Total Variation Constrained Non-Negative Matrix Factorization for Medical Image Registration

Chengcai Leng, Hai Zhang, Guorong Cai, Zhen Chen, and Anup Basu, *Senior Member, IEEE*

***Abstract*—This paper presents a novel medical image registration algorithm named total variation constrained graph-regularization for non-negative matrix factorization (TV-GNMF). The method utilizes non-negative matrix factorization by total variation constraint and graph regularization. The main contributions of our work are the following. First, total variation is incorporated into NMF to control the diffusion speed. The purpose is to denoise in smooth regions and preserve features or details of the data in edge regions by using a diffusion coefficient based on gradient information. Second, we add graph regularization into NMF to reveal intrinsic geometry and structure information of features to enhance the discrimination power. Third, the multiplicative update rules and proof of convergence of the TV-GNMF algorithm are given. Experiments conducted on datasets show that the proposed TV-GNMF method outperforms other state-of-the-art algorithms.**

***Index Terms*—Data clustering, dimension reduction, image registration, non-negative matrix factorization (NMF), total variation (TV).**

I. Introduction

IMAGE registration is an important research topic for aligning two or more images of the same scene taken at different times, viewpoints, or sensors [1]. Registration is widely used in computer vision and medical image processing, including multimodal image fusion, medical image reconstruction, and the monitoring of tumors. For example, the fusion of multimodal information can be realized by registering two images, which provides better visualization of anatomical structures and functional changes to facilitate diagnosis and treatment [2]. Area-based registration methods [3] mainly uses gray level information to optimize the maximum similarity measure, including mutual information (MI), by adapting optimization algorithms for registration [4]. Gong *et al*. [5] proposed a novel image registration method including the pre-registration and a fine-tuning process based on scale-invariant feature transform (SIFT) and MI. Woo *et al*. [6] presented a novel registration method based on MI by incorporating geometric and spatial context to compute the MI cost function in large spatial variation regions for multimodal image registration. However, these methods are very sensitive to intensity variations and suffer from noise interference. Feature-based methods for image registration directly detect salient features and construct feature descriptors, which are robust and invariant to noise, illumination, and distortion. SIFT [7] is one of the most popular methods invariant to rotation, scale, translation, and illumination changes. Rister *et al*. [8] extended SIFT to arbitrary dimensions by adjusting the orientation assignment and gradient histogram of key points.

We can often treat the feature matching problem as a graph matching problem in image registration, since spectral graph theory [9] is widely used for image segmentation [10], [11], graph matching [12]–[15], and image registration [16]–[21]. In order to make many algorithms practical in several real-life applications, dimensionality reduction is necessary. In order to avoid the curse of dimensionality, some dimensionality reduction matching or registration methods have been introduced [22]–[24]. Xu *et al*. [24] proposed such a method for high-dimensional data sets using the Cramer-Rao lower bounds to estimate the transformation parameters and achieve data set registration. In addition, some manifold learning methods [25] have also been presented, such as ISOMAP [26], locally linear embedding (LLE) [27], and Laplacian Eigenmap [28]. However, many of these algorithms have high computational complexity, and deal poorly with large data sets [29]. Liu *et al*. [30] proposed the text detection method based on morphological component analysis and Laplacian dictionary, which can reduce the adverse effects of complex backgrounds and improve the discrimination power of dictionaries.

Recently, some low-rank matrix factorization methods have been introduced in data representation [31]. Among these

Manuscript received June 2, 2020; revised July 2, 2020, July 29, 2020; accepted August 6, 2020. This work was supported by the National Natural Science Foundation of China (61702251, 41971424, 61701191, U1605254), the Natural Science Basic Research Plan in Shaanxi Province of China (2018JM6030), the Key Technical Project of Fujian Province (2017H6015), the Science and Technology Project of Xiamen (3502Z20183032), the Doctor Scientific Research Starting Foundation of Northwest University (338050050), Youth Academic Talent Support Program of Northwest University (360051900151), and the Natural Sciences and Engineering Research Council of Canada, Canada. Recommended by Associate Editor Xin Luo. *(Corresponding author: Chengcai Leng.)*

Citation: C. C. Leng, H. Zhang, G. R. Cai, Z. Chen, and A. Basu, "Total variation constrained non-negative matrix factorization for medical image Registration," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 5, pp. 1025–1037, May 2021.

C. C. Leng is with the School of Mathematics, Northwest University, Xi'an 710127, and with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada (e-mail: ccleng@nwu.edu.cn).

H. Zhang is with the School of Mathematics, Northwest University, Xi'an 710127, China (e-mail: zhanghai@nwu.edu.cn).

G. R. Cai is with the College of Computer Engineering, Jimei University, Xiamen 361021, China (e-mail: guorongcai.jmu@gmail.com).

Z. Chen is the School of Measuring and Optical Engineering, Nanchang Hangkong University, Nanchang 330063, China (e-mail: chenzhen@nchu.edu.cn).

A. Basu is with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada (e-mail: basu@ualberta.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2021.1003979

methods, non-negative matrix factorization (NMF) [32] achieves a part-based representation for non-negative data sets with applications in data clustering [33], [34] and data or image analysis [35], [36]. Some researchers, such as [37]−[40], also incorporated manifold learning information into NMF. Li *et al*. [39] proposed a graph regularized non-negative low-rank matrix factorization (NLMF) method by adding graph regularization into NLMF to exploit the manifold structure information and utilizing robust principal components analysis (PCA). Shang *et al*. [40] proposed a novel feature selection method by adding sparse regression and dual-graph regularization to NMF to improve the feature selection ability. Ghaffari and Fatemizadeh [41] presented a new image registration method by introducing correlation into the low rank matrix theory based on rank-regularized sum-of-squared-differences (SSD) to improve the similarity measures. In addition, there are still very few NMF based methods used for image matching or image registration. Luo *et al*. have also published several papers [42]–[46] using non-negative latent factor models for high-dimensional and sparse matrices, which can be widely used in industrial applications and highly accurate web service QoS predictions. We will introduce a special sparse matrix factorization method for image registration called total variation constrained graph-regularization for non-negative matrix factorization (TV-GNMF).

Rudin *et al*. [47] first proposed the total variation (TV) method, which is effective for image denoising and can enhance the boundary features of large data sets. It can be used for various pattern recognition tasks, such as hyperspectral unmixing [48]–[50], data clustering [51], image restoration or image fusion [52]–[54], and face recognition [55], [56]. Thus, TV regularization is incorporated into NMF to enhance the details or features of the data. Graph regularization can also be added to NMF, which can discover the intrinsic geometric and structural information of the data. In the differential form of TV regularization, a diffusion coefficient is used to control the diffusion speed. This coefficient can denoise in smooth regions and preserve details in edges regions based on the gradient information. Therefore, our approach is a good part-based data representation that improves the data discrimination ability for clustering big data sets. We exploit this part-based data representation method to find better feature point matches for image registration.

