# Pattern Visualization and Recognition Using Tensor Factorization for Early Differential Diagnosis of Parkinsonism

Rui Li[1,6], Ping Wu[2], Igor Yakushev[3], Jian Wang[4], Sibylle I. Ziegler[3],
Stefan Förster[3], Sung-Cheng Huang[5], Markus Schwaiger[3], Nassir Navab[1],
Chuantao Zuo[2], and Kuangyu Shi[3(✉)]

[1] Department of Computer Science,
Technische Universität München, Munich, Germany
[2] Huashan Hospital, PET Center, Fudan University, Shanghai, China
[3] Department of Nuclear Medicine,
Technische Universität München, Munich, Germany
k.shi@tum.de
[4] Department of Neurology, Huashan Hospital, Fudan University, Shanghai, China
[5] Department of Molecular and Medical Pharmacology,
University of California, Los Angeles, USA
[6] Alibaba Cloud, Hangzhou, China

**Abstract.** Idiopathic Parkinsons disease (PD) and atypical parkinsonian syndromes may have similar symptoms at the early disease stage. Pattern recognition on metabolic imaging has been confirmed of distinct value in the early differential diagnosis of Parkinsonism. However, the principal component analysis (PCA) based method ends up with a unique probability score of each disease pattern. This restricts the exploration of heterogeneous characteristic features for differentiation. There is no visualization of the underlying mechanism to assist the radiologist/neurologist either. We propose a tensor factorization based method to extract the characteristic patterns of the diseases. By decomposing the 3D data, we can capture the intrinsic characteristic pattern in the data. In particular, the disease-related patterns can be visualized individually for the inspection by physicians. The test on PET images of 206 early parkinsonian patients has confirmed differential patterns on the visualized feature images using the proposed method. Computer-aided diagnosis based on multi-class support vector machine (SVM) shown improved diagnostic accuracy of Parkinsonism using the tensor-factorized feature images compared to the state-of-the-art PCA-based scores [Tang et al. Lancet Neurol. 2010].

## 1 Introduction

Atypical parkinsonian syndromes, including multiple system atrophy (MSA) and progressive supranuclear palsy (PSP) present very similar clinical symptoms to idiopathic Parkinson's disease (PD) especially in their early stages [10]. It

has been reported that approximately 20%–30% of PD patients were misdiagnosed [11]. This diagnostic error has significant consequences for clinical patient care and causing inadequate treatment [6]. Positron emission tomography (PET) detect abnormal functional alterations [2,7,12,17] far before structural damages to the brain tissue are present [3,12,18,20]. Although $^{18}$F-FDG PET is effective in the diagnosis of parkinsonism by visualizing brains glucose metabolism [8], the complex spatial abnormalities make the exploration of its differential potential challenging. Sophisticated pattern recognition has been developed to boost the performance in early differential diagnosis of parkinsonism [5,16]. Specific metabolic patterns of PD, MSA and PSA were extracted using principal component analysis (PCA). For each pattern, a score is derived to describe the probability of the PET images of an individual subject presenting this pattern. These pattern scores have been found as surrogates to accurately discriminate between the different types of parkinsonian syndromes in early stage. However, the unique probability score of each disease pattern provides limited information of the heterogeneity of the underlying pathophysiological abnormalities, leading to bottleneck to further improve the diagnostic accuracy. Furthermore, there is no visualization of the characteristic features, which restricts the possibility of diagnostic inspection and approval by radiologist/neurologists.

In contrast to vector-based dimension reduction of PCA, tensor factorization can represent data by a two-dimensional matrix directly and provides a powerful tool for factor analysis. Various successful applications have been reported. A work [15] extracts features from a tensor by higher order discriminant analysis. The class discriminant information in the high dimensional data is captured by the within-class scatter matrix and between-class scatter matrix, which can be seen as an extension to the well-known linear discriminant analysis (LDA). An error analysis [4] of the tensor decomposition was proposed to provide error bounds. The experiments showed improved performance on the video compression and classification. Another work [13] suggested the tensor factorization for context-aware collaborative filtering. A recent study [14] applied tensor factorization to find meaningful latent variables that predict brain activity with competitive accuracy.

In this paper, we propose a tensor factorization based method to extract the characteristic patterns of PD, MSA and PSA. This is achieved by decomposing the 3D data into 2D planes containing the determinant information. The pattern related features can be then represented in the 2D visual space. Thus it can be visualized individually for the inspection by physicians. The method was tested on $^{18}$F-FDG PET images of 206 patients suspected with parkinsonism. The computer-aided diagnosis on the derived 2D feature images were compared with that on the state-of-the-art PCA-based pattern scores [16].

