

REMOVING OCCLUSION IN IMAGES USING SPARSE PROCESSING AND TEXTURE SYNTHESIS

Bincy Antony M¹ and K A Narayanankutty²

¹Department of Computer Science, Amrita Vishwa Vidyapeetham University,
Coimbatore, India
bincyndl@gmail.com

²Department of Electronics and Engineering, Amrita Vishwa Vidyapeetham University,
Coimbatore, India
ka_narayanankutty@cb.amrita.edu

ABSTRACT

We provide a solution to problem of occlusion in images by removing the occluding region and filling in the gap left behind. Inpainting algorithms fail in filling occlusions when the occluding region is large since there is loss of both structure and texture. We decompose the image into structure and texture images using a decomposition method based on sparseness of the image. The sparse reconstruction of the decomposed images result in an inpainted image with all the structures made intact. A texture synthesis is performed on the texture only image. Finally the structure and texture images are combined to get an image where the occlusion is filled. The performance of our algorithm in terms of visual effectiveness is compared with other algorithms used for inpainting.

KEYWORDS

Occlusion, Inpainting, Image Restoration, Sparse Representation, Texture Synthesis

1. INTRODUCTION

Occlusion is anything that obstructs vision. Removing occlusion is very important for better perception and object recognition. Occlusion is a big challenge in many of the computer vision algorithms. For instance, in 3D reconstruction from images occlusion makes reconstruction ambiguous for object parts not visible in any of the images. These parts must be reconstructed in a visually plausible way. Occlusion also reduces the aesthetic beauty of the picture. There are situations where unexpected and unappreciated people stand in front of historic places and monuments thus reducing the historic beauty of the picture.

When we look at the real world we see many objects, but we do not see each object one by one. Parts of them are occluded by others all the time. Still we have an intuitive understanding of what is covered. This is because we have seen similar objects before and know what they look like, and we can imagine what is missing. We humans have an intuitive understanding of what is missing but computers do not. To date there is no mathematical algorithm which can create things that were not captured at all. Therefore in removing occlusion using a single image we try to do an intelligent inpainting of the occluding region. Image inpainting is a process of restoring the original image from the observed data that are damaged by some degradation process. This

process of obtaining a version of image that is very close to the original is called image restoration. This kind of restoration problems is called inverse problems. It is hard to solve the inverse problems because they are not guaranteed to have a unique solution.

Reconstructing a damaged image in a way that looks natural for the eye can be done by using information from surrounding areas and merging the inpainted region into the image so that a typical viewer is not aware of it. The quality of the result will depend on what is missing. If the inpainting region is small and the surrounding area is without much texture, the result will be good.

Bertalmio et al. in [2] proposed digital inpainting algorithm based on PDE. The algorithm iteratively propagates information from outside the user selected area to be inpainted, along the isophotes (line of equal gray values). This method preserve edges across the missing regions, but when repairing large regions it introduces some blur easily. Inspired by the work of Bertalmio et al.[2], Chan and Shen proposed two image-inpainting algorithms [3], [4]. The Total Variational (TV) inpainting model [3] uses an Euler-Lagrange equation. The model simply employs anisotropic diffusion based on the contrast of the isophotes. This model was designed for inpainting small regions. The Curvature-Driven Diffusion (CDD) model [5] extended the TV algorithm to also take into account geometric information of isophotes when defining the strength of the diffusion process, thus allowing the inpainting to proceed over larger areas. CDD can connect some broken edges, but the resulting interpolated segments usually look blurry.

Efros and Leung [6] proposed a method that, although initially intended for texture synthesis, has proven most effective for image inpainting problem. The image gap is filled in recursively, inwards from the gap boundary. Each unknown pixel P at the boundary is filled with the value of the pixel Q (lying outside the image gap, i.e. Q is a pixel with valid information) such that the neighborhood $\Psi(Q)$ of Q (a square patch centered in Q) is most similar to the (available) neighborhood $\Psi(P)$ of P . Once P is filled in, the algorithm marches on to the next pixel at the boundary of the gap, never going back to P (whose value is, therefore, not altered again).The main shortcomings of this algorithm are its computational cost, the selection of the neighborhood size, and the fill order. The results are poor when the image gap is large.

Another set of algorithms combine both PDE based and texture synthesis methods to perform inpainting. Bertalmio et al. in [11] decompose the original image into textural and structural sub images. The textural sub image is restored by a texture synthesis and the structural sub image is reconstructed by a structure algorithm. They are then combined to form the inpainted image

Elad et al. [12] decompose image into structure and texture and inpaint the missing region during reconstruction. This method is however suitable only when inpainting region is small. We propose to extend this method to deal with even large occlusion by incorporation a texture synthesis on the texture only image.

The paper is organized as follows. Section 2 details the proposed method. Experimental results are discussed in section 3 followed by the conclusion in section 4.

