

EXTREME IMAGE COMPLETION

Radhakrishna Achanta, Nikolaos Arvanitopoulos, Sabine Süsstrunk

School of Computer and Communication Sciences
École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

It is challenging to complete an image whose 99% pixels are randomly missing. We present a solution to this extreme image completion problem. As opposed to existing techniques, our solution has a computational complexity that is linear in the number of pixels of the full image and is real-time in practice. For comparable quality of reconstruction, our algorithm is thus almost 2 to 5 orders of magnitude faster than existing techniques.

Index Terms— Image completion, real-time, linear complexity, parameterless.

1. INTRODUCTION

If 99% of the pixels in an image are randomly removed, it is almost impossible for a human to interpret the image (Fig. 1). However, due to redundancies of natural images, we can reconstruct the missing pixels using techniques such as interpolation and inpainting [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. These completed images, even if not at the same visual quality as the original, are easily interpretable by a human or a machine vision system.

The drawback of most algorithms in the literature that can achieve such image completion is that they are computationally quite expensive. In this paper, we present a parameterless algorithm we call *Efficient Filtering by Adaptive Normalization* (EFAN) that is capable of completing images in real-time. The complexity of our algorithm is linear in the number of pixels contained in the final image and is independent of the number of missing (or available) pixels.

The rest of the paper is organized as follows. After summarizing the relevant state-of-the-art of image completion, we present a computationally efficient method for image completion in extreme cases using a low-pass filtering technique. We then present important modifications to this algorithm, which improve both the computational efficiency, resulting in a real-time algorithm, as well as the quality of the output. We then compare our algorithms with the state-of-the-art and show that for comparable visual quality and measurable quality metrics (MSE, PSNR, SSIM), we obtain 2-5 orders of magnitude speed-up.

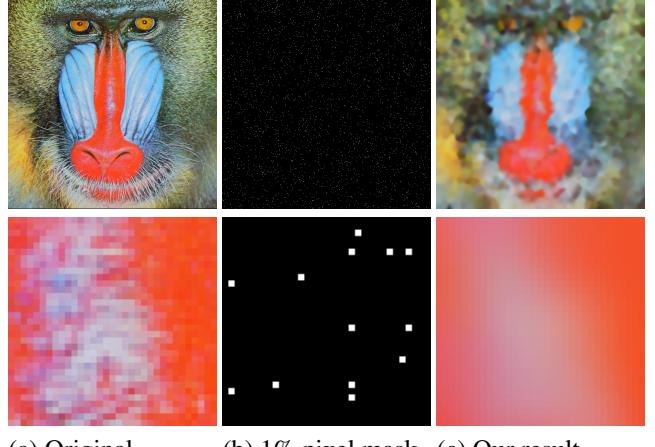


Fig. 1. Extreme image completion using our real-time algorithm. Top row shows original images of size 512×512 and bottom row shows a zoomed-in central region of size 32×32 pixels. Along each column, (a) is the original image, (b) is the mask of 1% randomly chosen pixels, and (c) is the result of completion using our algorithm. In the absence of the original image, it is very hard to interpret the image from only 1% retained pixels. However, it is still possible to complete the image.

2. RELATED WORK

The literature related to image completion is vast since it pertains to a wide range of applications such as image-scaling, super-resolution, noise-removal, demosaicing, artifact removal, and image editing. We briefly cover the literature that addresses randomly-sampled image completion as it is closely related to our problem. We leave out hole-filling or inpainting solutions unless they can also address extreme image completion.

To complete a randomly sampled image, existing methods fall into two broad categories - those that use local pixel information, and those that use global information from the entire image to exploit image self-similarities. Some of these methods were presented as hole-filling or inpainting solutions but can also do extreme image completion. Local methods are usually computationally simpler than global methods but global methods usually provide higher quality.

A total variation (TV) approach like the one used for noise removal [1, 2, 11] can be used for extreme image completion using local pixel-level information. Such methods tend to blur contours due to the nature of the TV regularizer.

Partial differential equation (PDE) based techniques [3, 4, 12, 13] also use local pixel information for image completion. PDE-based schemes generate high quality edge-preserving outputs. Despite the complexity of these techniques, they can be implemented in an efficient manner using coarse-to-fine strategies [13].

Global methods of inpainting use image patches to take advantage of image self-similarity [7, 8, 9]. Patch-based methods may not always be suitable for extreme image completion because there are almost no connected pixels to obtain patches from. As an exception, Facciolo et al. [10] present a method for interpolating sparsely sampled images, which uses a variational formulation to transfer information between similar image patches. This method performs high quality image completion, like most global methods, but has a prohibitive computational cost.

Another category of global methods relies on matrix completion [14, 15, 16]. However, these techniques are computationally intensive and require a large number of retained pixels (30% or more). The approach we present in this paper relies on local pixel information. It is suitable for completing images with very small numbers of retained pixels and delivers real-time performance.

