LETTER

Hole-Filling by Rank Sparsity Tensor Decomposition for Medical Imaging

Lv GUO[†a)], ***Student Member*, Yin LI**[†], **Jie YANG**[†], ***Nonmembers*, *and* Li LU**[†], ***Student Member***

SUMMARY Surface integrity of 3D medical data is crucial for surgery simulation or virtual diagnoses. However, undesirable holes often exist due to external damage on bodies or accessibility limitation on scanners. To bridge the gap, hole-filling for medical imaging is a popular research topic in recent years [1]–[3]. Considering that a medical image, e.g. CT or MRI, has the natural form of a tensor, we recognize the problem of medical hole-filling as the extension of Principal Component Pursuit (PCP) problem from matrix case to tensor case. Since the new problem in the tensor case is much more difficult than the matrix case, an efficient algorithm for the extension is presented by relaxation technique. The most significant feature of our algorithm is that unlike traditional methods which follow a strictly local approach, our method fixes the hole by the global structure in the specific medical data. Another important difference from the previous algorithm [4] is that our algorithm is able to automatically separate the completed data from the hole in an implicit manner. Our experiments demonstrate that the proposed method can lead to satisfactory results.

key words: *tensor analysis, hole-filling, medical image processing*

1. Introduction

Surface integrity of 3D medical image sets is important for modern medical operations such as surgery simulation, model processing and virtual diagnosis [5]–[7]. Unfortunately, undesired holes are often brought to the surface because of imperfection of prototypes or accessibility limitation of scanning devices. Therefore it is of significant importance to design efficient approaches for detecting and filling holes on the surface from 3D medical image sets.

Essentially, the problem of hole-filling is to build up the relationship between the known elements and the unknown ones. So far, the problem is still an open question [1]–[3]. Traditional methods usually follow a strictly local approach, by the assumption that the missing entries mainly depend on their neighbors and the dependency decay fast as the distance increases. However, the global structure indeed plays a key role in estimating the missing entries. Thus, it is advantageous to develop a tool to directly fill holes according to the global information in medical data.

Typically, the medical data, either the CT or the MRI data, has the natural form of a multi-dimensional array, namely the tensor. For the 2D case, i.e. a matrix, the "rank" provides useful cues to capture the global information. The "rank" itself is non-convex, but it can be approximated by its convex envelop, namely the trace norm. Another useful prior for hole-filling is "sparsity", since the hole usually occupies a small portion of the data. It is until recently that much attention have been focused on the rank-sparsity problem [8], [9], [12], namely the Principal Component Pursuit (PCP). These works seek to directly decompose the data into a low-rank part plus a sparse part.

[4] proposes tensor completion for missing value estimation. However, their method is not applicable in medical imaging, since tensor completion requires the locations of missing data while in many real world medical cases, the location of the hole is unknown. Inspired by the work of [4] and [8], and based on the previous work of [12], we extend the PCP problem for matrix case to tensor case as a solution to the problem of hole-filling for medical data. Our solution could automatically detect the location of the missing values and recover the completed data, In addition, the problem of tensor PCP is formulated as a convex optimization as in [12] and a block coordinate descent (BCD) based algorithm is presented to efficiently solve this problem.

The paper is organized as follows: In Sect. 2, we introduce the PCP problem in the tensor case and present our algorithm. In Sect. 3, we show the experiment of our algorithm and provide our analysis on the results. Finally, Sect. 4 concludes the paper.

2. Algorithm Description

In this section, we consider the hole-filling problem as an extension of PCP problem from matrix to tensor. By our method, the original medical data is recognized as a 3-mode tensor and decomposed into a recovered "low-rank" data plus a small (sparse) hole. And we use the relaxation technique towards an efficient solution.

2.1 Hole-Filling by Tensor Decomposition

In general, the medical data (CT or MRI) has the form of a three dimensional array, with each slice as a grey level image. Such a format follows the natural form of a three-mode tensor. Formally, a three-mode tensor is defined as $X \in R^{I_1 \times I_2 \times I_3}$, with its elements $x_{i_1 i_2 i_3} \in R$. A basic operation for the tensor is to convert a tensor into a matrix, also called matricizing or unfolding. The "unfold" operation along the k-th mode on a tensor X is defined as $unfold(X, k) := X_{(k)} \in R^{I_k \times (I_1 \cdots I_{k-1} I_{k+1} \cdots I_3)}$. Accordingly, its inverse operator $fold$ can be defined as $fold\,(X_{(k)}, k) := X$. Moreover, denote $\|X\|_F =$

Manuscript received May 24, 2010.
Manuscript revised October 17, 2010.

