INPAINTING COLOR IMAGES IN LEARNED DICTIONARY

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

Sparse representation of natural images over redundant dictionary enables solution of the inpainting problem. A major challenge, in this regard, is learning of a dictionary that is well adapted to the image. Efficient methods are developed for grayscale images represented in patch space by using, for example, K-SVD or independent component analysis algorithms. Here, we address the problem of patch space-based dictionary learning for color images. To this end, an image in RGB color space is represented as a collection of vectorized 3D patch tensors. This leads to the state-of-the-art results in inpainting random and structured patterns of missing values as it is demonstrated in the paper.

Index Terms – learned dictionary, independent component analysis, color image, inpainting.

1. INTRODUCTION

Information recovery from incomplete or partially observed data is ubiquitous in biomedical signal processing, computer vision, chemometrics, communication networks, etc. We consider the problem of inpainting color images that can be damaged or corrupted by noise, [1]. It is also possible that certain number of pixels is saturated in some of the color channels, [2, 3]. An image in RGB color space, which is a representation used in this paper, is a 3D tensor. Hence, color image inpainting is a 3D tensor completion problem. One approach to solve this problem is based on the minimization of trace norm (also called nuclear norm) of the matricized tensor [4, 5, 6]. The concept of nuclear norm for tensors appeared for the first time in [4]. Nuclear norm (defined as the sum of the singular values of a matrix) is the tightest convex lower bound of the rank of the matrix on the set of matrices $\{Y: \|Y\|_2 \leq 1\}$ (matrix rank minimization is a non-convex problem). Hence, the nuclear norm minimization assumes that a tensor unfolded in selected mode has low-rank representation. However, fulfilment of the low-rank assumption is data dependent and may fail in some applications. When it comes to RGB color images, experimental checking demonstrates that the rank of an unfolded image tensor in each of the three modes mostly equals tensor dimension in corresponding mode. Thus, for RGB color images, low-rank assumption is rarely satisfied. Moreover, it has been demonstrated in [7] that the minimization of the nuclear norm can lead to multiple solutions of the related matrix completion problem. As shown in [7], the inpainting based on the minimization of the nuclear norm fails to recover a color image damaged by a thick line pattern of missing pixels.

As opposed to the trace norm minimization, an approach that uses sparse representation of an image in learned dictionary enables inpainting of random and structured pattern of missing pixels, as it is demonstrated in [1] for denoising color images and in [8] for inpainting grayscale images. The method in [1] extends the patch space K-SVD algorithm-based dictionary learning, developed and demonstrated in [9] for denoising of grayscale images, to denoising and inpainting of color images. As discussed in [1], denoising of color images requires adaptation of the orthogonal matching pursuit algorithm, used in the dictionary learning process, to overcome artefacts that occur in color image processing. That is due to the fact that 3D patches collected for dictionary learning do not represent well diversity of colors in natural images. In this paper, we propose to use the independent component analysis (ICA) for dictionary learning on a matrix of vectorized 3D patches collected randomly from a training set of color image tensors. Provided that the training set is rich enough, this approach is expected to represent diversity of colors in natural images. Moreover, the dictionary learning (sparse coding) and image reconstruction (inpainting) stages developed for grayscale images, [8, 9], are directly extendable to color images. The color image tensor in patch space representation has to be vectorized before "grayscale-like" processing and the image reconstruction result has to be tensorized. Hence, the focus is not on finding decomposition of a given image tensor, that is the case in [10] where a tensor is decomposed in CP/PARAFAC model, but to recover color image tensor from incomplete data by assuming that it is sparse in the predefined (learned) dictionary. Thus, within the inpainting context, saturated pixels in color images can be declared as missing and recovered without imposing/assuming [2]: (*i*) correlation between the color channels, (*ii*) that most of the pixels are not saturated, and (*iii*) that not all color channels are saturated simultaneously at some pixel location. Since a learned dictionary yields more efficient (sparser) representation than a fixed dictionary, see [8] for details related to the inpainting experiments, the learned dictionary approach proposed here is expected to yield a better reconstruction of the saturated pixel values than the fixed dictionary method proposed in [3].

The rest of the paper is organized as follows. Section 2 presents basics of tensor notation. Dictionary learning in the patch space representation of matricized color images as well as sparseness constrained image reconstruction are presented in section 3. Section 4 presents experimental results related to inpainting of color images with random and thick lines patterns of missing values. Conclusion is presented in section 5.

