INPAINTING OF MULTI-CHANNEL IMAGES USING A DICTIONARY LEARNED FROM A SINGLE IMAGE

Marko Filipović[1], Ivica Kopriva[1] and Andrzej Cichocki[2]

[1]Ruđer Bošković Institute/Division of Laser and Atomic R&D
Bijenička cesta 54, 10002, Zagreb, Croatia
filipov@irb.hr, ikopriva@irb.hr

[2]RIKEN Brain Science Institute/Laboratory for Advanced Brain Signal Processing
Wako-shi, Saitama, 351-0198, Japan
cia@brain.riken.jp

ABSTRACT

Image inpainting consists in recovering missing parts of an image. Since a color image is a 3D array, tensor completion methods are applicable to this problem. Tensor completion approach based on trace norm minimization can be useful when the fraction of missing pixels is not large, with the advantage that the training set is not required. Here, we demonstrate that the dictionary for sparse representation of multichannel image patches can be learned from a single (clean) image, yielding results comparable to those as obtained by a dictionary learned on a training set of images. We show that the learned dictionary-based approach performs considerably better than tensor completion both on color images of natural scenes and multi-phase computed tomography images.

Index Terms – dictionary learning, independent component analysis, color image, computed tomography image, inpainting, tensor completion.

1. INTRODUCTION

Image inpainting problem consists in recovering missing parts of an image, wherein the locations of missing pixels are known or selected by the user. A related problem is removal of salt-and-pepper noise. Salt-and-pepper noise occurs when some image pixels are saturated, meaning that they have minimal or maximal value in the range of pixel values, wherein the locations of noise-corrupted pixels are unknown. For images of natural scenes, salt-and-pepper noise removal problem can be reduced to the inpainting problem by declaring all saturated pixels as missing [1]. This approach works well even if some pixels are mistakenly declared as missing. In [1], the dictionary for sparse representation of image patches learned by independent component analysis (ICA) was used for image inpainting and salt-and-pepper noise removal. Good results were demonstrated even for 80% of missing or corrupted pixels. This approach was extended for color images in [2]. There, the tensor completion approach to color image inpainting based on minimization of trace (nuclear) norm was also discussed. Trace norm of a matrix, defined as the sum of its singular values, is convex function and an often used approximation of matrix rank. Low-rank assumption is often used to regularize the ill-posed tensor completion problem. Low-rank tensor completion based on trace norm minimization was used for color image inpainting in [3]. However, it was argued in [2] that the low-rank assumption is data-dependent and often fails for color images. Therefore, tensor completion approach doesn't perform well for color image inpainting when the fraction of missing pixels is large. When the fraction of corrupted or missing pixels is not large, tensor completion approach can give visually pleasing results, despite the low-rank assumption not being satisfied. Since the tensor completion approach doesn't require a training set of images, it could be considered as an advantage compared to learned dictionary-based approach, which generally needs training set of images to learn the dictionary on. In this paper, which extends our previous papers [1, 2], we demonstrate that the dictionary can be learned on a *single image* only. At the same time, significantly better results are obtained than with the tensor completion approach. This could be important in applications where possibly not many images are available for dictionary learning.

In above cited papers, only images of natural scenes were considered. Although obtained results are interesting, it could be argued that they are of small practical relevance. Perhaps more practically important applications are in the area of medical imaging. Therefore, in this paper we also use three-phase computed tomography (CT) images for the demonstration of inpainting algorithms.

In the following section we briefly describe tensor completion and learned dictionary approaches to color and three-phase CT image inpainting. The experimental results are presented in Section 3. Conclusions are drawn in Section 4.

2. APPROACHES TO COLOR AND THREE-PHASE CT IMAGE INPAINTING

In this paper we consider only images with 3 channels: RGB color images and three-phase CT images. However, the same approach, as described in the sequel, could be used for images with more channels.

RGB color image is a 3-D tensor. Therefore, color image inpainting is a special instance of tensor completion problem. In [3], tensor completion algorithms based on minimization of trace norm were used for inpainting of some color and magnetic resonance (MR) images. Therein, low-rank assumption (expressed through the use of the trace norm) was used as a regularization for the ill-posed tensor completion problem. However, since RGB color and MR images are not low-rank, significantly better results can be obtained with the learned dictionary approach, described in the following.

