

IMAGE RECOVERY USING SPARSE RECONSTRUCTION BASED TEXTURE REFINEMENT

Haricharan Lakshman¹, Martin Köppel¹, Patrick Ndjiki-Nya¹ and Thomas Wiegand^{1,2}

¹Image Processing Department
Fraunhofer Institute for Telecommunications -
Heinrich Hertz Institute
Einsteinufer 37, 10587 Berlin, Germany

²Image Communication Chair
Department of Telecommunication Systems
Technical University of Berlin
Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT

We present a robust algorithm for spatial recovery of missing region in images. The algorithm consists of two stages: sparse modeling and patch based refinement. We note that a model based image recovery might not be able to reconstruct the richness or details in a signal unless the signal truly fits that model. We show that the reconstruction using a sparse model provides enough information about the inherent features present in the unknown area, using which, a patch based refinement process can replicate the structure and the natural texture from the surrounding available samples. The developed algorithm is tested on a variety of image characteristics. Significant objective and subjective gains are observed compared to the state-of-the-art.

Index Terms— Image recovery, Texture refinement, Sparse reconstruction, Error concealment

1. INTRODUCTION

Recovery of missing regions of images are required for many applications. For example, image and video communications using block-based coders over error-prone channels may lead to block losses. Similarly, imperfections in capture, storage or other processes in an image processing system produce errors which necessitate the use of restoration algorithms that estimate missing regions. Decoder side recovery techniques work on the received data without the need for any error correction data to be transmitted by the encoder. Most algorithms reported in literature are best suited for either pure structure or pure texture areas. Although an attempt is made in [1] to classify the lost blocks as structure or texture and use the suitable algorithm for each case, the issue of structure within texture blocks or vice versa is not well addressed. With the popularity of high resolution content, block sizes larger than the traditional 16×16 are becoming increasingly important for coding [2] and hence the decoder should be able to deal with large block errors.

Algorithms for image recovery can be broadly classified into two categories - Model driven and Data driven. In model driven recovery techniques, the image samples are assumed

to be emanating from a parametric model. The surrounding samples of the unknown block are used to estimate the parameters of the model, which are then utilized to generate samples in the unknown area. In recent years, considerable interest has been paid to sparse image modeling techniques. These algorithms operate on the intuition that natural images can be decomposed into sparse combination of basic elements [3]. Sparse modeling techniques are able to successfully capture the inherent structure but produce blurring artifacts when the dictionary elements are not tuned to current texture. In order to counter this effect, techniques have been suggested to update the dictionary elements based on the observed data [4], adaptive thresholding [5] etc.

In data driven approaches, known data is directly used to control the filling of unknown areas [6, 7]. In [8], we proposed Structure-Aware Inpainting, which propagates the edges detected in known area into the unknown area and replicates the structure and texture from appropriate locations. However, it relies on the masks generated by segmentation, which can produce visible artifacts when filling happens from an incorrect segment. Data driven approaches provide excellent quality for some images but fail to produce reasonable results for some others.

In this paper, we combine the main ideas of model driven and data driven techniques to arrive at a powerful algorithm for recovering missing data from nonstationary signals. The proposed approach performs a Markov Random Field (MRF) based refinement step on the samples recovered by sparse modeling. The algorithm starts by building a sparse model of the known surrounding area and extrapolates into the missing area. Then, using the extrapolated samples as an initialization for the missing area, a patch based refinement is performed, replicating the natural structure & texture from the surrounding into the missing area. Unlike prevalent data driven algorithms, our approach does not require any segmentation or edge detection steps. Moreover, the algorithm can be implemented using fixed dictionary elements, thereby eliminating dictionary updation or adaptive thresholding steps. The algorithm is tested on an extensive set of images with missing regions containing edges, textures and other image features.