2. BASICS OF TENSOR NOTATION

Tensor, also called multi-way array, is a generalization of vectors (1D array) and matrices (2D array). The analysis presented in this paper is related to color images that are 3D arrays: $\underline{\mathbf{X}} \in \mathbb{R}_{0+}^{I_1 \times I_2 \times I_3}$, where I_1 and I_2 represent the number of pixels in horizontal and vertical directions respectively, and I_3=3 represents number of spectral (R, G and B) channels. Each tensor index is called way or mode, and the number of levels on a certain mode is called dimension in that mode. This is the standard notation adopted in multi-way analysis [11]. Since the focus of our paper is tensor reconstruction from incomplete data, and not tensor decomposition, we shall not discuss here the tensor models. Details about this topic can be found in [12]. Dictionary learning approach proposed here is based on the representation of the image tensor $\underline{\mathbf{X}}$ in the patch space: $\underline{\mathbf{X}}_p \in \mathbb{R}_{0+}^{\sqrt{l} \times \sqrt{l} \times 3}$, where p represents a patch index and the size of (color) patches is $\sqrt{l} \times \sqrt{l} \times 3$. Each patch is vectorized, yielding a column vector $\mathbf{x}_p \in \mathbb{R}_{0+}^{3l}$. Thus, the vectorization maps an image tensor into a matrix: $\underline{\mathbf{X}} \mapsto \mathbf{X} \in \mathbb{R}_{0+}^{3l \times P}$, where P represents overall number of patches collected from $\underline{\mathbf{X}}$. Hence, a mapped color image $\mathbf{X}$ can be processed by dictionary learning and image reconstruction algorithms already developed for grayscale images, [8, 9], whereas the reconstructed image has to be tensorized.

3. DICTIONARY LEARNING AND IMAGE RECONSTRUCTION IN PATCH SPACE

We cast the inpainting problem in the following mathematical framework. It is assumed that an image is sparse in a dictionary $\mathbf{D}$, that is learned from an ensemble of patches collected randomly from color images belonging to the training set. The assumption of sparsity of color images (or more precisely, image patches) in an appropriate dictionary has already been justified empirically, see for example [1]. Random collection of 3D patches is expected to represent color diversity of natural images, which has been noticed as a problem in [1]. Formally, assuming an image patch $I \in \mathbb{R}^{\sqrt{l} \times \sqrt{l} \times 3}$, the ensemble of T vectorized patches forms a training matrix $\mathbf{Y} \in \mathbb{R}^{n \times T}$, where n=3l. Each column vector represents a vectorized patch, assumed to be sparse in the dictionary $\mathbf{D} \in \mathbb{R}^{n \times m}$: $\mathbf{y}_t \approx \mathbf{D}\mathbf{c}_t$, $\mathbf{c}_t \in \mathbb{R}^m$, $\|\mathbf{c}_t\|_0 \ll m$, $m \geq n$ and $t \in \{1, ..., T\}$. Here, $\|\mathbf{c}_t\|_0$ stands for the ℓ_0-quasi norm that counts the number of nonzero elements in $\mathbf{c}_t$. Hence, learning a dictionary $\mathbf{D}$ is a sparse coding problem. It is implemented through the sparseness constrained factorization of $\mathbf{Y}$=$\mathbf{DC}$, whereas sparseness constraint is imposed on the code $\mathbf{C}$.

Many algorithms can be used to implement sparse coding whereas the most often used are the K-SVD algorithm [9], sparseness constrained nonnegative matrix factorization (NMF) algorithm, [13], and recently ICA algorithm, [8]. Here, as in [8], we shall use the FastICA algorithm, [14], with *tanh* nonlinearity to learn dictionary $\mathbf{D}$. The *tanh* nonlinearity induces the code distributed according to the Laplacian-like probability density function, and that ensures the code $\mathbf{C}$ to be sparse. Moreover, the FastICA algorithm enables, in sequential mode, to learn overcomplete dictionary $\mathbf{D}$. Thus, for the purpose of dictionary learning, sparse coding is interpreted as the blind source separation (BSS) with predefined distribution of the code (sources in BSS vocabulary). Detailed comparative performance analysis related to the inpainting of grayscale images in [8] has demonstrated that FastICA learned dictionary yields comparable or better results than K-SVD algorithm. Before dictionary learning, the training matrix $\mathbf{Y}$ was preprocessed by making every column zero-mean and by multiplying the resulting matrix by the matrix of the form $I + \frac{a}{l}K$, where I denotes $3l \times 3l$ identity matrix, K is a matrix of the form

$$K = \begin{pmatrix} J & 0 & 0 \\ 0 & J & 0 \\ 0 & 0 & J \end{pmatrix}$$

where J is the $l \times l$ matrix of ones, and a is an appropriate constant. In this way, learned basis vectors were forced to take into account the average colors. This idea was taken from the paper [1], where it was used in a somewhat different way, by modifying the OMP algorithm. The constant a was set to $a = \sqrt{\gamma + 1} - 1$, where $\gamma = 5.25$, see [1] for details.