We denote an RGB color or three-phase CT image as $\underline{\boldsymbol{X}} \in \mathbb{R}^{I_1 \times I_2 \times 3}$, where I_1 and I_2 denote the number of rows and columns, respectively. It has been demonstrated in many papers ([4, 5, 6, 7, 1], to cite only several) that by learning a dictionary for (approximately) sparse representation of image *patches*, excellent results can be obtained in problems of image inpainting, denoising or reconstruction from small number of measurements. Image patch is a small image block, $\underline{\boldsymbol{X}}_p \in \mathbb{R}^{\sqrt{l} \times \sqrt{l} \times 3}$, where $\sqrt{l} \times \sqrt{l}$ is the spatial size of the patch, and p denotes the patch index. To learn the dictionary for sparse representation of image patches, a collection of randomly selected and vectorized patches from image(s) in the training set is stored columnwise in matrix $\boldsymbol{Y} \in \mathbb{R}^{n \times T}$, where $n = 3l$ and T denotes the number of selected patches. The dictionary $\boldsymbol{D} \in \mathbb{R}^{n \times m}, n \leq m$, is learned such that every vectorized patch $\boldsymbol{y}_t \in \mathbb{R}^n$ can be represented as $\boldsymbol{y}_t \approx \boldsymbol{D}\boldsymbol{c}_t$, where $\|\boldsymbol{c}_t\|_0 \ll m$. Here, $\|\boldsymbol{c}_t\|_0$ stands for the number of nonzero elements of vector $\boldsymbol{c}_t$. Therefore, $\boldsymbol{D}$ is learned through the factorization $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{C}$, where sparsity is imposed on the columns of $\boldsymbol{C}$. That is known as sparse coding. In our previous papers [1, 2] we have used ICA-based probabilistic approach for this purpose. Sparsity of matrix $\boldsymbol{C}$ was imposed implicitly through the choice of parameter(s) in ICA algorithm, see [1, 2]. In this way, good results were obtained in image inpainting and removal of salt-and-pepper noise experiments.

It was argued in [2] that, if the training set is rich enough, the learned dictionary atoms represent the diversity of colors in natural images. However, here we show that comparable results can be obtained by learning the dictionary from a *single* image. This eliminates the need for many training images. Of course, the choice of an image used for dictionary learning is important. In the experiments presented in the following section we have used several images to learn different dictionaries and compared the results of inpainting experiments.

Once the dictionary is learned, the inpainting is performed by processing every image patch, as described in [2]. Namely, for every image patch, the following problem is solved:

$$\hat{\boldsymbol{c}}_p = \arg\min_{\boldsymbol{c}} \left\{ \|\boldsymbol{c}\|_0 : \left\| \boldsymbol{M}_p \boldsymbol{D}\boldsymbol{c} - \boldsymbol{v}_p \right\|_2^2 \leq \epsilon \right\} \quad (1)$$

where $\boldsymbol{M}_p \in \mathbb{R}^{k_p \times n}, k_p < n$, denotes the projection onto indexes of k_p observed (available) pixels in current patch, p is the patch index, $\boldsymbol{v}_p \in \mathbb{R}^{k_p}$ denotes the vector of observed pixels, and $\epsilon > 0$ is the allowed error of the representation (it is set heuristically). Reconstructed patch $\hat{\boldsymbol{x}}_p$ is obtained (in vectorized form) as $\hat{\boldsymbol{x}}_p = \boldsymbol{D}\hat{\boldsymbol{c}}_p$. After processing all patches, overlapping regions are averaged.

Here we note that the usual procedure for grayscale images before dictionary learning *and* inpainting (denoising) is to preprocess all patches to make them zero-mean. In other words, before processing every patch, mean value $m(\boldsymbol{v}_p)$ of the available pixels (i.e., mean of vector $\boldsymbol{v}_p$ above) is subtracted from $\boldsymbol{v}_p$. Then, the modified problem is solved: $\tilde{\boldsymbol{c}}_p = \arg\min_{\boldsymbol{c}} \left\{ \|\boldsymbol{c}\|_0 : \left\| \boldsymbol{M}_p \boldsymbol{D}\boldsymbol{c} - \left(\boldsymbol{v}_p - m(\boldsymbol{v}_p) \right) \right\|_2^2 \leq \epsilon \right\}$. The reconstructed patch is then obtained as $\hat{\boldsymbol{x}}_p = \boldsymbol{D}\tilde{\boldsymbol{c}}_p + m(\boldsymbol{v}_p)$. This leads to good results for grayscale images, and when the spatial distribution of missing pixels is uniform. However, for *color* images, this approach leads to color artifacts. Therefore, mean should not be subtracted when processing color image patches. This applies to both inpainting and dictionary learning. Namely, dictionary is learned on original color patches, without making them zero-mean. Preprocessing of patches suggested in [2] to avoid color artifacts is then not necessary.

3. EXPERIMENTAL RESULTS

In this section we present the experiments performed on natural images and three-phase CT images.