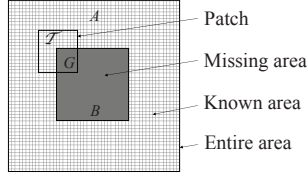


Fig. 1. Labeling of different regions of an image. \mathcal{A} : Known area; \mathcal{B} : Missing area; \mathcal{T} : Template; \mathcal{G} : Target

2. BACKGROUND

We consider a rectangular neighborhood \mathcal{R} that contains unknown samples in area \mathcal{B} surrounded by known samples in area \mathcal{A} . We define a *patch* to consist of a small subset of samples in \mathcal{R} such that some area comprises of known (or already estimated) samples \mathcal{T} and the others are unknown \mathcal{G} , as shown in Fig. 1. In a certain patch, the known area and the unknown area form *template* and *target* respectively.

Patch based image recovery (PBR) is based on homogeneous property of textures that for a generated texture to be similar to an input texture it is sufficient that all neighborhoods in the generated texture be similar to some neighborhood in the input [9]. In PBR, the location that best matches the template is searched in the known area and the samples corresponding to target area are copied. The main issue with PBR is that the samples in the current target area might not be correlated to the matched patch location. For instance, there could be singularities in the form of edges or segment boundaries inside the target region that is not captured in the matched location but is evident from the spatially adjacent samples in known area. Therefore, such a PBR algorithm is often controlled by segmentation or edge detection. Additionally, the optimal patch size depends strongly on structure and texture of the image. In [8], the identified edges are first propagated into unknown area before patching happens. The process of edge detection in known area and edge propagation into unknown area contribute to the instability of such algorithms because the perceivable segments and edges cannot be well defined mathematically.

In this paper, we propose an alternate solution without explicitly using segmentation or edge detection. We obtain the necessary structural information from a model based approach and use this to control the patch based filling. We recognize that a model based approach provides an estimate of structure within the missing area, but might not be able to capture the richness or details in the non-stationary signal. We postulate that the reconstruction using a sparse model provides enough information about the inherent structures present in the unknown area so that the template *and* target areas can be matched in the patch based texture refinement process. The use of initial estimate of target samples along with template samples can bring back the missing details, making the reconstruction visually plausible.

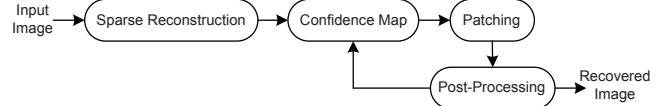


Fig. 2. Block diagram of proposed algorithm. First, the missing samples are estimated by sparse combination of dictionary elements. Next, patches are selected according to the order specified by confidence map [6] and texture refinement is performed. Finally, in-loop post processing of matched target area provides seamless transitions across adjacent patches. Recovered image is obtained after completing all patches.

3. FRAMEWORK FOR RECOVERY

The pixel intensities of region \mathcal{R} can be interpreted as a column vector $\mathbf{f} \in \mathbb{R}^N$, where N denotes the total number of samples in \mathcal{R} . For capturing the inherent structure present in known samples of \mathbf{f} , it is approximated using only a few basic elements from a dictionary \mathcal{D} . The entries of \mathcal{D} could be overcomplete and are denoted as vectors \mathbf{d}_k of dimension $N \times 1$. A parametric model consisting of a linear combination of vectors \mathbf{d}_k is used for generating the approximation vector \mathbf{g} , so that

$$\mathbf{g} = \sum_{\forall \mathbf{d}_k \in \mathcal{K}} c_k \cdot \mathbf{d}_k, \quad (1)$$

where $k \in \mathcal{K}$ consists of the vectors chosen from \mathcal{D} used for modeling and c_k are the model parameters to be estimated. A vector \mathbf{h} of dimension $N \times 1$ is formed by taking known samples from area \mathcal{A} of \mathbf{f} and the estimated samples from area \mathcal{B} of \mathbf{g} ,

$$\mathbf{h} = \mathbf{M}_1 \cdot \mathbf{f} + \mathbf{M}_2 \cdot \mathbf{g}, \quad (2)$$

where \mathbf{M}_1 and \mathbf{M}_2 are matrices of dimension $N \times N$ consisting of 1s at diagonal entries when a sample is selected from the vector it is operating on and zeros elsewhere. Finally, texture refinement operates on the vector \mathbf{h} by copying samples from appropriate locations in \mathcal{A} to the locations in \mathcal{B} resulting in output samples \mathbf{s} ,

$$\mathbf{s} = \mathbf{T} \cdot \mathbf{h}, \quad (3)$$

where \mathbf{T} is a $N \times N$ texture refinement matrix. In order to produce the final image, the unknown parameters c_k and \mathbf{T} are to be determined. A joint estimation of c_k and \mathbf{T} is an extremely complex task, hence we resort to a simple way of determining c_k s first and then calculating \mathbf{T} using MRF approach.