## 2   Methods

### 2.1   Introduction to Tensor Factorization

In this study, we denote the imaging data as a order-3 tensor $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$, where $I, J, K$ are the dimensions. $i, j, k$ can take on the specific value in the $I, J$

and $K$ respectively. Thus, $\mathcal{T} = [t_{ijk}]_{i,j,k=1}^{I,J,K}$, where $t_{i,j,k}$ is an entry of the tensor at the position $(i, j, k)$. CANDECOMP/PARAFAC (CP) [9] decomposition[1] is employed to decomposing the tensor in this study [19]. With respect to an order-3 tensor $\mathcal{T}$, TF factorizes $\mathcal{T}$ into three components bases (factor matrices $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$), which are often constrained to a unit length vector associated with a weight vector $\lambda = [\lambda_1, \dots, \lambda_R]$. These rank-one components can re-express the original tensor. Thus, the factorized form is as follows:

$$\mathcal{T} \approx \sum_{r=1}^{R} \lambda_r \times a_r \circ b_r \circ c_r, \tag{1}$$

in which the component basis $\mathcal{A} = [a_1, a_2, \dots, a_r]$, $\mathcal{B} = [b_1, b_2, \dots, b_r]$, $\mathcal{C} = [c_1, c_2, \dots, c_r]$. The weight vector is $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_r]$. Outer product is denoted by the $\circ$, and $R$ is the rank. After the CP tensor factorization, the factorized model can then be compactly expressed as $\mathcal{T} \approx \{\boldsymbol{\lambda}, \mathcal{A}, \mathcal{B}, \mathcal{C}\}$. By reducing the 3D tensor into a tensor with smaller dimension (such as 2D matrix), we can visualize a 3D image by a 2D representation. As a higher dimensional extension to the singular value decomposition (SVD), TF can be solved by the CP-ALS (alternating least squares). The idea of CP-ALS is to minimize the least square term $\min_{\mathcal{A}, \mathcal{B}, \mathcal{C}} \|\mathcal{T} - \sum_{r=1}^{R} \lambda_r \times a_r \circ b_r \circ c_r\|$, where $\|\cdot\|$ is the vector 2-norm. We used the Matlab tensor toolbox [1] implementation in the experiments.
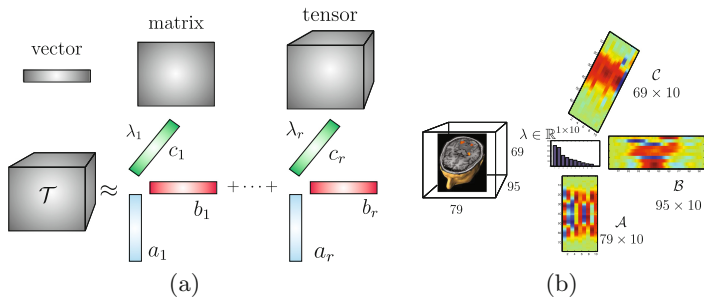


**Fig. 1.** (a) Illustration of vector, matrix, tensor and CP decomposition. (b) Demonstration of factor matrices and associated weights by $R = 10$ in Eq. 1.

In Algorithm 1, the data reduction of $\mathcal{T} \Rightarrow \mathcal{T}'$ is possible because the model $\mathcal{M}_{\mathcal{T}}$ can be expressed by the product of the decomposed factor matrices sharing a common dimension size of $R^2$. Furthermore, we can select the top-$m$ components from the $R$ to reduce the data. Thus, reducing the data by using the top-$m$ bases in the factor matrices is mathematically feasible. We, in fact, can shrink the data

---

[1] We name CP decomposition as tensor factorization (TF) in this work.

[2] The $R$ was chosen by gradually reducing the rank number until the minimal lambda is not smaller than 5% of the maximum lambda after decomposition.

---

**Algorithm 1**: Learning features in high dimensional data via tensor factorization

---

1 **Inputs**: Tensor data $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$, rank $R$, top-$m$ components ($m \leq R$)

2 **Outputs**: Feature vector for each 3D image

3 **Model training**: Train a tensor model $\mathcal{M}_{\mathcal{T}}$ using a given rank $R$ for each class MSA, PSP and PD respectively

4 **Deriving features**: Given an image, use the $\mathcal{M}_{\mathcal{T}}$ to reduce the data along the second mode $J$, such that $\mathcal{T} \in \mathbb{R}^{I \times J \times K} \Rightarrow \mathcal{T}' \in \mathbb{R}^{I \times p \times K}$, where $m \leq R \leq J$. Flatten the $\mathcal{T}'$ into a one dimensional vector of size $I \cdot m \cdot K$ that is used as a feature vector to train a classification classifier such as support vector machines (SVM).