3.1. Natural images. Figure 1 shows several images used for dictionary learning. The images were downloaded from [8].

Figure 1. Training images. Four dictionaries were learned, from the above images.

For every image, 20000 patches of size $8 \times 8 \times 3$ were extracted. Flat patches (i.e., those with small variance) were discarded. Vectorized patches were stacked as columns of matrix $\boldsymbol{Y}_i$ (here, i denotes the image index).

The complete $(m = n)$ dictionary is learned from every image using the FastICA algorithm [9] with *gauss* nonlinearity, which yields similar results as when using *tanh* nonlinearity (used in [2]), while making FastICA algorithm faster.

Two examples of learned dictionaries are shown in Figure 2.

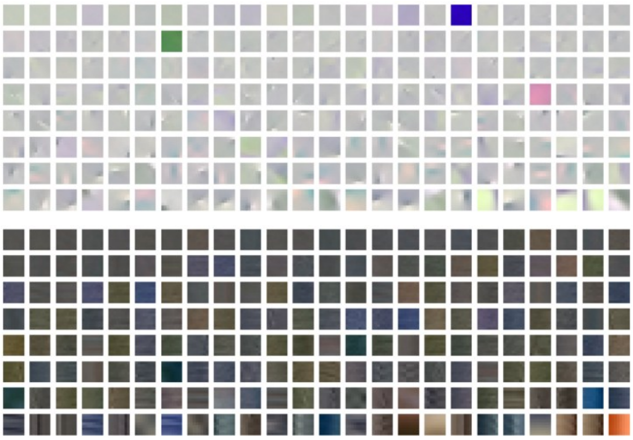

Figure 2. Tensorized columns (*atoms*) of size $8 \times 8 \times 3$ of learned dictionaries. The upper part corresponds to the image in top left of Figure 1, while the lower part corresponds to the image in top right of Figure 1.

Robust SL0 algorithm [10] was used for solving the sub-problems (1). Parameter ϵ was set to 1. Other values were also tried, without improving the quality of the results.

We have compared the learned dictionary-based approach with tensor completion. Tensor completion method described in [3] was used (the implementation is available at[1]). Default values of the parameters were used. Namely, the weights of the trace norm terms were $[100, 100, 0]$. The number of iterations was set to 3000, which was enough for the algorithm to reach the stationary point. Other values were also tested, yielding similar results. Figure 3 shows the images used for comparing the methods. Tables 1 and 2 show the results, comparing the ICA-learned dictionary-based method and tensor completion method. Values in the table are peak signal-to-noise ratios (PSNR-s) in decibels (dB). The experiments were repeated for several random masks (i.e. distributions of missing pixels), without significant change in the comparative performance of methods. Namely, variations in PSNR for different random masks were less than 1.5 dB in all simulations.

Figure 3. Images used for the comparison of methods.

Table 1. Results on test images from Figure 3, with 50% missing pixels. For learned dictionary approach, 4 results are presented for each image, since 4 dictionaries were learned. Values in the table are peak signal-to-noise ratios (PSNR-s) in decibels (dB).

Method \ Image	Fig.3(a)	Fig.3 (b)	Fig.3 (c)
ICA-learned dictionary 1	37.79	37.5	42.7
ICA-learned dictionary 2	37.85	36.4	42
ICA-learned dictionary 3	37.97	37.61	42.92
ICA-learned dictionary 4	37.57	37.1	42.34
Tensor completion	29.99	27.6	33.53

Table 2. Results on test images from Figure 3, with 80% missing pixels. Values in the table are PSNR-s in dB.

Method \ Image	Fig.3 (a)	Fig.3 (b)	Fig.3 (c)
ICA-learned dictionary 1	28.53	28.96	35.1
ICA-learned dictionary 2	29.43	28.35	34.1
ICA-learned dictionary 3	28.8	28.62	34.26
ICA-learned dictionary 4	28.9	28.8	34.59
Tensor completion	23.84	20.72	25.48

It is obvious that the learned dictionary-based approach performs significantly (up to 10 dB) better than the tensor completion approach. This is true both in terms of PSNR and visually, as shown in Figures 4 and 5.

In [2], the result of 29.36 dB (mean over 5 random masks) was achieved on the castle image (Figure 3 (a)) with 80% missing pixels. It can be seen from Table 2 that, when the dictionary is learned on a single image, the best result was

1 http://peterwonka.net/Publications/LRTC_Package_Ji.zip

29.43 dB, which is even better than the result reported in [2], while the average over 4 different dictionaries is 28.91 dB. Therefore, it can be seen from this example that comparable results in image inpainting can be achieved when the dictionary is learned on a single image only (relative to the case when the dictionary is learned on a training set of images).