In this paper, we propose a special part-based matrix factorization method, called TV-GNMF, which extends our previous work in [57]. The manifold graph regularization enhances and efficiently reveals the intrinsic geometric and structural information of the data, and the TV regularization denoises and preserves the sharp edges or boundaries to enhance the features of an image. We now explain why we incorporate TV regularization into TV-GNMF. In the TV regularization terms, the diffusion coefficient $1/|\nabla H|$ is used to control the diffusion speed, which can denoise and enhance the edges or details based on the gradient information. If $|\nabla H|$ has a large value in the neighborhood of a point, this point is considered to be an edge and the diffusion speed is lowered to preserve the edges. Otherwise, if $|\nabla H|$ has a small value in the neighborhood of a point, and the diffusion is strong, it helps remove noise. We develop novel iterative update rules, prove the convergence of our optimization technique and give a matching algorithm based on TV-GNMF. Experimental results demonstrate the discrimination ability and better performance of our algorithm.

The remaining sections are organized as follows: Background work is introduced in Section II. Section III proposes the TV-GNMF method, detailed multiplicative update rules and proof of convergence of our optimization method. Section IV presents the image matching algorithm based on TV-GNMF. Experimental results are presented in Section V, before the conclusions in Section VI.

II. Preliminaries

A. Symbols

First, we list some necessary symbols used in this paper in Table I.

TABLE I
Some Necessary Symbols

Symbol	Description
$V_{m\times n}$	Non-negative matrix of size $m\times n$
$W_{m\times r}$	Basis matrix of size $m\times r$
$H_{r\times n}$	Coefficient matrix of size $r\times n$
m, n, r	The number of features, sample points, factors respectively
$\|\cdot\|_F$	Frobenius norm
T	Transpose
$\|\cdot\|_2$	L_2 norm
$\|\cdot\|_{BV}$	TV semi-norm
α,β,λ	Parameters
D, S, L	Diagonal matrix, weight matrix, Laplacian matrix respectively
$\mathrm{Tr}(\cdot)$	Trace of matrix
$\|\cdot\|_{TV}$	TV norm
φ_{ik},ϕ_{kj}	Lagrange multipliers
$div(\cdot)$	Divergence function
∇	Gradient
p	The number of nearest neighbors

B. NMF

Non-negative matrix factorization attempts to find an approximate factorization $V_{m\times n}=[v_1,v_2,\ldots,v_n]\approx W_{m\times r}H_{r\times n}$. Equation (1) measures the similarity between V and WH [58]:

$$O_F=\|V-WH\|_F^2,\quad \text{s.t. } W\geq 0, H\geq 0. \tag{1}$$

The multiplicative update rules are formulated for (1) by Lee and Seung [59], to find a locally optimal solution:

$$W_{ik}\leftarrow W_{ik}\frac{\left(VH^T\right)_{ik}}{(WHH^T)_{ik}},\ H_{kj}\leftarrow H_{kj}\frac{\left(W^TV\right)_{kj}}{(W^TWH)_{kj}}.$$

C. TV-NMF

Yin and Liu [56] proposed a new NMF model with bounded

TV regularization to solve the following optimization problem:

$$O_{TV-NMF} = \frac{1}{2}\|V - WH\|_2^2 + \alpha|H|_{BV} + \beta|W|_{BV} \quad \text{s. t.} \quad W \geq 0,\ H \geq 0. \tag{2}$$

Further definitions of the symbols above can be found in [56].

D. GNMF

Graph regularization is introduced into NMF, i.e., GNMF, to reveal the geometric information of the data [37], as follows:

$$O_{GNMF} = \|V - WH\|_F^2 + \lambda \mathrm{Tr}(HLH^T) \quad \text{s. t.}\ W \geq 0,\ H \geq 0 \tag{3}$$

$L = D - S$ is the graph Laplacian matrix, where S is a weight matrix and D is a diagonal matrix, i.e., $D_{ii} = \sum_{j=1}^{n} S_{ij}$.

Note that S is a weight matrix that we need to construct. Consider a graph with n vertices, we first use the 0-1 weighting scheme to construct a p-nearest neighbor data graph each of whose vertex corresponds to a data point, i.e., v_j. Therefore, the weight matrix S is defined as [37]:

$$S_{jk} = \begin{cases} 1, & \text{if } v_k \in n_p(v_j) \\ 0, & \text{otherwise} \end{cases} \qquad j,k = 1,2,\ldots,n$$

where $n_p(v_j)$ denotes the set of p-nearest neighbors of v_j. In addition, related theory and definitions can be found in [9], [37].

III. TV-GNMF

In this section, we outline the idea behind the total variation method for enhancing or preserving edge features of data sets (images). The TV method is a form of anisotropic diffusion, which smoothens by selectively using diffusion coefficients based on the gradient information to retain image features while eliminating noise. Therefore, TV regularization and graph regularization are integrated with the NMF model to preserve edge features of the intrinsic geometry and structure information of the data. The proposed novel model called TV-GNMF can enhance the intrinsic geometry and preserve edge characteristics of the data to improve discrimination ability for data clustering and image matching.

A. Total Variation

In order to enhance the edge features of the data, we introduce TV [47] regularization in this paper, defined as

$$E(H) = \|H\|_{TV} = \int_{\Omega} |\nabla H| \, dxdy. \tag{4}$$

A similar discrete form can be found in [51], [57]. The TV method can remove noise and preserve edge features [60].

B. Multiplicative Update Rules

Based on TV regularization and graph regularization, the TV-GNMF model with TV regularization is given by

$$O_{TV-GNMF} = \|V - WH\|_F^2 + \lambda \mathrm{Tr}(HLH^T) + 2\beta\|H\|_{TV} \quad \text{s. t.}\ W \geq 0,\ H \geq 0 \tag{5}$$

where $\lambda, \beta \geq 0$ are parameters that can balance the reconstruction error in the first term of the objective function $O_{TV-GNMF}$ of TV-GNMF in (5).

The iterative updating algorithm can achieve a locally optimal solution to $O_{TV-GNMF}$ as follows:

$$O_{TV-GNMF} = \mathrm{Tr}(VV^T) - 2\mathrm{Tr}(VH^TW^T) + \mathrm{Tr}(WHH^TW^T) + \lambda \mathrm{Tr}(HLH^T) + 2\beta\|H\|_{TV}. \tag{6}$$

The Lagrange multipliers φ_{ik} and ϕ_{kj} are given for constraints $w_{ik} \geq 0$ and $h_{kj} \geq 0$, and $\Psi = [\varphi_{ik}]$, $\Phi = [\phi_{kj}]$. The Lagrange function L_L is given by

$$\begin{aligned} L_L &= O_{TV-GNMF} + \mathrm{Tr}(\Psi W^T) + \mathrm{Tr}(\Phi H^T) \\ &= \mathrm{Tr}(VV^T) - 2\mathrm{Tr}(VH^TW^T) + \mathrm{Tr}(WHH^TW^T) \\ &\quad + \lambda \mathrm{Tr}(HLH^T) + 2\beta\|H\|_{TV} + \mathrm{Tr}(\Psi W^T) + \mathrm{Tr}(\Phi H^T). \end{aligned} \tag{7}$$

The partial derivatives of L_L with respect to W and H are:

$$\frac{\partial L_L}{\partial W} = -2VH^T + 2WHH^T + \Psi \tag{8}$$

$$\frac{\partial L_L}{\partial H} = -2W^TV + 2W^TWH + 2\lambda HL - 2\beta div\left(\frac{\nabla H}{|\nabla H|}\right) + \Phi. \tag{9}$$

Using the Karush-Kuhn-Tucker (KKT) conditions $\varphi_{ik}w_{ik} = 0$ and $\phi_{kj}h_{kj} = 0$, the following equations are given for w_{ik} and h_{kj}:

$$-(VH^T)_{ik}w_{ik} + (WHH^T)_{ik}w_{ik} = 0 \tag{10}$$

$$-2(W^TV)_{kj}h_{kj} + 2(W^TWH)_{kj}h_{kj} + 2\lambda(HL)_{kj}h_{kj} - 2\beta div\left(\frac{\nabla H}{|\nabla H|}\right)_{kj} h_{kj} = 0. \tag{11}$$

The following multiplicative update rules are obtained based on (10) and (11):

$$w_{ik} \leftarrow w_{ik}\frac{(VH^T)_{ik}}{(WHH^T)_{ik}} \tag{12}$$

$$h_{kj} \leftarrow h_{kj}\frac{\left(W^TV + \lambda HS + \beta div\left(\frac{\nabla H}{|\nabla H|}\right)\right)_{kj}}{(W^TWH + \lambda HD)_{kj}}. \tag{13}$$

The detailed updating procedure of TV-GNMF is summarized in Algorithm 1.