2. PROPOSED METHOD

Real images are a combination of a structure and texture. When a large occlusion is to be filled in we have to reconstruct both the structure and the texture. This motivated Bertalmio et al. [11] to decompose image into structure and texture images and do an inpainting on these images separately. Our proposed method works in a similar way utilizing sparsity, for decomposition and inpainting for retaining lost structure followed by a patch based texture synthesis on the texture image for restoring the lost texture. In texture synthesis we copy patches of texture from known areas in the image. This method of image completion does not introduce any blur effect since actual pixel values are copied. The following sections explain our proposed method based on sparse representation and texture synthesis.

2.1. Sparse representation

A signal or image is said to be sparse when it is represented by very few significant values. Most of the signals are sparse either in time or frequency domains or both. In the last decade, sparsity has emerged as one of the leading concepts in a wide range of signal processing applications (restoration, feature extraction, source separation, and compression, to name only a few applications). Sparsity has long been an attractive theoretical and practical signal property in many areas of applied mathematics (such as computational harmonic analysis, statistical estimation, and theoretical signal processing). With advancements in optimization theory sparse based processing proved very effective in many image processing tasks.

2.2. Sparse Signals/Images

A signal x , considered as a vector in a finite-dimensional subspace of R^N , where $x = x[1], x[2], \dots, x[N]$ is sparse if most of its entries are equal to zero. If a signal is not sparse, it may be sparsified in an appropriate transform domain. For instance, if x is a sine, it is clearly not sparse, but its Fourier transform is extremely sparse. Concepts of atoms and dictionary form basis of sparse processing.

An atom is an elementary signal representing template also called bases. Examples include sinusoids, monomials, wavelets, and Gaussians. Using a collection of atoms as building blocks, one can construct more complex waveforms. A dictionary is an indexed collection of atoms. In discrete time, finite length signal processing, a dictionary is viewed as a $N \times T$ matrix whose columns are the atoms. When the dictionary has more columns than rows ($T > N$), it is called overcomplete or redundant. The overcomplete case is the setting which amounts to an underdetermined system of linear equations by linear superposition. In the proposed system we form dictionaries by amalgamating two transforms, one for the texture image and the other for the structure image. Analysis and synthesis operations using dictionaries perform sparse decomposition and reconstruction. Analysis is the operation that associates with each signal x a vector of coefficients α , $\alpha = \Phi^T x$. Synthesis is the operation of reconstructing x by superposing atoms: $x = \Phi \alpha$.

The interest in sparsity has arisen due to the new theory of signal processing called compressed sensing which states that a signal or an image can be reconstructed with a sampling rate lesser than that suggested by Nyquist provided the signal is sparse. By emphasising on sparsity compressed sensing paves way to many of the signal and image processing problem.

2.3. Image Inpainting using Sparse Representation

The input image of N pixels be represented using a 1D vector of length N by lexicographic ordering. We use a matrix $T_t \in M^{N \times L}$, $L > N$, to model image containing only texture as

$$x_t = T_t \alpha_t, \text{ where } \alpha_t \text{ is sparse}$$

Similarly we use another transform, $T_n \in M^{N \times L}$ to model the image containing only structure as

$$x_n = T_n \alpha_n, \text{ where } \alpha_n \text{ is sparse}$$

l_0 norm $\|\alpha\|_0 = \#\{i : \alpha(i) \neq 0\}$ (i.e. Number of nonzero values in α) with small value of l_0 norm indicating sparsity Finding the l_0 norm is hard and hence we use the l_1 norm which is some of nonzero values instead of l_0 norm.

A sparse representation using dictionary and l_1 norm will be

$$\{\alpha_t^{opt}, \alpha_n^{opt}\} = \arg \min_{\{\alpha_t, \alpha_n\}} \|\alpha_t\|_1 + \|\alpha_n\|_1 \text{ subject to } x = T_t \alpha_t + T_n \alpha_n \quad (1)$$

A Total Variation (TV) penalization imposed on the structure image does better recovery of edges

$$\{\alpha_t^{opt}, \alpha_n^{opt}\} = \arg \min_{\{\alpha_t, \alpha_n\}} \|\alpha_t\|_1 + \|\alpha_n\|_1 + \lambda \|x - T_t \alpha_t - T_n \alpha_n\|_2^2 + \gamma TV(T_n \alpha_n) \quad (2)$$

Here $TV(T_n \alpha_n)$ ensures that we get a piece wise smooth image and the fidelity is ensured by the term $(x - T_t \alpha_t - T_n \alpha_n)$