[†]The authors are with IPRAI, Shanghai Jiao Tong University, Shanghai, China.
a) E-mail: larrygreen1982@gmail.com

DOI: 10.1587/transinf.E94.D.396

$\left(\sum_{i_1,i_2,i_3} \left|x_{i_1 i_2 i_3}\right|^2\right)^{\frac{1}{2}}$ and $\|X\|_1 := \sum_{i_1,i_2,i_3} \left|x_{i_1 i_2 i_3}\right|$ as the Frobenius norm and l_1 norm of a tensor. Then, we have $\|X\|_F = \left\|X_{(k)}\right\|_F$ and $\|X\|_1 = \left\|X_{(k)}\right\|_1$ for any $1 \le k \le 3$.

The ultimate goal of hole-filling is to identify the hole and recover the completed data. A reasonable assumption is that the original corrupted data is an addition of the completed data and the (possible) hole. Thus, we could seek to a direct decomposition as proposed in [12]. Since the tensor is a higher dimensional extension of matrix, the PCP algorithm can be extended to the tensor case by solving the following optimization:

$$\begin{aligned} \min_{L,S} \quad & \|L\|_{tr} + \lambda \|S\|_1 \\ s.t. \quad & L + S = X \end{aligned} \tag{1}$$

where X, L and S are three-mode tensors with identical size in each mode. X is the original data tensor. L and S represent the correspondent low rank part (the recovered data) and sparse part (the hole) respectively.

However, the notion of trace norm for tensors of three-mode tensor is subtle. For example, there are alternative approaches for tensor decompositions [10], [11], leading to different definition of the trace norm. We propose the following definition for the tensor trace norm as in [4], [12]

$$\|X\|_{tr} := \frac{1}{n} \sum_{i=1}^{n} \left\|X_{(i)}\right\|_{tr} \tag{2}$$

Essentially, the trace norm of a tensor is the average of the trace norms of all matrixes unfolded along each mode. In particular, when the mode $n = 2$ (i.e. the matrix case), this definition is consistent with the matrix trace norm, since $X_{(1)}^T = X_{(2)}$. Under this definition, the optimization in (1) can be rewritten as:

$$\begin{aligned} \min_{L,S} \quad & \sum_{i=1}^{3} \left\|L_{(i)}\right\|_{tr} + \lambda \|S\|_1 \\ s.t. \quad & L + S = X \\ & unfold\left(L_{(i)}, i\right) = L \qquad \forall i \end{aligned} \tag{3}$$

2.2 Simplified Formulation

Problem (3) is hard to solve due to the interdependent trace norm and l_1 norm constraint. To simplify the problem, we adopt similar optimization scheme as in [12] by introducing additional auxiliary matrix $M_i = L_{(i)}$ and $N_i = S_{(i)}$. Thus, the formulation is rewritten as

$$\begin{aligned} \min_{L,S,M_i} \quad & \sum_{i=1}^{3} \|M_i\|_{tr} + \lambda \sum_{i=1}^{3} \|N_i\|_1 \\ s.t. \quad & L + S = X \\ & M_i = L_{(i)} \qquad N_i = S_{(i)} \quad \forall i \end{aligned} \tag{4}$$

In (4), the constraints $M_i = L_{(i)}$ and $N_i = S_{(i)}$ still enforce the consistency of all M_i and N_i. Thus, we could further relax the equality constrains $M_i = L_{(i)}$ and $N_i = S_{(i)}$ by $\left\|M_i - L_{(i)}\right\|_F \le \varepsilon_1$ and $\left\|N_i - S_{(i)}\right\|_F \le \varepsilon_2$. Furthermore, if we allow a dense noise term over X, we can relax $L + S = X$ by $\|L + S - X\|_F \le \varepsilon_3$. Therefore, we get the relaxed form:

$$\begin{aligned} \min_{L,S,M_i} \quad & \sum_{i=1}^{3} \|M_i\|_{tr} + \lambda \sum_{i=1}^{3} \|N_i\|_1 \\ s.t. \quad & \left\|M_i - L_{(i)}\right\|_F \le \varepsilon_1 \qquad \forall i \\ & \left\|N_i - S_{(i)}\right\|_F \le \varepsilon_2 \qquad \forall i \\ & \|L + S - X\|_F \le \varepsilon_3 \end{aligned} \tag{5}$$