---
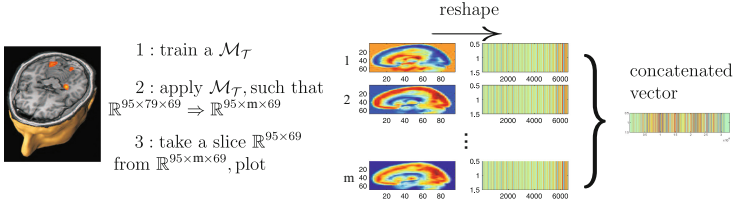


**Fig. 2.** Demonstration of feature learning via tensor factorization. The 2D image shows the third (middle) slice from the reduced image $\mathbb{R}^{95 \times m \times 69}$. m is the number of selected top-$m$ basis components from the model $\mathcal{M}_{\mathcal{T}}$. In this study, we set $m = 1$ for the purpose of visualization, although the proposed method is valid for any value of $m$.

along any mode. In this study, the reduction was performed along the second mode $J$, corresponding to sagittal plane.

Figure 2 depicts the feature learning process using the proposed tensor factorization (TF) method. In our experiments, we train a TF model for each disease type (MSA, PSP and PD) using the training images, resulting in three models denoted as $\mathcal{M}_{\mathcal{T}}^{\text{MSA}}$, $\mathcal{M}_{\mathcal{T}}^{\text{PSP}}$ and $\mathcal{M}_{\mathcal{T}}^{\text{PD}}$. Given a test image (new coming subject), we we just need to project the data to the established factorized bases to derive the respective feature vectors (illustrated as right hand side images in Fig. 2), which are then concatenated together as a one dimensional feature vector representing the test image. This is similar to the state-of-the-art method, where new data only need to be projected to the established PCA-bases. Thus, an image is characterized by the three trained models (factorized bases). Finally, a classifier can be trained based on the derived feature vectors to arrive at a predictive model.

## 2.2    Application of Tensor Factorization to 3D Images

In Fig. 2, the model $\mathcal{M}_{\mathcal{T}}$ is trained by the training data. $\mathcal{M}_{\mathcal{T}}$ consists the factor matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ that are the keys to perform data reduction. To reduce an image of size $95 \times 79 \times 69$ along the second dimension 79, we iterate over the
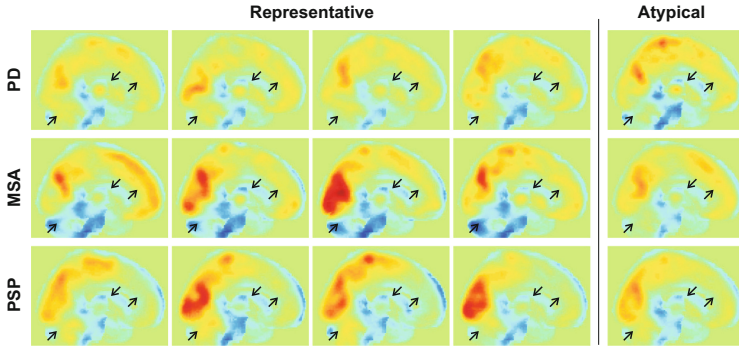
**Fig. 3.** Illustration of feature images of PD, MSA and PSP. 4 representative feature images and 1 atypical feature image for each disease are displayed.

third dimension 69. In each iteration, a 2D image of size $95 \times 79$ is selected, which is multiplied with the second factor matrix of size $79 \times p$. The resulting matrix is of size $95 \times m$ as shown in Fig. 2. After 69 iterations, a reduced 3D image $(95 \times m \times 69)$ is generated. The reduced 3D data is concatenated to form a feature vector representing the image. One may also iterate over the first dimension to perform the reduction, as long as the matrix operation is allowed. Two additional points must be stated: First, it is possible to further reduce the image to an even smaller size along other dimensions. Second, it is also possible to reduce the image by any dimension rather than the second dimension. In this work, we chose the second dimension (sagittal plane) to have the best visualization of all interested structures of 3D data reduction by tensor factorization, with $m$ set to one (second dimension is reduce to one).

## 3   Experiments and Results

206 patients with suspected parkinsonian clinical features were subjected to an $^{18}$F-FDG PET imaging. After the imaging, these patients were assessed by blinded movement disorders specialists for a mean of 2.1 years before a final clinical diagnosis of PD $(n = 136)$, MSA $(n = 40)$, and PSP $(n = 30)$ were made. PET images were normalized by global mean and then were spatially normalized to Montreal Neurological Institute (MNI) space using SPM8[3]. A Gaussian kernel (size $8 \times 8 \times 8$ mm) were applied to smooth the PET images. The mean of each image will be subtracted from the image.