Algorithm 1 TV-GNMF Algorithm

Input: $V \in \mathbb{R}^{m\times n}$, D, S and $1 \leq r \leq \min\{m,n\}$.

Initialization: W_0, H_0, λ, β and $k = 0$.

For $k = 0,1\ldots$ until convergence or maximum iteration.

Update H^{k+1} according to

$$H^{k+1} = H^k \frac{\left(W^TV + \lambda HS + \beta div\left(\frac{\nabla H}{|\nabla H|}\right)\right)^k}{(W^TWH + \lambda HD)^k}$$

Update W^{k+1} according to

$$W^{k+1} = W^k \frac{(VH^T)^k}{(WHH^T)^k}$$

$k = k + 1$

End

Output: $W \in \mathbb{R}^{m\times r}$, $H \in \mathbb{R}^{r\times n}$.

We will describe a theorem related to the above iterative update rules along with the detailed proof of convergence.

TABLE II
COMPUTATIONAL OPERATION COUNTS FOR EACH ITERATION FOR DIFFERENT METHODS

	F-norm formulation			
	Fladd	Flmlt	Fldiv	Overall
NMF	$2mnr+2(m+n)r^2$	$2mnr+2(m+n)r^2+(m+n)r$	$(m+n)r$	$O(mnr)$
GNMF	$2mnr+2(m+n)r^2+n(p+3)r$	$2mnr+2(m+n)r^2+(m+n)r+n(p+1)r$	$(m+n)r$	$O(mnr)$
TV-GNMF	$2mnr+2(m+n)r^2+n(p+3)r+12nr$	$2mnr+2(m+n)r^2+(m+n)r+n(p+1)r+7nr$	$(m+4n)r$	$O(mnr)$

C. Proof of Convergence

To prove convergence, the auxiliary function is constructed and used in the expectation-maximization algorithm [61]. Before the proof of convergence, we will introduce the following related definition and lemmas.

Definition 1: $G(x,x')$ is an auxiliary function for $F(x)$ if $G(x,x') \geq F(x)$ and $G(x,x) = F(x)$ are satisfied.

Lemma 1: If G is an auxiliary function of F, then F is non-increasing under the update rule:

$$x^{t+1} = \arg\min_x G(x, x^t). \tag{14}$$

Proof: $F(x^{t+1}) \leq G(x^{t+1}, x^t) \leq G(x^t, x^t) = F(x^t)$. ■

Lemma 2: The function

$$G(w, w_{ab}^t) = F_{w_{ab}}(w_{ab}^t) + F'_{w_{ab}}(w_{ab}^t)(w - w_{ab}^t) + \frac{(WHH^T)_{ab}}{w_{ab}^t}(w - w_{ab}^t)^2 \tag{15}$$

is an auxiliary function for $F_{w_{ab}}$, which is the part of $O_{TV-GNMF}$ that is only relevant to w_{ab}.

Proof: Since $G(w,w) = F_{w_{ab}}(w)$ is obvious, we only need to show that $G(w, w_{ab}^t) \geq F_{w_{ab}}(w)$. Consider the Taylor series expansion of $F_{w_{ab}}(w)$:

$$F_{w_{ab}}(w) = F_{w_{ab}}(w_{ab}^t) + F'_{w_{ab}}(w_{ab}^t)(w - w_{ab}^t) + [(HH^T)_{bb}](w - w_{ab}^t)^2. \tag{16}$$

By combining (15) with (16), we can show that $G(w, w_{ab}^t) \geq F_{w_{ab}}(w)$ is equivalent to

$$\frac{(WHH^T)_{ab}}{w_{ab}^t} \geq (HH^T)_{bb}. \tag{17}$$

We have

$$(WHH^T)_{ab} = \sum_{l=1}^{r} w_{al}^t (HH^T)_{lb} \geq w_{ab}^t (HH^T)_{bb}. \tag{18}$$

Thus, (17) holds and $G(w, w_{ab}^t) \geq F_{w_{ab}}(w)$. ■

Motivated by [37], [51], [56], [59], we will now describe Theorem 1.

Theorem 1: The objective function $O_{TV-GNMF}$ in (6) is non-increasing under the update rules in (12) and (13).

Proof: Replacing $G(w, w_{ab}^t)$ in (14) by (15), we have

$$w_{ab}^{t+1} = w_{ab}^t - w_{ab}^t \frac{F'_{w_{ab}}(w_{ab}^t)}{2(WHH^T)_{ab}} = w_{ab}^t \frac{(VH^T)_{ab}}{(WHH^T)_{ab}}.$$

We use an auxiliary function (15), $F_{w_{ab}}$ is non-increasing under the update rules. Similarly, we get

$$h_{ab}^{t+1} = h_{ab}^t \frac{\left(W^T V + \lambda HS + \beta div\left(\frac{\nabla H}{|\nabla H|}\right)\right)_{ab}}{(W^T WH + \lambda HD)_{ab}}.$$

Theorem 1 guarantees convergence under the update rules based on (12) and (13). ■

D. Complexity Analysis

The computational complexity of the TV-GNMF method will be discussed and compared with the NMF and GNMF methods. Since Cai *et al.* gave the arithmetic operations of NMF and GNMF for each iteration in [37], we follow their results as shown in Table II. The main difference between TV-GNMF and GNMF is the component of TV norm. Specifically, to generate the divergence function of the discrete gradient matrix H, we need to calculate the first and second derivatives of each element, which results in 9 floating-point additions and 3 floating-point divisions. Moreover, the update rule of the divergence function of each element needs 3 floating-point additions, 7 floating-point multiplications, and one floating-point division. In general, compared to GNMF, our method adds 12 floating-point additions, 7 floating-point multiplications, and 4 floating-point divisions for each iteration. Note that m denotes the rows of an input image, whose scale is much larger than 12. Therefore, the overall complexity of our TV-GNMF is also $O(mnr)$. Details of the complexity analysis are summarized in Table II.

In Table II, Fladd, Flmlt, Fldiv denote the number of floating-point additions, floating-point multiplications, and floating-point divisions, respectively; n represents the number of sample points; m is the number of features; and, r and p stand for the number of factors and the number of the nearest neighbors, respectively.

IV. IMAGE MATCHING ALGORITHM BASED ON TV-GNMF

To avoid confusion, the first part tests clustering performance of the data sets directly represented by the matrix V to compute matrices W and H based on (12) and (13), without using the following image registration algorithm. The data sets include images with many features or details and TV regularization can denoise and preserve details or edges of features to improve clustering performance.