For certain α_i, β_i and γ_i, (5) can be converted to its equivalent form by Lagrange multiplier.

$$\begin{aligned} \min_{L,S,M_i} \quad & \frac{1}{2} \sum_{i=1}^{3} \alpha_i \left\|M_i - L_{(i)}\right\|_F \\ & + \frac{1}{2} \sum_{i=1}^{3} \beta_i \left\|N_i - S_{(i)}\right\|_F \\ & + \frac{1}{2} \sum_{i=1}^{3} \gamma_i \left\|M_i + N_i - X_{(i)}\right\|_F \\ & + \sum_{i=1}^{3} \|M_i\|_{tr} + \lambda \sum_{i=1}^{3} \|N_i\|_1 \end{aligned} \tag{6}$$

Intuitively, the weights α_i, β_i and γ_i indicate the preference towards different 'unfold' operation, i.e. the configuration of the tensor. For example, we would prefer to explain the tensor representation of the CT data as the collection of slices. The optimization problem in (6) is convex but non-differentiable. Next, we focus on its solution.

2.3 The Proposed Algorithm

The alternative direction strategy is presented for the optimization (6), leading to a block coordinate descent (BCD) algorithm. The core idea of the BCD is to optimize a group of variables while fixing the other groups. Thus, to achieve the optimum solution, we estimate N_i, M_i, L and S sequentially in each iteration. For clarity, we first define the "shrinkage" operator $D_\tau(x)$ by

$$D_\tau(x) = \max\{\mathrm{sgn}(x)(|x| - \tau), 0\} \tag{7}$$

where $sgn(x)$ is the sign function. The operator is extended to the tensor case by performing the shrinkage towards each element.

Computing N_i: The optimal N_i with all other variables fixed is the solution to the following sub-problem

$$\begin{aligned} \min_{N_i} \quad & \frac{\beta_i}{2} \left\|N_i - S_{(i)}\right\|_F \\ & + \frac{\gamma_i}{2} \left\|M_i + N_i - X_{(i)}\right\|_F + \lambda \|N_i\|_1 \end{aligned} \tag{8}$$

By the well-known l_1 minimization [13], the global minimum of the optimization problem in (8) is given by

$$N_i^* = D_{\frac{\lambda}{\beta_i+\gamma_i}}\left(\frac{\beta_i S_{(i)} + \gamma_i \left(X_{(i)} - M_i\right)}{\beta_i + \gamma_i}\right) \tag{9}$$

where D_τ is the "shrinkage" operation.

Computing M_i: The optimal M_i with all other variables fixed is the solution to the following sub-problem

$$\begin{aligned}\min_{M_i} \quad & \frac{\alpha_i}{2}\left\|M_i - L_{(i)}\right\|_F \\ & + \frac{\gamma_i}{2}\left\|M_i + N_i - X_{(i)}\right\|_F + \|M_i\|_{tr}\end{aligned} \tag{10}$$

From current trace norm minimization literature [14], the global minimum of the optimization problem in (10) is given by

$$M_i^* = UD_{\frac{1}{\alpha_i+\gamma_i}}(\Lambda)\,V^T \tag{11}$$

where $U\Lambda V^T$ is the singular value decomposition given by

$$U\Lambda V^T = \frac{\alpha_i L_{(i)} + \gamma_i \left(X_{(i)} - N_i\right)}{\alpha_i + \gamma_i} \tag{12}$$

Computing S_i: The optimal S with all other variables fixed is the solution to the following sub-problem

$$\min_S \quad \frac{1}{2}\sum_{i=1}^{n}\beta_i \left\|N_i - S_{(i)}\right\|_F \tag{13}$$

It is easy to show that the solution to (13) is given by

$$S^* = \frac{\sum_{i=1}^{n}\beta_i fold\,(N_i, i)}{\sum_{i=1}^{n}\beta_i} \tag{14}$$