Once the dictionary $\mathbf{D}$ is learned, the actual inpainting is performed according to the following procedure. Let us assume that color image tensor $\underline{\mathbf{X}}$ contains damaged pixels on known locations. After transformation to patch space a vectorized patch with the vector of known pixels $\mathbf{v} \in \mathbb{R}^k$, $k \leq n$, is related to the patch with true but unknown pixels $\mathbf{x} \in \mathbb{R}^n$ through: $\mathbf{v} = \mathbf{Mx} = \mathbf{MDc}$. Here, $\mathbf{M}$ is an indicator matrix that is determined by the layout of missing pixels. Hence, recovery of the patch with true pixels $\mathbf{x}$

can be achieved through the sparseness-constrained minimization:

$$\arg\min_{\mathbf{c}} \left\{ \|\mathbf{c}\|_0 : \|\mathbf{MDc} - \mathbf{v}\|_2^2 \le \varepsilon \right\}$$

where the parameter ε depends on the noise variance. Provided that $k \ge 2\|\mathbf{c}\|_0$ and ε is small, the above problem has the unique solution. However, the above optimization problem is NP-hard, i.e. its computational complexity grows exponential with *m*, which makes it computationally intractable for practical purposes. Computationally feasible solutions are obtained by replacing the ℓ_0-quasi norm of **c** by ℓ_p-norm $\|\mathbf{c}\|_p = \left(\sum_i |c_i|^p\right)^{1/p}$. For $0<p<1$ this yields a non-convex optimization problem. Yet, algorithms that minimize ℓ_p-norm for $p<1$, [15] , outperform ℓ_1-norm minimization in practice. In this paper, as in [8], we have used the method proposed in [16], with MATLAB code available in [17]. The method minimizes smooth approximation of the ℓ_0-quasi norm of **c**: $\|\mathbf{c}\|_0 \approx n - F_\sigma(\mathbf{c})$, where $F_\sigma(\mathbf{c}) = \sum_i f_\sigma(c_i)$ and $f_\sigma(c_i) = \exp\left(-c_i^2/2\sigma^2\right)$ is an approximation of the indicator function of set {0}. The parameter σ regulates how close the approximation is to ℓ_0-quasi norm. Minimization of $\|\mathbf{c}\|_0$ is equivalent to the maximization of $F_\sigma(\mathbf{c})$ for a sequence $\sigma_1>\sigma_2>...>\sigma_K$. This approach outperforms methods based on the minimization of ℓ_1-norm in terms of accuracy and computational efficiency. This is especially the case when the code **c** contains several dominant coefficients and many coefficients with the magnitude close to zero. Such situation occurs in practice for real world signals such as images of natural scenes.

4. EXPERIMENTAL RESULTS

In the examples in this section the dictionary was learned from a matrix composed of vectorized 3D patches collected from 23 color images shown in Figure 1. The images were downloaded from [18]. Tensorized basis vectors (or *atoms*) learned by FastICA algorithm are shown in Figure 2. Most of the atoms is gray (color-less) since the dictionary was learned on a generic image database. These atoms represent spatial structure in images. The colored atoms should represent differences in structure in each of the three (R, G and B) color channels. As shown in [1], dictionary learning on a single image would result in more colored atoms since the dictionary is adapted to a single image. This was the approach taken in [1], where the dictionary was learned on the damaged image itself. However, here we show that comparable result can be obtained with generic dictionary.