A slight modification on (2) will provide an inpainting with decomposition. We use a mask matrix $K \in M^{N \times N}$, to indicate the occlusion region. The main diagonal of K indicates if a pixel is occluded or not. A value of 0 in main diagonal is occlusion. The integrated inpainting model becomes

$$\{\alpha_t^{opt}, \alpha_n^{opt}\} = \arg \min_{\{\alpha_t, \alpha_n\}} \|\alpha_t\|_1 + \|\alpha_n\|_1 + \lambda \|K(x - T_t \alpha_t - T_n \alpha_n)\|_2^2 + \gamma TV(T_n \alpha_n) \quad (3)$$

This formulation seeks the decomposition into structure and texture. The fidelity term is done for the existent pixels only. From the sparse decompositions $\{\alpha_t, \alpha_n\}$ we get the corresponding inpainted texture and structure image by applying the inverse transforms.

$$x_t = T_t^T \alpha_t \text{ and } x_n = T_n^T \alpha_n$$

The filling in happens during sparse reconstruction of texture and structure using information from existent pixels.

2.4. Texture Synthesis

We use the approach proposed by Effros et al [6, 18] for texture synthesis. The texture only image is taken and filled in recursively inwards from the occlusion boundary. Each pixel P at the

boundary of occlusion is filled with the value of the pixel Q lying outside the occlusion region such that the neighborhood $\Psi(Q)$ of Q which is square patch centered at Q is most similar to the available neighborhood $\Psi(P)$ of P . To deal with large occlusion we slightly modified the algorithm and copied entire patch instead of single pixel. Starting from an initial patch we looked for similar patches in the region outside occlusion and assigned a patch randomly from the set. The overlap region was minimized using minimum error boundary cut. The synthesis process proceeded in raster scan order.

The most popular method of texture synthesis for inpainting images was proposed by [7]. They also used concept from [6, 18] but required large computation time. Finding the most appropriate patch for a query region is fundamentally hard. Number of source image blocks (image blocks from which patches are chosen for replacement) is dependant on image resolution in [7]. Today's images are of very high resolution and [7] has very poor performance due to the matching process. In our method instead of using the entire image region we use a restricted source region for patch matching. The intuition is that the best matching patch is found always nearer to the source region. This greatly improves the computational efficiency of the method especially for high resolution images.

Patch size also affect the result of texture synthesis. If the patch size is too small each patch will have insufficient information to perform best match. However if the patch size is large it will result in fewer patches and the system will not have sufficient number of patches to perform best match' Large patch size has a result of reduction in computational time as more patches get filled at a time. We make patch size proportional to whole size and image content. Small patches are used for fine structures and large patches for large structures.

3. EXPERIMENTAL RESULTS

The proposed method gives a good filling in of large occlusion. The choice of transforms influenced the fill in process. We tried with different transforms and used the one that was most affective for natural images. The transforms used were curve lets and localDCT. The occluded bungee jumper image and the mask indicating occlusion are shown in Figure 1. We used this image because it is widely used in inpainting problems in past. Figure 2 illustrates the proposed method. It can be seen in Figure 2.a that all the structure in the missing region is created well by sparse inpainting. The texture is however not restored. Figure 2.b shows the reconstructed texture image. Figure 2.c shows the occlusion free image obtained by combining structure and texture images. Figure 3 shows the result of occlusion removal by other methods. See how the PDE method is introducing blur. The texture synthesis method failed in restoring structure. The output of [7] is visually good. Result on real life images shows the effectiveness of our method.



a) Original Image



b) Mask Image indicating occlusion

Figure 1. Input Images



a) Structure Image

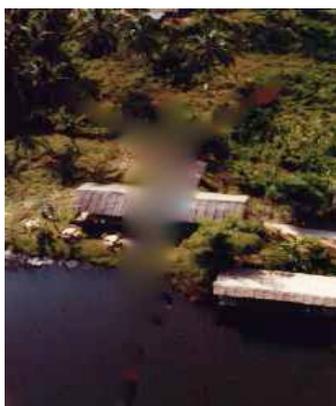


b) Texture Image

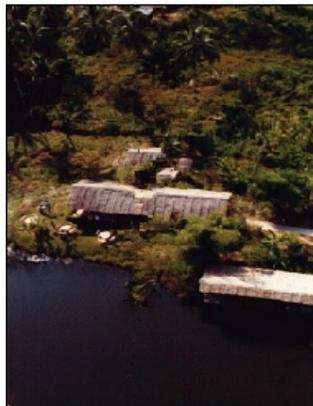


c) Structure and Texture Images

Figure 2. Proposed Method



a) PDE Inpainting



b) Texture Synthesis



c) Criminisi [7]