3. OUR APPROACH

We present a computationally efficient approach to complete a 1% pixel image, and then we present an improvement that is even faster. For the first method we take a low-pass filtering approach using a Gaussian filter.

In an image containing all its pixels, the use of a separable Finite Impulse Response (FIR) kernel in the time domain is easy. In such a case, the result of the convolution of an image row (or column) f at each position n by a Gaussian filter h of finite support in the set $\{-M, -M + 1, \dots, M - 1, M\}$ is given by:

$$(f * h)[n] = \frac{1}{W} \sum_{m=-M}^M f[n-m]h[m] \quad (1)$$

where

$$W = \sum_{m=-M}^M h[m] \quad (2)$$

is the sum of the filter coefficients, which is unity when the filter coefficients are normalized.

However, in an image where most pixels are randomly missing, when we perform such a convolution (assuming, say, zeros for the missing pixel values), the value of W in Eq. 1

adds up coefficients of missing pixels as well, resulting in incorrect values for the filtered output. To use an FIR filter as above, only the available pixels should be used, and W should be accumulated for only those pixels, requiring a search operation at each pixel position. The need to search prevents the use of computationally efficient ways of performing convolution in the frequency domain or by the use of Infinite Impulse Response (IIR) filters.

To overcome this problem we create another signal g containing ones and zeros of similar length as f such that:

$$g[n] = \begin{cases} 1, & \text{if } \exists f[n] \\ 0, & \text{otherwise} \end{cases}$$

Now, instead of searching for available values in f , we assume zeros for the missing values, we convolve both f and the derived signal g with the same kernel h using any technique, and obtain the desired filtered output F by performing an element-wise division of the two convolved signals as:

$$F[n] = \frac{(f * h)[n]}{(g * h)[n]} \quad (3)$$

This works because the denominator signal is affected in the same way as the numerator by zero values and therefore adaptively compensates for it. This scheme is easily extended to two or higher dimensions. Such an approach has been used in the past for linearizing the bilateral filter using the bilateral grid [17].

Another issue is to choose the standard deviation σ of the Gaussian filter. If K pixels are uniformly randomly retained, the average number of unknown pixels closest to each known pixel is N/K . The value of σ is therefore chosen to be the radius of a circle covering this area, i.e $\sigma = \sqrt{N/\pi K}$, which proves to be a robust heuristic in practice.

The result of filtering the entire image using the approach of Eq. 3 can be seen in Fig. 2e. We refer to this method as the *Filtering by Adaptive Normalization* (FAN) method since the weight normalization is adapted to the presence or absence of a value without the need for a search.

We now revisit the FAN approach and notice that we can obtain the same result if at each *unknown* pixel position i , we sum up the Gaussian-weighted contribution of each *known* pixel value at position k . That is,

$$J[i] = \frac{\sum_{k=1}^K w(i, k) I[k]}{\sum_{k=1}^K w(i, k)} \quad (4)$$

where $w(i, k) = e^{-0.5(i-k)^2/\sigma^2}$ is the Gaussian weight and J is the completed output. With this observation we can simplify the image completion process of FAN.

Firstly, we only compute the contributions known pixels can make at the unknown pixel locations. We rely on two maps G and D (initialized to zero) of the same dimensions as the input image. G and D accumulate, respectively, the

Gaussian weighted contribution and the Gaussian weight of each *known* pixel value at each *unknown* pixel location in the image plane. In the same way as FAN, we compute the completed image by performing an element-wise division of G and D . Notice that we are not doing a convolution operation any more on the entire image. Also notice that there is no signal of ones and zeros that needs to be filtered for the sake of normalization.

Secondly, we note that beyond the 3σ distance, Gaussian weights are negligible. We take advantage of it to restrict the computation within a $6\sigma \times 6\sigma$ window S centered at the *known* pixels. Without this restriction, for K known pixels the complexity is $O(NK)$ (as in Eq. 4). By limiting the computation to window S , whose size is proportional to σ^2 , the complexity becomes $O(\sigma^2 K)$ i.e., $O(N)$. In other words, the computation is linear in N and independent of K . We call this method *Efficient Filtering by Adaptive Normalization* (EFAN) since it is more than twice as fast as FAN (Fig. 2).

Algorithm 1 The EFAN algorithm.