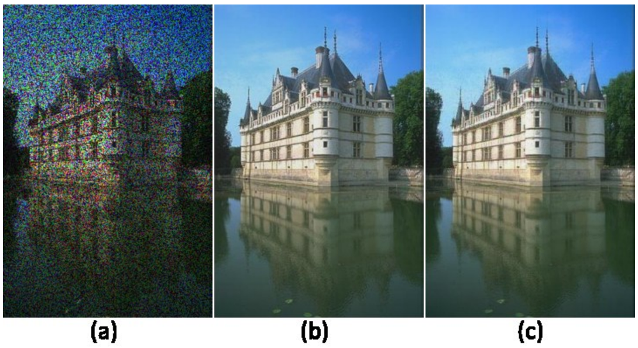

Figure 4. Comparison of the inpainting results using the learned dictionary and tensor completion approach. (a) Image with 50% missing pixels. (b) Image reconstructed using the ICA-learned dictionary. (c) Image reconstructed using the tensor completion approach.

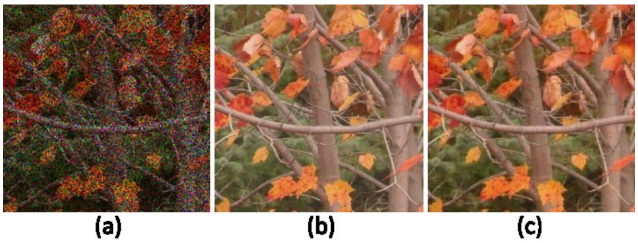

Figure 5. Comparison of the inpainting results using the learned dictionary and tensor completion approach. (a) Image with 50% missing pixels. (b) Image reconstructed using the ICA-learned dictionary. (c) Image reconstructed using the tensor completion approach.

3.2. Three-phase CT images. Proposed method was also tested on experimental three-phase CT images of an abdomen.

For CT images, the dictionary was learned on a *single* phase of one slice only. It is shown in Figure 6. Therefore, *grayscale* patches were used for dictionary learning in this case (as in [1]), and every phase-image was inpainted separately in the inpainting phase. Patch size was set to 12×12 pixels. The learned dictionary is shown in Figure 7.

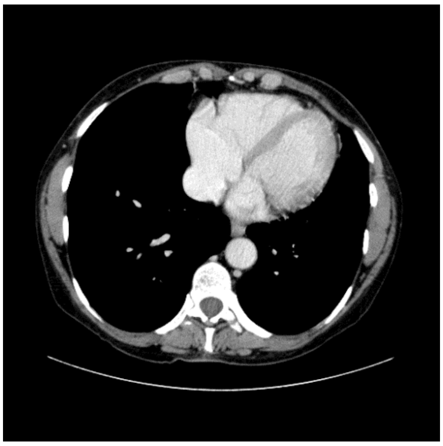

Figure 6. Phase-image on which the dictionary was learned. It was chosen from a different slice than the one used in inpainting experiments, shown in Figure 8.

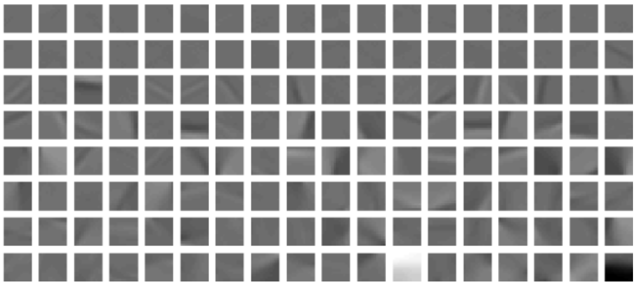

Figure 7. Matricized dictionary atoms of size 12×12.

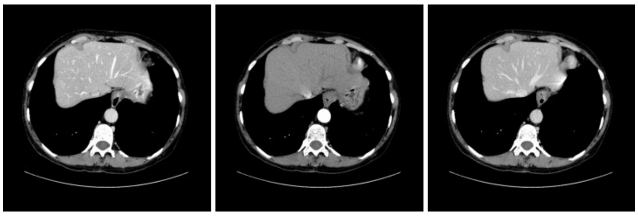

Figure 8. From left to right: portal-venous 1, arterial and portal-venous 2 phase CT images used for the inpainting experiments.