The second part evaluates matching performance on medical images. We construct the non-negative matrix, not images, by exploiting geometric positions of feature points. TV regularization can enhance and characterize the intrinsic relationship of feature points based on diffusion depending on the gradient information of points. Further details on the matching algorithm of TV-GNMF can be found in [57].

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we provide some experimental evaluation of

the proposed TV-GNMF method for image clustering and registration. There are two aspects in this study. We commence with an analysis of image data sets to demonstrate the clustering performance based on multiplicative update rules. The image matching performance is tested in the second part to show that the dimensionality reduction method has better discrimination ability for medical image registration. The clustering performance is evaluated in the first part.

A. Data Sets

The important statistics of the data sets used to evaluate the clustering performance are summarized in Table III. Further details can be found in [35], [37], [51].

TABLE III
INFORMATION ON THE DATA SETS

Data sets	Size	Dimensionality	# of classes
COIL20	1440	1024	20
PIE	2856	1024	68
ORL	400	1024	40
TDT2	9394	36771	30

B. Performance Evaluation and Comparisons

Performance is tested by comparing the labels obtained for each sample with the labels provided by the data sets. One metric is accuracy (*AC*), used to measure the percentage of correct labels obtained. The second metric is the normalized mutual information metric (*NMI*), used to measure how similar the two sets of clusters are. Detailed definitions of *AC* and *NMI* can be found in [33], [62].

In order to demonstrate our method's performance on the above data sets, we compared TV-GNMF with two other related methods, i.e., the NMF [32] and GNMF [37] methods. The Frobenius norm is used to measure the similarity for the above three methods. We construct the weight matrix S of (3) and (5) using the 0–1 weighting based on the p-nearest neighbor graph, with $p = 5$ for the GNMF and TV-GNMF methods. In addition, the regularization parameter λ is set to 100 for the GNMF method [37]; and the TV regularization parameter β is given and tested in the experiments for the TV-GNMF method.

Table IV gives the data clustering results on the above four normalized datasets. In the experiments, the different cluster numbers are given on the Columbia University Image Library (COIL20) and Olivetti Research Laboratory (ORL) datasets for 100 iterations, and on the NIST Topic Detection and Tracking (TDT2) corpus dataset for 50 iterations. The regularization parameter β is set to 2 for the above three datasets in our TV-GNMF method. The number of iterations is set to 50 on the Pose, Illumination and Expression (PIE) face dataset for testing, and the parameter β is set to 0.2 in our TV-GNMF method. The GNMF method outperforms NMF, indicating that GNMF preserves or reveals the geometric structure of the data in learning under varying angles on the COIL20 dataset and different lighting and illumination conditions on the PIE dataset. Surprisingly, the average of *AC* and *NMI* of the GNMF method is lower than the NMF method for the ORL dataset. The GNMF method does not reveal the geometric information because the ORL database consists of 40 distinct subjects with varying lighting, different facial expressions, and details. Our TV-GNMF method has high accuracy and normalized mutual information. The TV-GNMF method improves clustering performance because it combines the merits of graphs and TV regularization to discover the geometric structure information and enhances feature details. The best results are highlighted in bold. In most cases, our TV-GNMF method has the best performance. However, in a few situations, GNMF has higher *AC* and *NMI* than ours for the underlined cases, such as when $k = 30$ for PIE. Our method cannot preserve the sharp edges or boundaries to enhance the feature details because the PIE dataset consists of 68 distinct subjects with different lighting and illumination conditions. In addition, the parameter β also affects the clustering performance. Overall, our TV-GNMF method outperforms NMF and GNMF, and has better performance. Our TV-GNMF method preserves geometric structure information and enhances the edge features of the data as demonstrated in Table IV. Note that our method outperforms others in most cases, including every instance of the COIL20, ORL, and TDT2 datasets in Table IV. Even for the few situations where our method does not have the best score, we are within 2% of the top score. However, our model has two parameters, λ and β, which we need to choose adaptively or empirically.

In addition, we use the ORL dataset as an example to test the effectiveness of our method. We add Gaussian noise with mean 0, and variance 0.09, based on the NMF, TV-NMF, GNMF, and TV-GNMF methods under the same conditions, with parameter $\beta = 2$, and 50 iterations, for different cluster numbers. We can see that TV-GNMF has the best clustering results as shown in Table V. This happens because TV regularization can remove noise and preserve the details or features of the data, and graph regularization can discover the intrinsic geometric and structural information of the data while removing noise and enhancing features. TV-NMF has better results than NMF and GNMF, because GNMF cannot discover or reveal the intrinsic geometric and structural information of the data well in the presence of noise.

C. Parameter Evaluation

In this section, stability is tested based on our TV-GNMF method for various parameter settings. Our model has two important regularization parameters: λ and β. The GNMF method produces the best results when λ is set to 100. In our model, λ is also set to 100, and we vary the regularization parameter β to test stability. The performance of TV-GNMF varies with the β on COIL20 and PIE datasets as shown in Fig.1, which shows that TV-GNMF is very stable with respect to β. Fig. 1(a)−1(c) give the clustering performance when the regularization parameter β varies from 0.1 to 20 for different classes; such as 8, 13, and 18 on the COIL20 dataset. Fig. 1(d)−1(f) also present the clustering performance when the parameter β varies from 0.1 to 35 under different classes, such as 20, 35, and 50 on the PIE dataset. For a big range of the regularization parameter on the two data sets, TV-GNMF

TABLE IV
COMPARISONS ON COIL20, PIE, ORL AND TDT2 DATASETS

Data	k	*AC* (%)			*NMI* (%)		
		NMF	GNMF	TV-GNMF	NMF	GNMF	TV-GNMF
COIL20	4	58.403	74.722	**76.597**	70.044	86.723	**87.729**
	8	60.000	69.861	**75.417**	70.715	85.287	**88.508**
	13	62.778	72.361	**74.028**	71.822	87.017	**87.219**
	18	69.444	74.028	**75.972**	74.277	85.979	**89.127**
	20	66.736	79.306	**79.722**	74.361	88.515	**89.862**
	Avg.	63.472	74.056	**76.347**	72.244	86.704	**88.489**
PIE	10	24.055	69.188	**72.724**	52.170	85.752	**87.174**
	30	25.490	**71.534**	69.013	53.109	**86.362**	85.551
	50	22.759	**76.436**	75.595	50.686	87.173	**87.644**
	68	24.510	72.129	**76.120**	53.014	86.521	**87.909**
	Avg.	24.204	72.322	**73.363**	52.245	86.4520	**87.070**
ORL	3	48.750	50.250	**52.750**	67.527	68.900	**71.513**
	6	46.500	45.000	**51.750**	68.744	66.101	**69.128**
	9	48.500	46.750	**51.500**	67.976	66.099	**68.431**
	12	49.000	48.000	**53.000**	68.503	68.785	**70.870**
	15	46.500	47.000	**50.250**	68.025	67.332	**70.577**
	Avg.	48.350	47.400	**51.850**	68.155	67.443	**70.104**
TDT2	5	45.476	70.939	**81.616**	58.877	78.434	**84.509**
	10	43.773	81.563	**86.385**	57.446	82.651	**84.702**
	15	47.743	83.553	**85.501**	58.812	83.217	**83.587**
	20	44.922	84.288	**86.183**	58.978	83.594	**83.785**
	Avg.	45.479	80.086	**84.921**	58.528	81.974	**84.146**