Require: Single channel input image I containing K values, list L of indices of the K values, weighted pixel value accumulation map G , weight accumulation map D , output image J to contain the completed image.

```

1:  $\sigma = \sqrt{N/\pi K}$ 
2: Initialize  $G$  and  $D$  to zero.
3: for  $k \in L$  do
4:   Define window  $S$  centered at  $k$ 
5:   for  $i \in S$  do
6:      $w(i, k) = e^{-0.5(i-k)^2/\sigma^2}$ 
7:      $D[i] = D[i] + w(i, k)$ 
8:      $G[i] = G[i] + w(i, k)I[k]$ 
9:   end for
10: end for
11: for  $i \in I$  do
12:    $J[i] = \frac{G[i]}{D[i]}$ 
13: end for
```

These two steps of using only the known pixel values, and restricting the computation to a small window, drastically reduce the computation as well as provide $O(N)$ complexity for any number of missing pixels. Effectively, we turn the FAN solution around from performing a convolution operation at each *unknown* pixel location to only performing the computation in restricted square window regions around the *known* pixels. Algorithm 1 lays out the steps for a practical implementation¹ of EFAN. Color images are processed by simply applying Algorithm 1 to each color channel separately. It is worth noting that our algorithm easily extends to higher-dimensional data like videos or image stacks.

Even though the two solutions FAN and EFAN are equivalent, in practice EFAN outputs exhibit lower errors (see

Fig. 2). This is because efficient convolution performed using IIR or frequency-domain filtering may lead to minor artifacts at the image borders.

4. COMPARISONS

In order to compare our methods (FAN and EFAN) with the state-of-the-art, we choose two local approaches and one global approach. The local approaches chosen are Total Variation based (TV)² and PDE-based (PDE)³, while the global approach is the exemplar-based (Exemplar)⁴ of Facciolo et al. [10]. We compare our results with theirs visually as well as in terms of MSE, PSNR, SSIM, and speed in Fig. 2.

We also compare MSE values of EFAN against the best competing method of PDE for extreme image completion on the 30 images (of size 512×512 pixels) of the Categorical Subjective Image Quality (CSIQ) database [18]. Our method EFAN shows lower MSE values than PDE on all but one images (Fig. 3).

In terms of signal reproduction error, our method EFAN is vastly superior to TV, and has usually better completion quality than PDE (Fig. 2 and Fig. 3). Our method EFAN is almost two orders of magnitude faster than PDE, its closest competitor in quality. In addition, EFAN is about five orders of magnitude faster than Exemplar. As a note, despite the poor scores, visually, the quality of Exemplar is high, as is to be expected of a global method. The poor-scores are due to the noisy output (visible upon zooming Fig. 2d).

5. CONCLUSION

We introduce a parameter-free real-time solution for the extreme image completion problem of reconstructing an image using only 1% randomly retained pixels. Our algorithm is about 2 orders of magnitude faster than a comparable quality state-of-the-art pixel-based technique, and about 5 orders of magnitude faster than an exemplar-based technique. Such a solution can lead to interesting practical applications.

In the case of video transmission, for instance, if energy consumption is an issue during capture and transmission, extreme images can be transmitted. In such a case, only the pixel values need be sent. There is no need to send the pixel positions as long as the receiver is aware of the image dimensions, and uses the same random seed and generator. As a side effect, such transmission offers privacy protection since it is hard to put together the original image without knowing the pixel locations. The video can be recreated very efficiently at the receiving end using our method. A potential scenario for transmitting extreme images is basic motion or obstacle detection using low-detail videos. We will explore such applications in the future.

²Implemented by us.

³Using Matlab's *regionfill* function.

⁴Publicly available code: <http://gpi.upf.edu/static/vnli/>

¹http://ivrl.epfl.ch/research/image_completion

(a) Original	(b) TV	(c) PDE	(d) Exemplar	(e) FAN	(f) EFAN
MSE	1288.32	481.80	542.48	503.55	461.55
PSNR (dB)	17.03	21.30	20.79	21.11	21.49
SSIM	0.86	0.92	0.88	0.91	0.92
MSE	1694.05	665.54	1098.15	679.71	620.92
PSNR (dB)	15.84	19.90	17.72	19.81	20.20
SSIM	0.77	0.87	0.80	0.88	0.88
MSE	1692.16	1063.90	1595.77	1140.08	1127.57
PSNR (dB)	15.85	17.86	16.10	17.56	17.61
SSIM	0.38	0.52	0.44	0.51	0.52
Avg. speed (s)	1200+	2.64	7200+	0.07	0.03

Fig. 2. Visual comparison of extreme image completion - images are recreated using only 1% randomly selected original pixels (using the same seed for random number generation). In terms of signal reproduction quality and computational efficiency, the proposed methods of FAN and EFAN (columns e and f) outperform the state-of-the-art. All images are of size 512×512 . Images are best viewed on display screen by zooming-in.