The proposed method is first tested on a castle image, Figure 3 top left, with 80% of pixels removed randomly, Figure 3 top right. The same example has been used in Figure 12 in [1], whereas achieved performance in image reconstruction was PSNR=29.65 dB with patches of the size 7×7×3 pixels. There, the K-SVD algorithm was used for dictionary learning and a modified version of the orthogonal matching pursuit algorithm, [19], for image reconstruction. The method proposed herein achieves PSNR of 29.36 dB (average of 5 runs) with patches of the size 8×8×3 pixels, see Figure 3 bottom left. Hence, difference in achieved performance in image reconstruction is small and is, arguably, consequence of different training set used for dictionary learning. When each color channel is treated as a grayscale image, the obtained performance was inferior, with PSNR of 25.05 dB, see Figure 3 bottom right, relative to the case when the color image has been treated as a 3D tensor. The same result has been demonstrated in [1] when OMP method has been used to inpaint each color channel separately in the dictionary learned by the K-SVD algorithm. We have further tested the proposed method on four test images shown in Figure 4. These images were damaged by randomly removing 80% of pixels as well as by using a thick line pattern of missing values, see Figure 5. For images with random pattern of missing values, the inpainting method proposed here has achieved the PSNR performance of 28, 35.1, 31.9 and 35.9 dB, respectively, with respect to top left, top right, bottom left and bottom right images on Figure 4. Regarding the thick line pattern of missing values, it has been shown in [7] that trace norm minimization fails to recover color image with such pattern of missing values. On the contrary, as shown in Figure 6, the learned dictionary based method proposed here yielded images of satisfying visual quality from the images with the thick line pattern of missing values. Achieved PSNR values in the order defined previously were 34.3, 37.2, 39.8 and 37.3 dB.

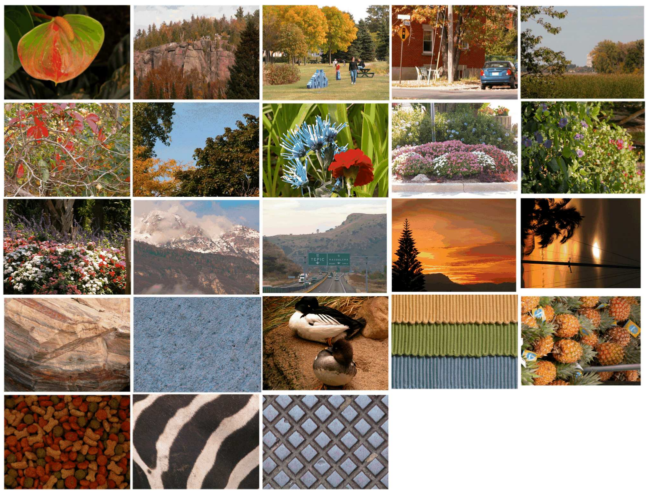

Figure 1. Training images used for dictionary learning.

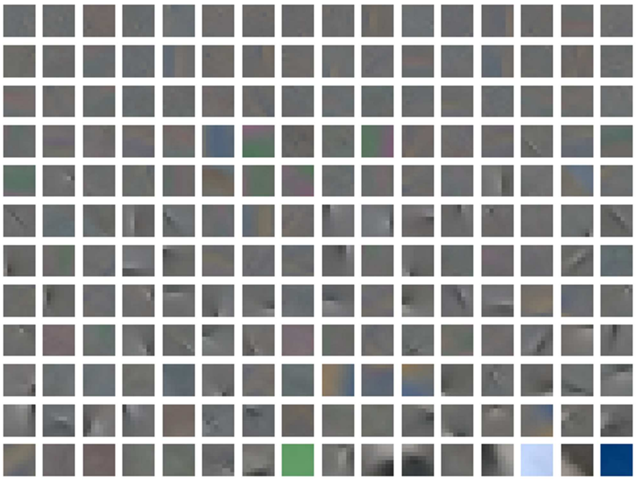

Figure 2. Tensorized basis vectors learned by FastICA algorithm from the training images shown in Figure 1.