TABLE V
COMPARISONS ON ORL DATASET WITH NOISE

Data	k	*AC* (%)				*NMI* (%)			
		NMF	TV-NMF	GNMF	TV-GNMF	NMF	TV-NMF	GNMF	TV-GNMF
ORL	25	14.750	15.500	13.000	**15.750**	34.311	35.892	29.313	**38.472**
	30	14.250	15.000	15.250	**17.000**	31.935	34.164	32.767	**39.898**
	35	15.750	15.500	14.500	**16.500**	34.432	35.984	31.818	**38.312**
	40	15.250	15.750	13.500	**16.000**	33.845	35.323	32.160	**40.450**
	Avg.	15.000	15.438	14.063	**16.313**	33.631	35.341	31.515	**39.283**

has consistently good and stable performance. For the COIL20 dataset, our method produces relatively big clustering results based on parameter evaluation when $\beta = 5$. The reason is that the randomness of the initial values of W and H affect the clustering performance. The initial values are randomly generated by non-negative constraints when we execute TV-GNMF with β varying from 0.1 to 20 for different numbers of clusters, which can only ensure convergence to a local optima as they are updated iteratively. However, from the second experimental results, the range of the regularization parameter is larger than the first and has higher accuracy.

In addition, we also use the ORL dataset as an example to test the effectiveness of our method with different p based on the GNMF and TV-GNMF methods under the same conditions with $\lambda = 100$ and 50 iterations, cluster number set to 16 classes and β set to 2 in the TV-GNMF method. As we have seen, GNMF and TV-GNMF use a p-nearest neighbor graph to capture the local geometric structure information on a scatter of data points. GNMF and TV-GNMF have better clustering performance based on the assumption that two neighboring data points share the same label. When there are more nearest neighbors p, this assumption is more likely to fail. This is the reason why the performance of GNMF and TV-GNMF declines, and TV-GNMF is still superior to GNMF as p increases, as shown in Table VI and in Fig. 2.

D. Medical Image Registration Performance

In this section, a novel low-rank preserving technique is proposed by matching feature points to verify the discrimination ability to achieve one-to-one correspondences. We must emphasize that feature point detection or feature point description is not our research focus. The key issue is

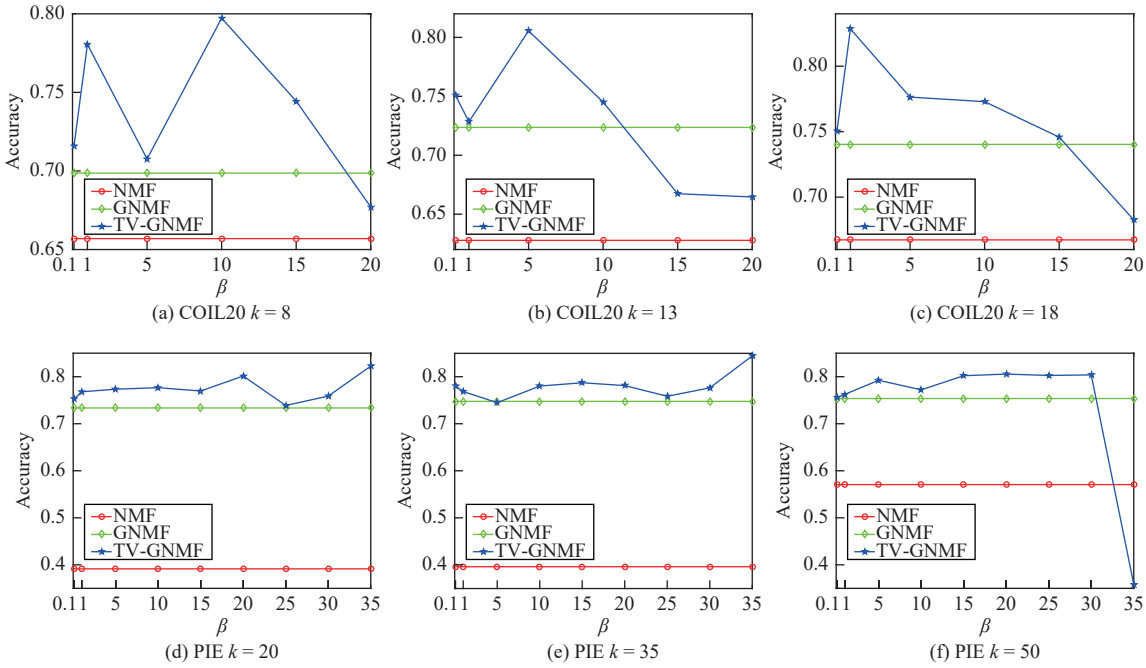

Fig. 1. Performance of TV-GNMF for varying regularization parameter β on COIL20 and PIE datasets.

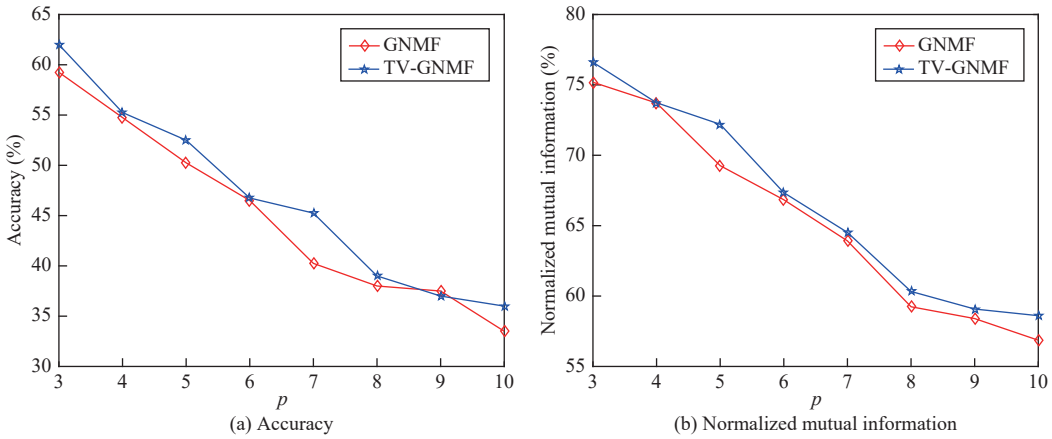

Fig. 2. The performance of GNMF and TV-GNMF decreases as p increases on ORL dataset.

TABLE VI
COMPARISONS ON ORL DATASET WITH THE DIFFERENT p

Data	p	*AC* (%)		*NMI* (%)	
		GNMF	TV-GNMF	GNMF	TV-GNMF
ORL	3	59.250	**62.000**	75.188	**76.605**
	4	54.750	**55.250**	73.709	**73.731**
	5	50.250	**52.500**	69.263	**72.200**
	6	46.500	**46.750**	66.840	**67.345**
	7	40.250	**45.250**	63.918	**64.521**
	8	38.000	**39.000**	59.240	**60.327**
	9	**37.500**	37.000	58.401	**59.078**
	10	33.500	**36.000**	56.863	**58.606**
	Avg.	45.000	**46.719**	65.428	**66.552**

how our TV-GNMF method exhibits discriminating power to capture the intrinsic geometry and structure information and finds one-to-one correspondences between feature points. In order to test the matching performance, we applied it to medical images demonstrating that the proposed method has the discriminating power to achieve stable one-to-one feature correspondences. Furthermore, we compare the results of our TV-GNMF method with the projection clustering matching method (Caelli's method) [23] and Zass' method [63] in terms of matching.