3.1. Initialization using sparse model

A model of the area \mathcal{R} , using the samples of \mathcal{A} is built by means of a masking vector \mathbf{m} which is defined to contain a value of one in \mathcal{A} and zero in \mathcal{B} . The estimation of c_k is done such that the approximation error between the samples

at known locations of \mathbf{f} and the corresponding samples produced by the model \mathbf{g} ,

$$E = (\mathbf{f} - \mathbf{g})^T \cdot \mathbf{m} \cdot \mathbf{m}^T \cdot (\mathbf{f} - \mathbf{g}), \quad (4)$$

is minimized. An isotropically decaying weighting function [10] is additionally employed in \mathbf{m} so that the known samples in the vicinity of the unknown area get a higher importance than the samples that are far from it.

Direct minimization of Eq. (4) by setting the partial derivatives of cost function E w.r.t c_k to zero leads to an underdetermined system of equations as the number of known samples in \mathcal{A} is less than the total number of samples in \mathcal{R} . For solving this underdetermined problem, a greedy approach is taken in which the signal is approximated in terms of one additional vector from \mathcal{D} per iteration [11, 12] resulting in a sparse representation. In each iteration, the additional vector is chosen in such a way that the reduction of the weighted residual energy is maximized. In case the dictionary is composed of basis vectors of \mathbb{R}^N , Frequency Selective Extrapolation [12], a fast and efficient algorithm that performs sparse modeling in transform domain, can be employed.

3.2. Patch based Refinement

We note that the patch based synthesis has the ability to reconstruct both structure and texture when a coarse estimate of structure in the target area is available. We utilize the filling order determination algorithm proposed in [6] and enhance it to include already estimated samples through sparse reconstruction. The gradient is computed for all the samples in the current patch thus leading to a better isophote direction (Please refer to [6] for the original algorithm). The priority of potential patches are computed and the patch with highest priority is selected for filling.

The filling algorithm searches for a patch in the source region that is most similar to the template *and* target area of the current patch. Through this combined matching, the assumption that the best template matched patch is also the best patch for copying target samples is relaxed. The best patch is found by minimizing the following cost function,

$$J = (\mathbf{x} - \mathbf{c})^T \mathbf{W}(\mathbf{x} - \mathbf{c}) + \lambda \cdot (\nabla \mathbf{x} - \nabla \mathbf{c})^T \mathbf{W}(\nabla \mathbf{x} - \nabla \mathbf{c}) \quad (5)$$

where, \mathbf{c} is the patch at current location to be filled; \mathbf{x} is the candidate patch in area \mathcal{A} ; $\nabla \mathbf{c}$ is the sample-wise gradient of \mathbf{c} ; $\nabla \mathbf{x}$ is the sample-wise gradient of \mathbf{x} ; \mathbf{W} is a diagonal matrix with a value of w at locations that multiply target samples and $1 - w$ at locations that multiply template samples; and λ is the relative importance of gradient component in the matching process.

Having found the best patch, a post-processing algorithm is applied to ensure smooth transition between adjacent patches [8]. This happens in-loop and influences the filling of future patches.

Image	TM	CM	SP+TM	SP+CM	Proposal
Airplane	21.27	20.40	21.96	22.08	22.66
Baboon	18.86	18.57	19.52	19.77	20.21
Girl2	26.36	25.10	27.66	27.81	28.52
Lake	19.03	18.77	20.07	20.23	20.72
Lena	24.98	24.92	26.85	26.91	27.53
Peppers	25.96	24.68	26.87	27.29	28.08

Table 1. PSNR(dB) results of recovery using different algorithms denoted as TM: Template Matching; CM: Confidence Map; SP: Sparse reconstruction.