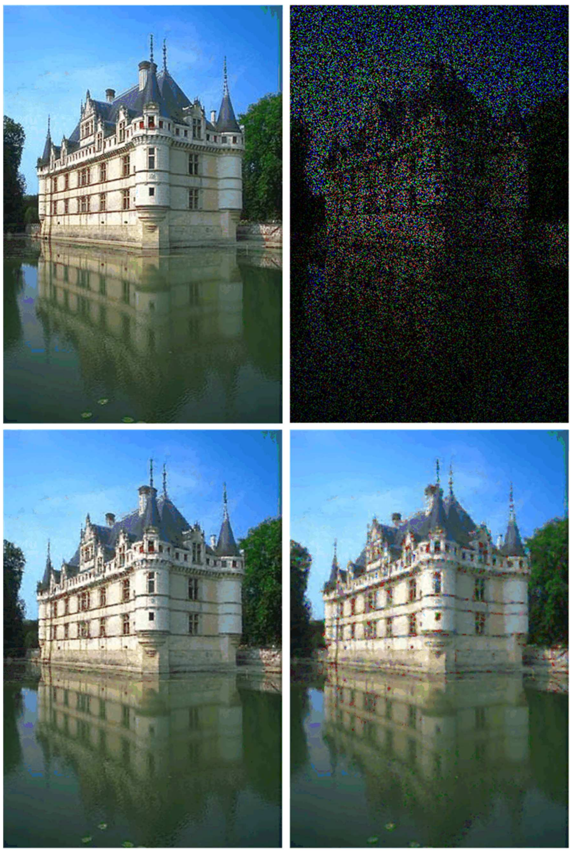

Figure 3. Top left: original image. Top right: 80% of pixels removed randomly. Bottom left: the result of inpainting matricized color image tensor: PSNR=29.36dB. Bottom right: the result of inpainting each spectral image separately as a grayscale image: PSNR=25.05dB.

Figure 4. Four test images.

Figure 5. Test images, shown in Figure 4, with a thick line pattern of missing values.

Figure 6. Images reconstructed by inpainting images shown in Figure 5 by method proposed in the paper.

5. CONCLUSION

Tensor completion for structured pattern of missing values is a challenge for state-of-the-art methods that minimize nuclear norm. Here we have demonstrated that inpainting of color image (3D tensor completion problem) can be accomplished successfully for random and thick line patterns of missing values. That is achieved by sparseness-constrained reconstruction formulated in space comprised of vectorized 3D patches of the image. The patches have sparse representation in dictionary learned in the space of vectorized 3D patch tensors collected randomly from the images in the training set.

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] J. Mairal, G. Sapiro, M. Elad, "Sparse representation for color image restoration," *IEEE Trans. Image Process.*, vol. 17, pp. 53-69, 2008.
[2] X. Zhang, D. H. Brainard, "Estimation of saturated pixel values in digital color imaging," *Journal of the Optical Society of America A*, vol. 21, pp. 2301-2310, 2004.
[3] H. Mansour, R. Saab, P. Nasiopoulos and R. Ward, "Color Image Desaturation Using Sparse Reconstruction," *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, USA, March 14-19, 2010, pp.778-781.
[4] J. Liu, P. Musialski, P. Wonka and J. Ye, "Tensor completion for estimating missing values in visual data," *Proceedings of the 2009 IEEE Conference on Computer Vision*, Kyoto, Japan, Sept. 26-Oct.4, 2009, pp.2114-2121.
[5] S. Gandy, B. Recht, I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, vol. 27, 025010, 2011.
[6] R. Tomioka, K. Hayashi, H. Kashima, "Estimation of Low-Rank Tensors via Convex Optimization," *arXiv:1010.07892v2*, March 2, 2011.
[7] X. Shi, P. S. Yu, "Limitations of Matrix Completion via Trace Norm Minimization," *SIGKD Explorations*, vol. 12, pp.16-20, 2011.
[8] M. Filipović, I. Kopriva, "A Comparison of Dictionary Based Approaches to Inpainting and Denoising with an Emphasize to Independent Component Analysis Learned Dictionaries," *Inverse Problems and Imaging*, vol. 5, pp. 815-841, 2011.
[9] M. Aharon, M. Elad, A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311-4322, 2006.
[10] E. Acar, D. M. Dunlavy, T. G. Kolda, M. Mørup, "Scalable Tensor Factorizations for Incomplete Data," *Chemometrics and Intelligent Lab. Systems*, vol. 106, pp. 41-56, 2011.
[11] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *Journal of Chemometrics*, vol. 14, pp.105-122, 2000.
[12] A. Cichocki, R. Zdunek, A. H. Phan, S. I. Amari, *Nonnegative Matrix and Tensor Factorization*, John Wiley & Sons, 2009.
[13] P. Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints," *Journal of Machine Learning Research*, vol. 5, pp.1457-1469, 2004.
[14] A. Hyvärinen, E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp.1483-1492, 1997.
[15] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Let.*, vol. 14, pp.707-710, 2007.
[16] H. Mohimani, M. Babaie-Zadeh, C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed norm," *IEEE Trans. Signal Process.*, vol. 57, pp. 289-301, 2009.
[17] http://ee.sharif.ir/~SLzero/
[18] A. Olmos, F. A. A. Kingdom, McGill calibrated colour image database, 2004., http://pirsquared.org/research/mcgilldb/
[19] S. Mallat, Z. Zhang, "Matching pursuit in a time-frequency dictionary," *IEEE Trans. Signal Proc.*, vol. 41, pp.3397-3415, 1993.