The 32nd slice of T1 and T2 of a magnetic resonance imaging (MRI) sequence is used to test, and the image matching results are given in Fig. 3. T1 denotes prominent tissue T1 relaxation (longitudinal relaxation) difference, which is used to observe anatomical structures. T2 denotes prominent tissue T2 relaxation (transverse relaxation) difference, which is used to show tissue lesions. We use the Harris Corner Detector [64] to extract 27 feature points and 38 feature points from Figs. 3(a), 3(c), and 3(e), and Figs. 3(b), 3(d), and 3(f), respectively. Obviously, Fig. 3(a) produces some two-to-one mismatches. Zass' and our methods achieve

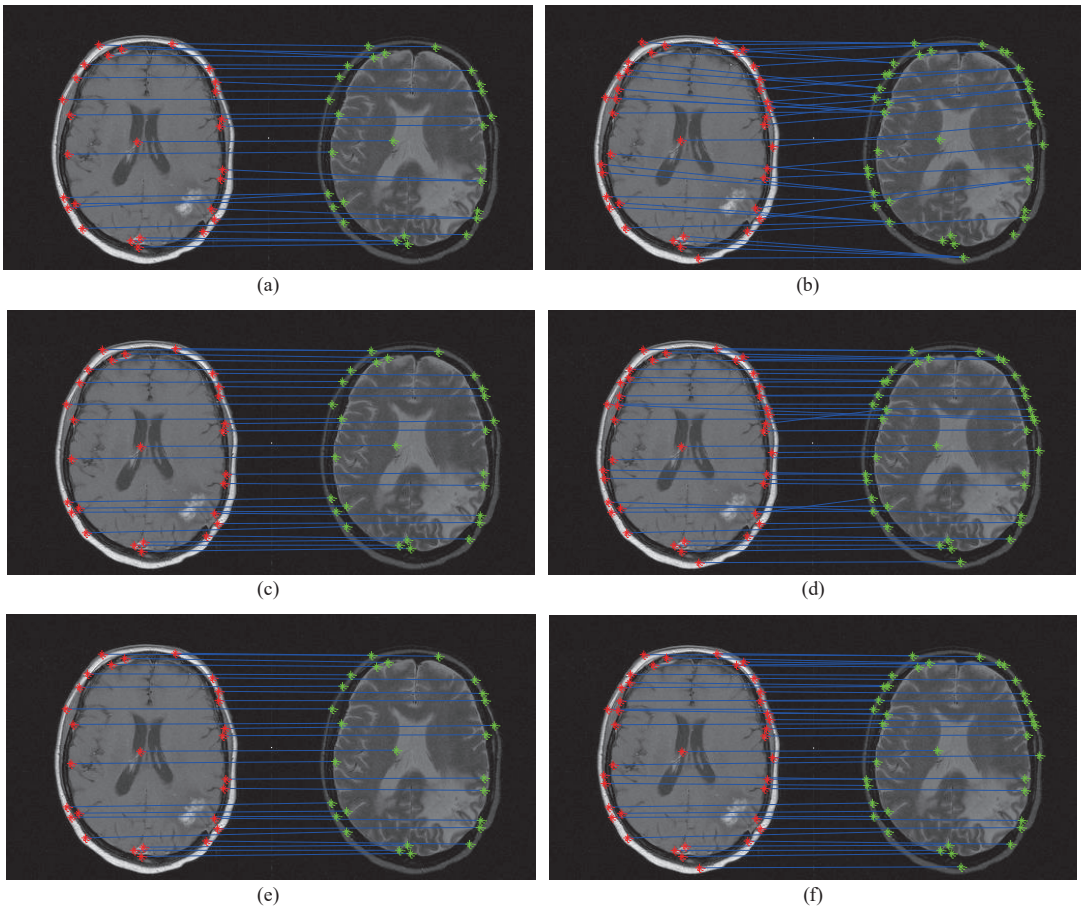

Fig. 3. Matching results: (a) and (b) Caelli's method; (c) and (d) Zass' method; (e) and (f) Our TV-GNMF method.

one-to-one correspondences in Figs. 3(c) and 3(e), respectively. In addition, some feature points are added as shown in Figs. 3(b), 3(d), and 3(f). More many-to-one correspondences are produced by Caelli's method in Fig. 3(b) when the number of feature points is increased. The reason is that if the distances between some of the extracted points are very close to each other, they are more likely to be in the same class. Thus, more many-to-one correspondences are produced. Zass' method is better than Caelli's method because the matching problem utilizes a probabilistic framework based on hypergraphs. However, this method also produces some mismatches as shown in Fig. 3(d). Despite some feature points being very close to each other, our method can still find one-to-one correspondences, as seen in Fig. 3(f). This result indicates that our method has better discrimination ability to improve matching performance, thereby achieving robust image registration. We also utilize the computation time to evaluate the quantitative analysis, and the computation time for the entire matching process, including feature point extraction, in Table VII. This table indicates that our method needs less computation time. Please see [57] for additional details.

In addition, we also use our method to test the matching ability compared to a more classical and effective method called the coherent point drift (CPD) method [65]. We use the Harris Corner Detector to extract 156 feature points in T1 (red "*") and T2 (blue "o") of the 24th slice of an MRI sequence. The test experiment is intended to show the effectiveness of our method. We execute the CPD algorithm and our matching algorithm on the feature point sets. Both methods have good matching performance based on the experimental matching results shown in Fig. 4. However, our method takes 0.481 s less than the 0.990 s needed by the CPD method. This indicates that we have introduced an effective matching method that is also computationally more efficient.

TABLE VII
COMPARISON OF THE COMPUTATION TIME (s) AS SHOWN IN FIG. 3

Figure and computation time	Figure		
	(a) and (b)	(c) and (d)	(e) and (f)
Fig. 3 (s)	0.573/0.682	0.536/0.650	**0.426/0.480**

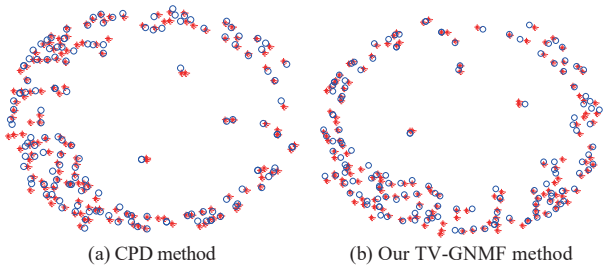

Fig. 4. Matching results for a feature point set.

