methods is the effect of priority. In currently presented version, a window-based priority which classifies patches in a more suitable way and leads to the pleasant results is introduced. Moreover, the adaptive patch selection in cases

where there are more two patches with the same similarity to source patch is introduced to limit the undesired results. This improvement can be applied to most exemplar-based methods. Comparison with some new and old methods of the second group is carried out and the results show that our approach produces better quality of output images. For future works, a perceptual patch similarity that are more stable and effective and a numerical estimation for quality of the restored image are being further studied.

REFERENCES

- [1] P. Arias, G. Facciolo, V. Caselles, G. Sapiro. A Variational Framework for Exemplar-Based Image Inpainting. In *International Journal of Computer Vision*, pp. 1-29, 2011.
- [2] T.F. Chan and J. Shen. Non-texture inpainting by Curvature-Driven Diffusions (CCD). In *Journal of Visual Communication and Image Representation*, vol 4. pp. 436 449, 2001.
- [3] P. Perez, M. Gangnet, and A. Blake. Poisson image editing. In *Transaction Graphics. ACM*, 2003, vol. 22, pp. 313 318.
- [4] D. Tschumperle. Fast anisotropic smoothing of multi-valued images using curvature-preserving pdes. In *International Journal of Computer Vision*, 2006, vol. 68, pp. 6582.
- [5] A. Efros and T. Leung. Texture synthesis by non parametric sampling. In *The Proceeding of the Seventh IEEE International Conference on. IEEE*, 1999, vol. 2, pp. 10331038.
- [6] A. Efros and W.T. Freeman. Image quilting for texture synthesis and transfer. In *Proceeding of the 28th Annual Conference on Computer Graphics and Interactive Techniques. IEEE*, 2001, pp. 341 346.
- [7] J. Jia and C.K.Tang. Image repairing: Robust image synthesis by adaptive nd tensor voting. In *Proceeding 2003 IEEE Computer Society Conference on Computer Vision Pattern Recognition*, 2003, pp. 643-650.
- [8] J. Sun, L. Yuan, J. Jia, and H-Y. Shum. Image completion with structure propagation. In *Proceeding of ACM SIGGRAPH*, 2005, vol. 24, pp. 861 868.
- [9] G.T. N. Komodakis and G. Tziritas. Image completion using global optimization. In *Proceeding of IEEE Computer Society Conference Computer Vision and Pattern Recognition*, pp. 442452, 2006.
- [10] G.T.N. Komodakis. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Transactions Image Process*, vol. 16, pp. 26492661, 2007.
- [11] A. Bugeau and M. Bertalmio. Combining texture synthesis and diffusion for image inpainting. In *Proceeding International Conference Computer Vision Theory and Applications*, 2009, pp. 2633.
- [12] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. In *Proceeding of Conference Computer Vision Pattern Recognition*, 2003, vol. 2, p. 707.
- [13] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. In *IEEE Transaction of Image Process. IEEE*, 2004, vol. 13 (9), pp. 12001212.
- [14] J. Wu and Q. Ruan. Object removal by cross isophotes exemplar based image inpainting. In *Proceeding of International Conference of Pattern Recognition*, 2006, pp. 810 813.
- [15] Q. Zhang and J. Lin. Exemplar-based image inpainting using color distribution analysis. In *Journal of Information Science and Engineering*, 2011.
- [16] W. Cheng, C. Hsieh, S. Lin, C. Wang, and J. Wu. Robust algorithm for exemplar-based image inpainting. In *Proceeding of International Conference on Computer Graphics, Imaging and Visualization. IEEE*, 2005, pp. 64 69.
- [17] A. Wong and J. Orchard. A nonlocal-means approach to exemplar-based inpainting. In *15th IEEE International Conference on, ICIP 2008, Image Processing*, pp. 2600 2603, 2008.
- [18] Anupam, P. Goyal, and S. Diwakar. Fast and enhanced algorithm for exemplar based image inpainting. In *Image and Video Technology PSIVT*, 2010, pp. 325 330.
- [19] J.B. Peter and H.A. Edward. The laplacian pyramid as a compact image code In *IEEE Transactions On Communications. IEEE*, 1983, vol. 31, pp. 532 540.
- [20] C. Harris and M.J. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pp. 147152, 1988.