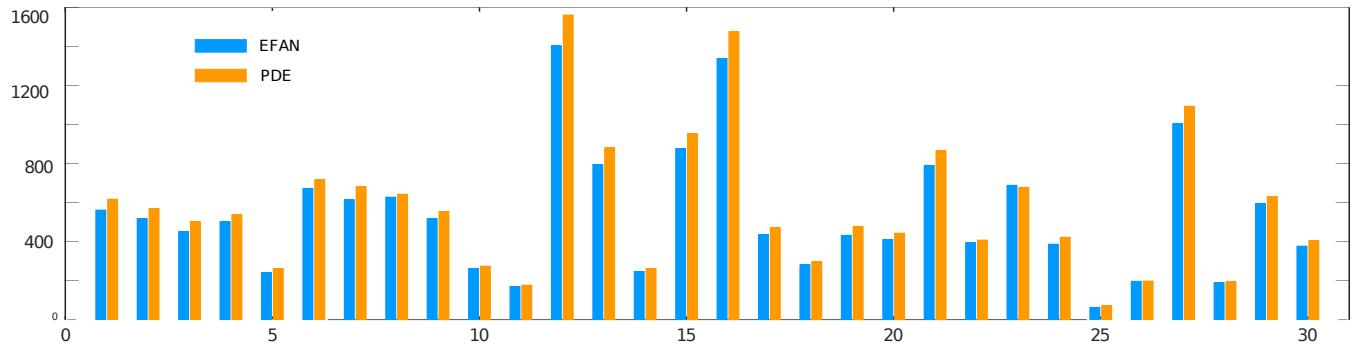


Fig. 3. Comparison of MSE values of our method EFAN against its closest competitor PDE on the 30 images of the CSIQ database [18]. EFAN shows lower MSE values in 29 of the 30 images. Similarly, PSNR and SSIM values of EFAN are higher than PDE in 29 and 24 images, respectively.

6. REFERENCES

- [1] S. Masnou and J. M. Morel, “Level lines based disocclusion,” in *International Conference on Image Processing (ICIP)*, Oct 1998, pp. 259–263 vol.3.
- [2] Pascal Getreuer, “Total Variation Inpainting using Split Bregman,” *Image Processing On Line*, vol. 2, pp. 147–157, 2012.
- [3] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester, “Image inpainting,” in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 2000, SIGGRAPH’00, pp. 417–424.
- [4] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, “Navier-stokes, fluid dynamics, and image and video inpainting,” in *Computer Vision and Pattern Recognition (CVPR). Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I–355–I–362 vol.1.
- [5] C.B. Schönlieb, *Modern PDE Techniques for Image Inpainting*, University of Cambridge, 2009.
- [6] A. Bourquard and M. Unser, “Anisotropic interpolation of sparse generalized image samples,” *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 459–472, Feb 2013.
- [7] A. Levin, A. Zomet, and Y. Weiss, “Learning how to inpaint from global image statistics,” in *IEEE International Conference on Computer Vision (CVPR)*, Oct 2003, pp. 305–312 vol.1.
- [8] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, Sept. 2004.
- [9] Jian Sun, Lu Yuan, Jiaya Jia, and Heung-Yeung Shum, “Image completion with structure propagation,” *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 861–868, July 2005.
- [10] Gabriele Facciolo, Pablo Arias, Vicent Caselles, and Guillermo Sapiro, *Exemplar-Based Interpolation of Sparsely Sampled Images*, pp. 331–344, Berlin, Heidelberg, 2009.
- [11] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.
- [12] Irena Galić, Joachim Weickert, Martin Welk, Andrés Bruhn, Alexander Belyaev, and Hans-Peter Seidel, “Image compression with anisotropic diffusion,” *Journal of Mathematical Imaging and Vision*, vol. 31, no. 2, pp. 255–269, 2008.
- [13] Sven Grewenig, Joachim Weickert, and Andrés Bruhn, “From box filtering to fast explicit diffusion,” in *DAGM Conference on Pattern Recognition*, Berlin, Heidelberg, 2010, pp. 533–542, Springer-Verlag.
- [14] Yao Hu, Debing Zhang, Jieping Ye, Xuelong Li, and Xiaofei He, “Fast and accurate matrix completion via truncated nuclear norm regularization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2117–2130, 2013.
- [15] Wei Li, Lei Zhao, Zhijie Lin, Duanqing Xu, and Dongming Lu, “Non-local image inpainting using low-rank matrix completion,” *Computer Graphics Forum*, vol. 34, no. 6, pp. 111–122, 2015.
- [16] Qing Liu, Zhihui Lai, Zongwei Zhou, Fangjun Kuang, and Zhong Jin, “A truncated nuclear norm regularization method based on weighted residual error for matrix completion,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 316–330, 2016.
- [17] Sylvain Paris and Frédéric Durand, “A fast approximation of the bilateral filter using a signal processing approach,” *International Journal of Computer Vision*, vol. 81, no. 1, pp. 24–52, Jan. 2009.
- [18] Eric C. Larson and Damon M. Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006, jan 2010.