Computing L_i: The optimal L with all other variables fixed is the solution to the following sub-problem

$$\min_L \quad \frac{1}{2}\sum_{i=1}^{n}\alpha_i \left\|M_i - L_{(i)}\right\|_F \tag{15}$$

Similar to (13), the solution to (15) is given by

$$L^* = \frac{\sum_{i=1}^{n}\alpha_i fold\,(M_i, i)}{\sum_{i=1}^{n}\alpha_i} \tag{16}$$

The algorithm is called Rank Sparsity Tensor Decomposition (RSTD) [12]. The pseudo-code of RSTD is summarized in Algorithm 1. We choose the difference of L and S in consecutive iterations as the stopping criterion. We can further show that BCD for RSTD is guaranteed to reach the global optimum, since the first three terms in (6) are differentiable and the last two terms are separable [15].

Algorithm 1 (RSTD: Rank Sparsity Tensor) Decomposition)
Input: X
Output: L, S
1. **Set** $L = X$, $S = 0$, $M_i = L_{(i)}$, $N_i = 0$
2. **while no convergence**
3. **for** $i = 1$ **to** 3
4. $N_i^* = D_{\frac{\lambda}{\beta_i+\gamma_i}}\left(\frac{\beta_i S_{(i)} + \gamma_i (X_{(i)} - M_i)}{\beta_i + \gamma_i}\right)$
5. $M_i^* = UD_{\frac{1}{\alpha_i+\gamma_i}}(\Lambda)\,V^T$
6. **end for**
7. $S^* = \frac{\sum_{i=1}^{n}\beta_i fold\,(N_i, i)}{\sum_{i=1}^{n}\beta_i}$
8. $L^* = \frac{\sum_{i=1}^{n}\alpha_i fold\,(M_i, i)}{\sum_{i=1}^{n}\alpha_i}$
9. **end while**

3. Experiments

We have implemented the hole-filling algorithm using C++ and tested the algorithm on different real world medical data. Two of the cases are shown in Fig. 1. Due to the fact that a good initialization would greatly benefit the efficiency of the iterations, the local hole-filling algorithm [1] is first performed for the initialization of our method. Moreover, for better visual effect, a surface reconstruction algorithm [16] is used to re-mesh the model after the tensor data is completed. Since deformation often exists on the surrounding region of the hole in medical images, leading to deformed result in recovered data, a point-deletion algorithm [17] is used to eliminate the influence by removing the deformed points from the surrounding region.

The top two rows show the result of RSTD on a CT data and the bottom two rows show the results on a MRI data. From the experiments, we can see that the algorithm can blindly separate a reasonable low-rank completed data from the hole. To highlight the effect of RSTD, we show the result data from different viewpoints in the first and fourth rows. Notice that the hole in the upper rows is fairly large (approximately 120 cm^2) and the one in the lower rows locates in the region of high curvature (brow ridge). These two kinds of data are representative in real world application and our algorithm demonstrates a surprising result.

Since the proposed method follows a global approach, the algorithm deals with all data in each iterations, yielding

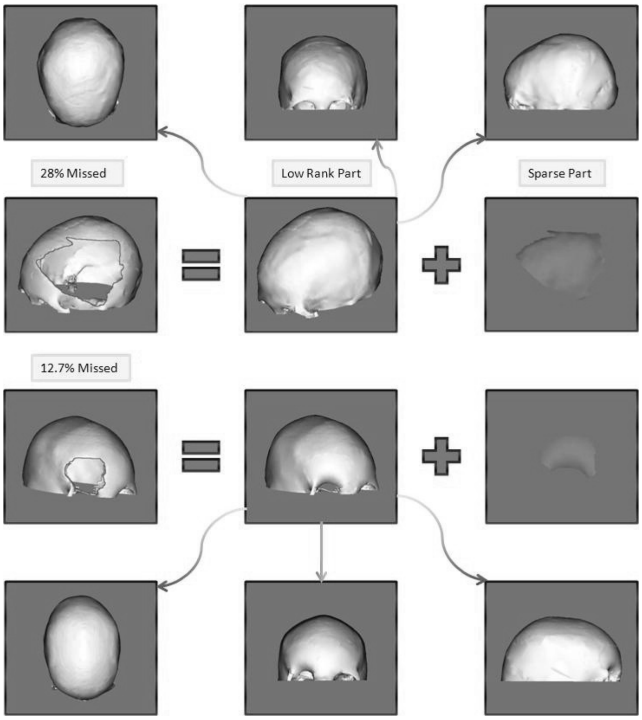

Fig. 1 Hole-filling by RSTD for medical imaging. The top two rows show the result on a CT data and the bottom two rows show the result on a MRI data. To highlight the effect of our algorithm, we show the result data from different viewpoints in the first and fourth rows.

a slow solution. However, during the experiments, we find that for most of the cases, the algorithm is able to converge reasonably fast in dozens of iterations with the tolerance of 1×10^{-6}.