A group of 20 PD, 20 MSA and 20 PSP images were randomly selected to generate mean image for tensor factorization. The factorization algorithm generated a set of base images (factor matrices). Afterwards, the PET image of each patient was projected to the factorized bases to generate 2D feature images. These feature images represent the characteristic patterns and were displayed for visual

---

[3] Statistical Parametric Mapping, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/, 2009.

inspection. For most patients the clinical diagnosis of PD, MSA or PSP can be visually differentiated. Figure 3 shows 4 representative feature images of PD, MSA and PSP. For representative PD patterns, no difference between frontal, parietal and occipital lobes was observed. Visible cerebellum and striatum activities can be observed. For MSA, vanishing cerebellum activity was observed and there were also reduced activities in striatum on the pattern images. For representative PSP, decreasing striatum activities were observed, while cerebellum activities were still visible. There were observable declining activities in frontal lobe. However, these typical findings do not represent all the images. Overall, 13 (9.6%) PD, 5 (12.5%) MSA and 6 (20%) PSP were found to be ambiguous in visual inspection. Examples of atypical pattern images were illustrated in Fig. 3.
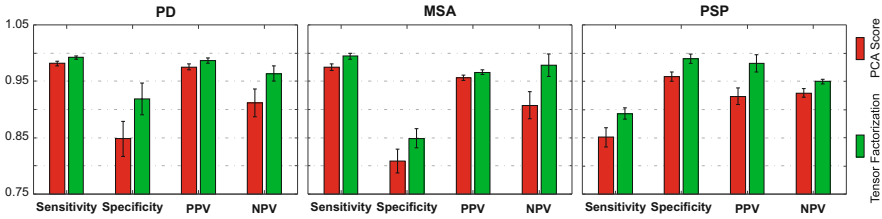


**Fig. 4.** Comparing the computer-aided diagnosis results on PCA-based pattern score [16] and the proposed tensor factorized feature images.

In addition to visual inspection, multi-class SVM were applied to the feature images. A linear kernel was used, with a grid search for parameter optimization. Grid search considers only the optimization of the penalty parameter $C$ in the linear SVM, selecting the value of $C$ yielding the best classification result based on the training data. After the best value of $C$ was found, we applied it to the test data. A 10-fold cross-validation was applied. The tensor factorization and cross-validation were repeated for 256 times. The results were compared with the conventional PCA-based pattern scores with the same setting of cross validation. Figure 4 displays the comparison of the proposed tensor factorization and pattern scores for the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Paired t-test of the results of 256 repeated tests show that visual representations of parkinsonian patterns lead to significant (p of all the comparisons are less than $1.8 \times 10^{-59}$) improvement compared to pattern scores.

## 4     Discussions and Conclusion

This paper proposed a new pattern representation method using tensor factorization, which allows both visual inspection and computer-aided diagnosis. The visualization of the derived feature images demonstrates different representative patterns for PD, MSA and PSP. This provides the potential for physicians to

inspect the diagnosis in reading room. The improved result of computer-aided diagnosis using the tensor-factorized patterns over the PCA-based scores confirmed that the new method can capture more characteristic features for differential diagnosis.

In this study, only one sagittal representation image was chosen for the visualization. This is based on the consideration that it may give a view covering maximum information of reported characteristic anatomical structures of parkinsonism, such as striatum, cerebellum and brainstem [16]. The generation of more anatomical planes and the increase of number of feature images can further increase the performance of SVM-based computer-aided diagnosis. A test of including 5 feature images ($m = 5$) for computer-aided diagnosis can overall improve 2.7% of specificity for PD and 0.5% of specificity for MSA. However, the increased number of images may incur additional burden for neurologist to resolve the critical information in the diagnosis. Further clinical test needs to be made to find optimized number of pattern images considering both the feasibility of visual inspection and accuracy of computer-aided diagnosis. Furthermore, the derived pattern images after tensor factorization has limited anatomical correspondence and the visual inspection may be different from conventional anatomy-guided diagnosis. Special training of the physicians is necessary to make it feasible in clinical practice. Nevertheless, the evolving applicability of PCA-based pattern scores after a series of international clinical trials makes a good example for clinical translation of the proposed concepts. Considering the high challenge of early differential diagnosis of parkinsonism, the exploration of more characteristic features for both visual and computer-aided diagnosis may change the state of art of parkinsonism management.