4. SIMULATION SETUP AND RESULTS

For evaluating the proposed algorithm, we generate input images by cutting out blocks from standard test images and images from ‘The Berkeley Segmentation Dataset’. The input images consist of missing blocks of size 16×16 and 32×32 samples. In the case of TM without CM, the filling is done in a helical order. For TM & CM, patch sizes of 7×7 and 11×11 are found to be optimal for missing blocks of size 16×16 and 32×32 respectively. The same patch sizes are also employed for the refinement step in our proposal.

The relative weight w in Eq. (5) was varied between 0 and 1 and it was found that a value of 0.5 (equal weights to \mathcal{T} and \mathcal{G}) gives good subjective results. The value of λ that controls the emphasis of gradient component in patch search was set to 0.7. For SP, the dictionary elements are composed of Fourier basis functions of size 64×64 and the sparse modeling selects 100 bases out of 64^2 possible bases. Fourier basis has better extrapolation properties compared to other bases as shown in [10]. The results of recovery comparing different methods are summarized in the form of PSNR metric in Tab. 1.

For subjective evaluation of recovery results, portions of test results are depicted. Fig. 3 shows results for test images ‘Lena’ and ‘Baboon’ with 16×16 missing blocks (left) and recovered images (right). In order to examine the recovery results closer, small image regions of size with single block loss are shown along with recovered results in Fig. 4. The images are composed of different features and consists of both textured and structured regions. Fig. 4(a)-(b) are of size 64×64

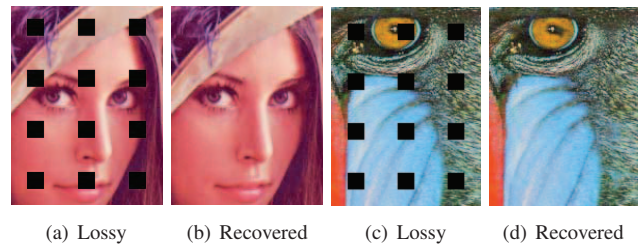


Fig. 3. Recovery results for 16×16 missing blocks.

with loss size of 32×32 while Fig. 4(c)-(f) are 48×48 with loss size of 16×16 . Notice that in Fig. 4(a)-(b), not only the texture of flower/trees but also the structure in the form of outline is synthesized by our algorithm. Fig. 4(c) contains missing region of background and foreground rocks. The missing block in Fig. 4(d) consists of a complex texture on the left half, relatively flat area on the right half and the edge separating these regions. Some inaccuracies in the structure reconstruction can be noticed in this case. Fig. 4(e) & Fig. 4(f) are composed of cloth texture and linear structure respectively. The resulting images look natural and do not contain blurring artifacts.