4. Conclusion and Future Work

In this paper, we propose a hole-filling solution for medical imaging by extending the Principal Component Pursuit to the tensor case and designing a highly efficient algorithm for the extension. To our best knowledge, we are the first to use global information for medical hole-filling problem. Both the theoretical derivation and practical experiment have demonstrated that the extension is suitable for filling holes in medical imaging. We are working on the efficiency of our algorithm, e.g. numerical approximations of a few largest singular values, since large-scale full SVD is slow and unnecessary. We would also like to evaluate the effectiveness of the different methods on synthetic data in future work.

References

[1] J.S. Suri, K. Liu, S. Singh, S.N. Laxminarayan, X. Zeng, and L. Reden, "Shape recovery algorithms using level sets in 2-d/3-d medicalimagery: A state-of-the-art review," IEEE Trans. Inf. Technol. Biomed., vol.6, no.1, pp.8–28, 2002.
[2] R.M. Willett and R.D. Nowak, "Platelets: A multiscale approach for recovering edges and surfaces in photon-limited medical imaging," IEEE Trans. Med. Imaging, vol.22, no.3, pp.332–350, 2003.
[3] B.J. Liu, M.Z. Zhou, and J. Documet, "Utilizing data grid architecture for the backup and recovery of clinical image data," Computerized Medical Imaging and Graphics, vol.29, no.2-3, pp.95–102, 2005.
[4] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," Proc. ICCV, Oct. 2009.
[5] J. Berkley, G. Turkiyyah, D. Berg, M. Ganter, and S. Weghorst, "Real-time finite element modeling for surgery simulation: An application to virtual suturing," IEEE Trans. Vis. Comput. Graphics, vol.10, no.3, pp.314–325, 2004.
[6] L.C. Hieu, E. Bohez, J. Vander Sloten, P. Oris, H.N. Phien, E. Vatcharaporn, and P.H. Binh, "Design and manufacturing of cranioplasty implants by 3-axis cnc milling," Technology and Health Care, vol.10, no.5, pp.413–423, 2002.
[7] M. Klein and C. Glatzer, "Individual cad/cam fabricated glass-bioceramic implants in reconstructive surgery of the bony orbital floor," Plastic and Reconstructive Surgery, vol.117, no.2, p.565, 2006.
[8] J. Wright, A. Ganesh, S. Rao, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," Proc. Conference on Neural Information Processing Systems (NIPS), 2009.
[9] E.J. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," Arxiv preprint arXiv:0912.3599 2009.
[10] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," SIAM J. Matrix Anal. Appl., vol.21, no.4, pp.1253–1278, 2000.
[11] M. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," Computer Vision—ECCV 2002, pp.447–460, 2002.
[12] Y. Li, Y. Zhou, J. Yan, and J. Yang, "Tensor error correction for corrupted values in visual data," IEEE International Conference on Image Processing, 2010.
[13] E.T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for l1 minimization: Methodology and convergence," SIAM J. Optimization, vol.19, p.1107, 2008.
[14] J.F. Cai, E.J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," SIAM J. Optimization, vol.20, p.1956, 2010.
[15] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," J. Optim. Theory Appl., vol.109, no.3, pp.475–494, 2001.
[16] R. Kolluri, J.R. Shewchuk, and J.F. O'Brien, "Spectral surface reconstruction from noisy point clouds," Proc. Eurographics/ACM SIGGRAPH Symposium on Geometry Processing, 2004.
[17] M.A. Mostafavi, C. Gold, and M. Dakowicz, "Delete and insert operations in voronoi/delaunay methods and applications," Computers & Geosciences, vol.29, no.4, pp.523–530, 2003.