# References

1. Brett W.B., Tamara, G.K., et al.: Matlab tensor toolbox version 2.6., February 2015
2. Bi, L., Kim, J., Feng, D., Fulham, M.: Multi-stage thresholded region classification for whole-body PET-CT lymphoma studies. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 569–576. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_71
3. Chen, K., Langbaum, J.B., Fleisher, A.S., Ayutyanont, N., Reschke, C., Lee, W., Liu, X., Bandy, D., Alexander, G.E., Thompson, P.M., Foster, N.L., Harvey, D.J., de Leon, M.J., Koeppe, R.A., Jagust, W.J., Weiner, M.W., Reiman, E.M.: Twelve-month metabolic declines in probable alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the alzheimer's disease neuroimaging initiative. Neuroimage **51**(2), 654–664 (2010)

4. Huang, H., Ding, C., Luo, D.J.: Tensor reduction error analysis-applications to video compression and classification. In: IEEE Conference on Computer Vision and Pattern Recognition (2008)

5. Eidelberg, D.: Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. Trends Neurosci. **32**(10), 548–557 (2009)

6. Fahn, S., Oakes, D., Shoulson, I., Kieburtz, K., Rudolph, A., Lang, A., Olanow, C.W., Tanner, C., Marek, K.: Levodopa and the progression of Parkinson's disease. New Engl. J. Med. **351**(24), 2498–2508 (2004)

7. Gao, F., Liu, H., Shi, P.: Patient-adaptive lesion metabolism analysis by dynamic PET images. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7512, pp. 558–565. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33454-2_69

8. Hellwig, S., Frings, L., Amtage, F., Buchert, R., Spehl, T.S., Rijntjes, M., Tuscher, O., Weiller, C., Weber, W.A., Vach, W., Meyer, P.T.: 18f-FDG PET is an early predictor of overall survival in suspected atypical Parkinsonism. J. Nucl. Med. **56**(10), 1541–1546 (2015)

9. Hitchcock, F.L.: The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. **6**(1), 164–189 (1927)

10. Hughes, A.J., Ben-Shlomo, Y., Daniel, S.E., Lees, A.J.: What features improve the accuracy of clinical diagnosis in parkinson's disease: a clinicopathologic study. Neurology **57**(10 Suppl 3), S34–S38 (2001)

11. Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y., Lees, A.J.: The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. Brain **125**(Pt 4), 861–870 (2002)

12. Jiao, J., Searle, G.E., Tziortzi, A.C., Salinas, C.A., Gunn, R.N., Schnabel, J.A.: Spatio-temporal pharmacokinetic model based registration of 4D PET neuroimaging data. NeuroImage **84**, 225–235 (2014)

13. Amatriain, X., Baltrunas, L., Karatzoglou, A., Oliver, N.: Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: Proceedings of the Fourth ACM conference on Recommender Systems, pp. 79–86 (2010)

14. Mitchell, T.M., Papalexakis, E.E., et al.: Turbo-SMT: accelerating coupled sparse matrix-tensor factorizations by 200x. In: SIAM International Conference on Data Mining. SIAM (2014)

15. Phan, A.-H., Cichocki, A.: Tensor decompositions for feature extraction and classification of high dimensional datasets. Nonlinear Theor. Appl. IEICE **1**(1), 37–68 (2010)

16. Tang, C.C., Poston, K.L., Eckert, T., Feigin, A., Frucht, S., Gudesblatt, M., Dhawan, V., Lesser, M., Vonsattel, J.P., Fahn, S., Eidelberg, D.: Differential diagnosis of Parkinsonism: a metabolic imaging study using pattern analysis. Lancet Neurol. **9**(2), 149–158 (2010)

17. Xu, Z., Bagci, U., Seidel, J., Thomasson, D., Solomon, J., Mollura, D.J.: Segmentation based denoising of PET images: an iterative approach via regional means and affinity propagation. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8673, pp. 698–705. Springer, Cham (2014). doi:10.1007/978-3-319-10404-1_87

18. Zhang, D., Shen, D.: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. Neuroimage **59**(2), 895–907 (2012)

19. Zhou, G.X., Zhao, Q.B., Cichocki, A., Zhang, Y., Wang, X.Y.: Fast nonnegative tensor factorization based on accelerated proximal gradient and low-rank approximation. Neurocomputing **198**, 148–154 (2016)
20. Zhou, L., Salvado, O., Dore, V., Bourgeat, P., Raniga, P., Villemagne, V.L., Rowe, C.C., Fripp, J.: MR-less surface-based amyloid estimation by subject-specific atlas selection and Bayesian fusion. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 220–227. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33418-4_28