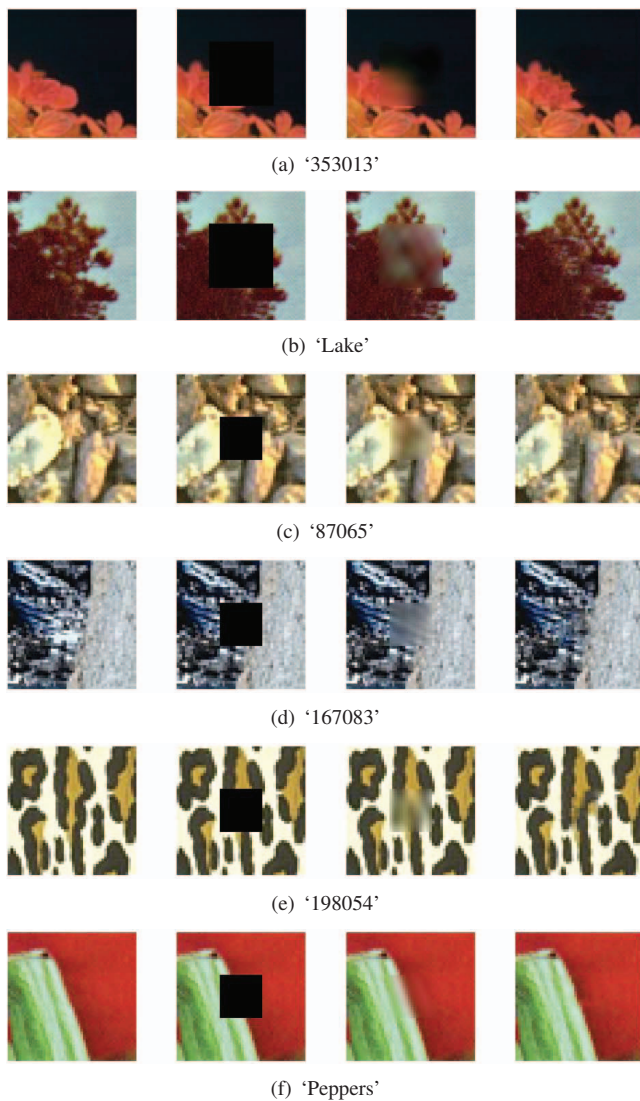


Fig. 4. Recovery results for 32×32 and 16×16 missing blocks. From left: Original image; Original image with missing block; Recovery using sparse reconstruction; Recovery using proposed method.

5. CONCLUSION

We presented a powerful algorithm by combining model-based and data-based recovery techniques that can be used in a variety of applications like prediction, error concealment, inpainting etc. We showed that the patch based synthesis technique has the ability to reconstruct both structure and texture when a coarse estimate of structure in the target area is available. For obtaining the coarse estimate, we used sparse modeling as it can robustly capture the inherent features in missing regions. The patch based refinement step brings back the richness and the recovered images appear visually plausible. The algorithm has both subjective and objective gains compared to the state-of-the-art.

6. REFERENCES

- [1] S. D. Rane, G. Sapiro, and M. Bertalmio, "Structure and Texture Filling-In of Missing Image Blocks in Wireless Transmission and Compression Applications," *IEEE Trans. Image Process.*, vol. 12, no. 3, pp. 296–303, March 2003.
- [2] P. Chen, Y. Ye, M. Karczewicz, "Video coding using extended block sizes", *VCEG-AJ23*, San Diego, USA, 8-10 Oct., 2008
- [3] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Research*, vol. 37, pp. 3311–3325, 1997.
- [4] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation", *IEEE Trans. On Sig. Proc.*, Vol. 54, no. 11, pp. 4311–4322, November 2006.
- [5] O. G. Güleriyüz, "Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising - Part I & II", *IEEE Tran. Image Proc.*, March 2006.
- [6] A. Criminisi, P. Perez, and K. Toyama, "Region Filling and Object Removal by Exemplar-based Image Inpainting," *IEEE Trans. Img. Proc.*, vol. 13, no. 9, pp. 1200-1212, 2004.
- [7] A. Efros, and W.T. Freeman. "Image quilting for texture synthesis and transfer." *Proc. of ACM SIGGRAPH*, pp. 341-346, Eugene Fiume, August 2001.
- [8] P. Ndjiki-Nya, M. Köppel, D. Doshkov and T. Wiegand, "Automatic Structure-Aware Inpainting for Complex Image Content," *Int. Sym. on Visual Computing*, Nov 2008.
- [9] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, "Texture optimization for example-based synthesis," *Proc. of ACM SIGGRAPH*, vol. 24, no. 3, pp. 795 - 802, July 2005.
- [10] K. Meisinger, "Selective Signal Extrapolation and its Application in Image and Video Communications," *PhD Thesis.*, University Erlangen-Nuremberg, Germany, Jan. 2008.
- [11] S. Mallat, and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [12] A. Kaup, K. Meisinger and T. Aach, "Frequency selective signal extrapolation with applications to error concealment in image communication," *Int. J. Electron. Commun.*, vol. 59, pp. 147–156, March 2005.