To test the accuracy of registration, the root mean square

TABLE VIII
COMPARISON OF COMPUTATION TIME AND ACCURACY FOR DIFFERENT IMAGE MODALITIES

Patients and methods		Computation time (s)		Accuracy (%)	
		T1 and PD	PD and T2	T1 and PD	PD and T2
19th patient	Caelli	0.5960	0.5651	21/21 (100.0%)	13/16 (81.3%)
	TV-GNMF	0.5290	0.5197	21/21 (100.0%)	15/16 (93.8%)
21st patient	Caelli	0.4713	0.4683	9/10 (90.0%)	**3/7 (42.9%)**
	TV-GNMF	0.4895	0.4824	10/10 (100.0%)	7/7 (100.0%)
27th patient	Caelli	0.5740	0.6064	**4/16 (25.0%)**	20/21 (95.2%)
	TV-GNMF	0.5331	0.5697	16/16 (100.0%)	20/21 (95.2%)
37th patient	Caelli	0.4378	0.3878	7/7 (100.0%)	**0/4 (0.0%)**
	TV-GNMF	0.4593	0.4530	7/7 (100.0%)	4/4 (100.0%)
45th patient	Caelli	0.4827	0.4949	12/12 (100.0%)	14/14 (100.0%)
	TV-GNMF	0.4870	0.4816	12/12 (100.0%)	14/14 (100.0%)
50th patient	Caelli	0.5356	0.4817	17/17 (100.0%)	10/11 (90.9%)
	TV-GNMF	0.5112	0.4808	17/17 (100.0%)	10/11 (90.9%)
60th patient	Caelli	0.5589	0.5634	15/15 (100.0%)	14/15 (93.3%)
	TV-GNMF	0.5240	0.5331	15/15 (100.0%)	14/15 (93.3%)
65th patient	Caelli	0.5179	0.5284	10/10 (100.0%)	10/10 (100.0%)
	TV-GNMF	0.5220	0.5266	10/10 (100.0%)	10/10 (100.0%)

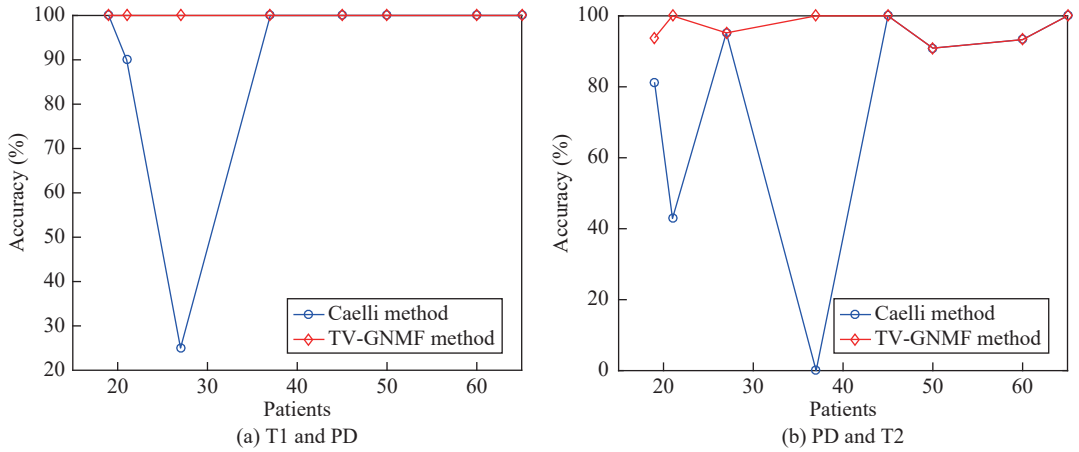

Fig. 5. Plot of accuracy considering different slices for the different patients of Table VIII.

error (*RMSE*) is used to evaluate the accuracy. Detailed results can be found in [57].

Finally, to verify the discrimination ability and robustness under different medical image modalities, we give the accuracy which is defined as N_c/N, where N_c denotes the number of correct matches and N denotes the total number of feature points. Table VIII summarizes eight experiments including different patients, with each patient repeated twice. It also shows the computation time and accuracy for different patients from the brain datasets [66], and is compared with Caelli's method. These experiments show that the proposed method has better discrimination ability in finding one-to-one correspondences and has good matching results. The reason for large fluctuation in accuracy in Caelli's method for different patients is that if the distances between feature points are extracted very close to each other, it is more likely that these points are in the same class. This produces many-to-one correspondences to create a large fluctuation in accuracy. Fig. 5 shows the matching performance for different patients in Table VIII. For this figure, the *y*-axis represents the accuracy and the *x*-axis denotes the patients. The different numbers of feature points are obtained by using the Harris Corner Detector under the same condition for different patients, and the number of feature points detected is relatively small. Therefore, the computation time is less than 1.0 s and the time difference is not big. However, the results (in bold) are not as good for some patients based on Caelli's method, as shown in Table VIII and Fig. 5. Thus, the experimental results indicate that our method is robust and has more discrimination ability than Caelli's.

Fig. 6 also shows the discrimination ability and robustness on the 65th patient of PD and T2 by increasing the number of feature points. PD reflects the difference in hydrogen proton content for different tissues, i.e., comparison of hydrogen