Three-phase CT image used for the inpainting experiments is shown in Figure 8. Obtained results were as follows. For 50% missing pixels, ICA-learned dictionary-based approach achieved root mean squared error (RMSE) of 9.37, while tensor completion approach achieved RMSE of 24.77. For 80% missing pixels, learned dictionary-based approach achieved RMSE of 22.58, while tensor completion approach achieved 75.6. Reconstructed images are shown in Figure 9. Experiments were repeated for several random masks, but the results did not vary significantly. It is clear that the learned dictionary-based approach greatly outperformed the tensor completion approach. The main reason for this is that the low-rank assumption is not satisfied.

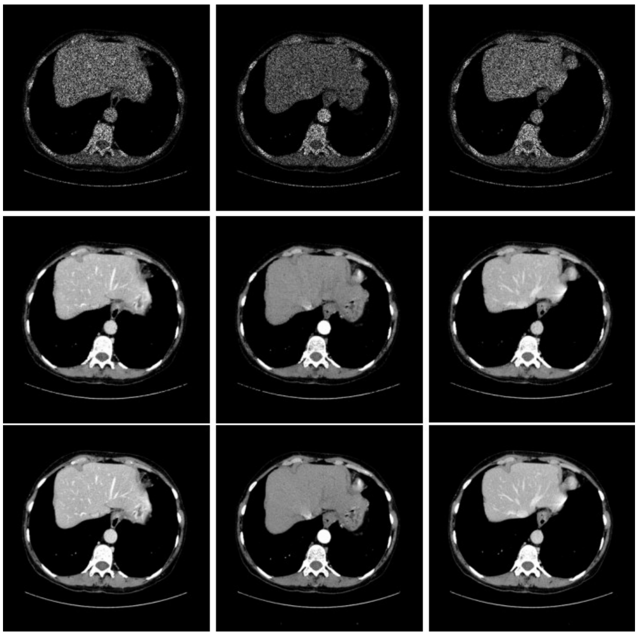

Figure 9. Results of inpainting of 3-phase CT image. Columns correspond to phases of a CT image. Top row: images with 50% missing pixels. Middle row: inpainting results using the tensor completion approach. Third row: inpainting results using the ICA-learned dictionary.

4. CONCLUSION

In this paper we have extended our previous results [1, 2]. Namely, we have shown that, by learning the dictionary on a single image only, comparable results in inpainting can be achieved relative to the case when the dictionary is learned on a training set of images. Detailed comparison of the learned dictionary approach and the tensor completion approach based on minimization of trace norm was presented, both on color (RGB) images of natural scenes and three-phase CT images. The approach to inpainting based on ICA-learned dictionary greatly outperformed the tensor completion approach in all experiments. We conjecture that similar results could also be obtained for other types of imaging modalities, like multispectral magnetic resonance imaging (MRI) or functional MRI.

ACKNOWLEDGMENT

This work has been supported through grant 098-0982903-2558 funded by the Ministry of Science, Education and Sports, Republic of Croatia.

REFERENCES.

[1] M. Filipović, I. Kopriva, "A comparison of dictionary based approaches to inpainting and denoising with an emphasis to independent component analysis learned dictionaries", *Inverse Probl. Imaging,* vol. 5, pp. 815-841, 2011.

[2] M. Filipović, I. Kopriva, A. Cichocki. "Inpainting color images in learned dictionary", 20*th* *European Signal Processing Conference (EUSIPCO 2012)*, Bucharest, Romania, August 27-31, 2012.

[3] J. Liu, P. Musialski, P. Wonka, J. Ye. "Tensor completion for estimating missing values in visual data", *IEEE Trans. Pattern Anal. Machine Intelligence*, to appear.

[4] J. Mairal, G. Sapiro, M. Elad, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, pp. 53-69, 2008.

[5] M. Elad, M. Aharon. "Image denoising via sparse and redundant representations over learned dictionaries", *IEEE Trans. Image Process.*, vol. 15, pp. 3736-3745, 2006.

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro. "Online learning for matrix factorization and sparse coding", *J. Mach. Learn. Res.*, vol. 11, pp. 19-60, 2010.

[7] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, G. Wang. "Low-dose X-ray CT reconstruction via dictionary learning", *IEEE Trans. Medical Imaging*, vol. 31, pp. 1682-1697, 2012.

[8] A. Olmos, F. A. A. Kingdom, McGill calibrated colour image database, 2004.,

http://pirsquared.org/research/mcgilldb/

[9] A. Hyvärinen, E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation,* vol. 9, pp.1483-1492, 1997.

[10] A. Eftekhari, M. Babaie-Zadeh, C. Jutten, H. A. Moghaddam. "Robust-SL0 for stable sparse representations in noisy settings", *Proc. IEEE Conf. Acoustics, Speech and Signal Process.*, 2009.