Figure 3. Result of other methods

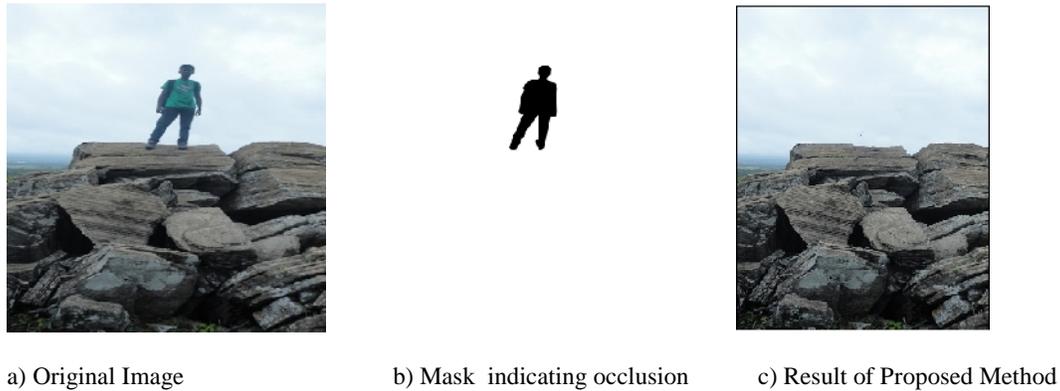


Figure 4. Result on Real Life Images

4. CONCLUSION

We proposed a method to remove large occlusions in images by sparse image inpainting and texture synthesis. The novel contribution in the paper is in combining the sparse reconstruction with inpainting and texture synthesis. The large small and dispersed occlusions could be effectively filled in by our method. The structure of the missing region is correctly retained by sparse inpainting. The texture synthesis process fills the lost texture. We also improved on the computational efficiency by restricting the source region for searching matching patches. The visual quality was improved by choosing patch size according to image content. The problem with our approach is the need to tune the patch size based on image content. We are focusing on improving the sparse inpainting process. We aim is to extend this method to inpaint videos in future.

REFERENCES

- [1] K A Narayanankutty, K P Soman, "Understanding Theory Behind Compressed Sensing" Int. J. Sensing, Computing & Control, Vol. 1, No. 2, pp. 81-92, 2012
- [2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. "Image inpainting." In Proc. SIGGRAPH'00, New Orleans, USA, pages 417-424, 2000.
- [3] T.F. Chan, S.H. Kang, and J. Shen. Euler's elastica and curvature based inpainting. SIAM Journal of Applied Math., 63(2):564-592, 2002
- [4] T. Chan, J. Shen, "Mathematical models for local nontexture inpainting," SIAM J. Appl. Math., vol. 62, no. 3, pp. 1019-1043, 2001.
- [5] T. Chan, and J. Shen, "Non-texture inpainting by curvature-driven diffusions (CDD)," J. vis. Commun. Image Representn., vol. 12, no. 4, pp. 436-449, 2001.
- [6] A. Efros and T. Leung, "Texture synthesis by non-parametrics sampling," proc. IEEE international conference on computer vision, pp.1033-1038, Corfu, Greece, 1999
- [7] A.Criminisi, P.Perez, and K. Toyama "Region filling and object removal by exemplar-based inpainting". IEEE Transactions on Image Processing, 13(9):1200-1212, 2004.
- [8] M. Ashikhmin, "Synthesizing natural textures", Proc. ACM Symp. On Interactive 3D Graphics, pp. 217-186, Research Triangle Park, NC, Mar 2001.
- [9] J. Sun, L. Yuan, J. Jia, and H.Y. Shum. "Image completion with structure propagation". In ACM SIGGRAPH 2005 Papers, page 868.
- [10] C. Barnes, E. Shechtman, A. Finkelstein, and D.B. Goldman. Patchmatch: "A randomized correspondence algorithm for structural image editing", ACM Transactions on Graphics, 28(3):2, 2009.
- [11] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. "Simultaneous structure and texture image inpainting". IEEE Transactions on Image Processing, 12(8):882-889, 2003.

- [12] M. Elad, J. L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis," *Appl. Comput. Harmon. Anal.*, vol. 19, pp. 340–358, 2005.
- [13] M. J. Fadili, J. L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *The Comput. J.*, vol. 52, no. 1, pp. 64–79, 2009.
- [14] Devasruthi D, Hema P Menon, K A Narayanankutty "FE-BEMD and exemplar based hybrid image inpainting for occlusion removal", *International Journal of Computer Applications*, No-8, Article 7, IJCA Aug 2011
- [15] Zongben Xu, Jian Sun, Sch. of Sci, Xi'an Jiaotong Univ. Xi'an , "Image Inpainting by Patch Propagation Using Patch Sparsity" , *IEEE Transactions on Image Processing*, Volume 19 Issue 5, May 2010 IEEE Press Piscataway, NJ, USA .
- [16] E. Candes, M. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE Sig. Proc. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008
- [17] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006
- [18] A. Efros, W.T. Freeman, "Image quilting for texture synthesis and transfer", In *Proc. ACM Conf. Comp. Graphics (SIGGRAPH)*, pp. 341–346, Eugene Fiume, Aug 2001.