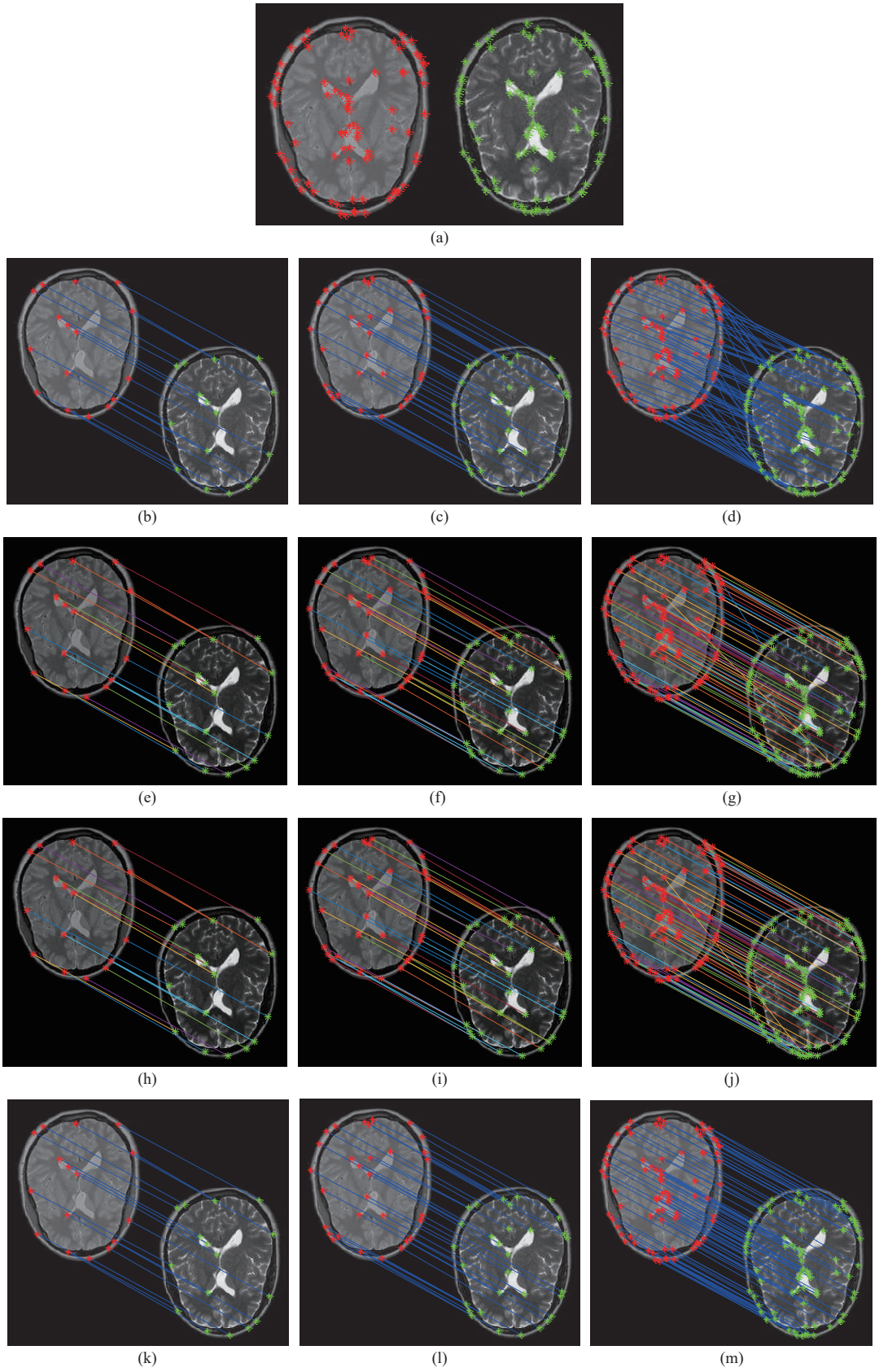

Fig. 6. Discrimination and robustness considering the same patient for different number of feature points. (a) Feature point extraction results of reference and sensed images; (b)−(d) Caelli's method; (e)−(g) TV-NMF method; (h)−(j) GNMF method; (k)−(m) TV-GNMF method.

proton density in prominent tissues. Fig. 6 compares the matching results based on Caelli's method, TV-NMF method, GNMF method, and our method. We can see that the matching results, whether correct or incorrect, more clearly,

when there are relatively few feature points and matching lines. For example, for Figs. 6(d), 6(g), 6(j), and 6(m), the matching results have many matching lines used to connect the reference and sensed images. This makes it difficult to see the texture of images due to many mismatches, as shown in Fig. 6(d). In order to avoid this problem, we show decomposition results with the feature points of the reference image (left) and the sensed image (right); which are first extracted as shown in Fig. 6(a). Then, these points are used for image matching as shown in Figs. 6(d), 6(g), 6(j), and 6(m). For these experimental results considering more feature points, our method still has better matching results, as shown in Fig. 6(m), than the TV-NMF method (Fig. 6(g)) and the GNMF method (Fig. 6(j)). However, Caelli's method has completely different results, as shown in Fig. 6(b)−6(d), for different number of feature points. This indicates that our method has good discrimination ability and robustness, and achieves one-to-one correspondences regardless of the number of feature points.

E. Summary

Based on the theory and empirical studies, we summarize that:

1) The proposed TV-GNMF model is able to accurately achieve data clustering and image registration in a low dimensional feature space. Hence, TV-GNMF outperforms other state-of-the-art algorithms in accuracy of clustering, registration, and time efficiency.

2) Total variation constraint and graph regularization can control the diffusion speed to denoise and preserve the features or details of the data. This is achieved by a diffusion coefficient based on the gradient information to reveal intrinsic geometric and structural information of features to enhance the discriminating power.

3) Iterative update rules are developed and a proof of convergence for the TV-GNMF algorithm is given.

VI. CONCLUSIONS

In this paper, we proposed a novel matrix factorization method called TV-GNMF, which can effectively remove noise and preserve the data features utilizing total variation. Our method can also reveal the intrinsic geometric and structural information of the data well to improve discrimination ability. Experimental results on data sets and images indicate that TV-GNMF is a better low-rank representation method for data clustering and image registration. There are two parameters, λ and β, that play a key role in our model. How to adaptively choose the values of λ and β will be investigated in our future work.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Cai, in the College of Computer Science at Zhejiang University, China, for providing his code for implementing GNMF.

REFERENCES

[1] D. D. Yu, F. Yang, C. Y. Yang, C. C. Leng, J. Cao, Y. N. Wang, and J. Tian, "Fast rotation-free feature-based image registration using improved N-SIFT and GMM-based parallel optimization," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 8, pp. 1653–1664, Aug. 2016.

[2] A. A. Goshtasby, *2D and 3D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Hoboken, USA: John Wiley & Sons, Inc., 2005.

[3] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imaging*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.

[4] C. C. Leng, J. J. Xiao, M. Li, and H. P. Zhang, "Robust adaptive principal component analysis based on intergraph matrix for medical image registration," *Comput. Intell. Neurosci.*, vol. 2015, Article No. 829528, Apr. 2015.

[5] M. G. Gong, S. M. Zhao, L. C. Jiao, D. Y. Tian, and S. Wang, "A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4328–4338, Jul. 2014.

[6] J. Woo, M. Stone, and J. L. Prince, "Multimodal registration via mutual information incorporating geometric and spatial context," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 757–769, Feb. 2015.

[7] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[8] B. Rister, M. A. Horowitz, and D. L. Rubin, "Volumetric image registration from invariant keypoints," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4900–4910, Oct. 2017.

[9] F. R. K. Chung, *Spectral Graph Theory*. Providence, USA: American Mathematical Society, 1997.

[10] A. K. Bhandari, A. Ghosh, and I. V. Kumar, "A local contrast fusion based 3D Otsu algorithm for multilevel image segmentation," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 1, pp. 200–213, Jan. 2020.

[11] L. He, N. Ray, Y. S. Guan, and H. Zhang, "Fast large-scale spectral clustering via explicit feature mapping," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1058–1071, Mar. 2019.

[12] C. C. Leng, W. Xu, I. Cheng, and A. Basu, "Graph matching based on stochastic perturbation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4862–4875, Dec. 2015.

[13] J. Tang, L. Shao, X. L. Li, and K. Lu, "A local structural descriptor for image matching via normalized graph Laplacian embedding," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 410–420, Feb. 2016.

[14] X. Yang and Z. Y. Liu, "Adaptive graph matching," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1432–1445, May 2018.

[15] J. C. Yan, C. S. Li, Y. Li, and G. T. Cao, "Adaptive discrete hypergraph matching," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 765–779, Feb. 2018.

[16] J. Y. Ma, J. Zhao, J. W. Tian, A. L. Yuille, and Z. W. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.

[17] J. Chen, J. Y. Ma, C. C. Yang, L. Ma, and S. Zheng, "Non-rigid point set registration via coherent spatial mapping," *Signal Process.*, vol. 106, pp. 62–72, Jan. 2015.

[18] J. Y. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.

[19] J. Y. Ma, J. J. Jiang, H. B. Zhou, J. Zhao, and X. J. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, Aug. 2018.

[20] J. Y. Ma, J. Zhao, J. J. Jiang, H. B. Zhou, and X. J. Guo, "Locality preserving matching," *Int. J. Comput. Vision*, vol. 127, no. 5, pp. 512–531, May. 2019.

[21] H. C. Chen, X. Zhang, S. Y. Du, Z. Z. Wu, and N. N. Zheng, "A correntropy-based affine iterative closest point algorithm for robust point set registration," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 4, pp. 981–991, Jul. 2019.

[22] L. Xu and I. King, "A PCA approach for fast retrieval of structural patterns in attributed graphs," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 31, no. 5, pp. 812–817, Oct. 2001.

[23] T. Caelli and S. Kosinov, "An eigenspace projection clustering method for inexact graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 515–519, Apr. 2004.

[24] M. Xu, H. Chen, and P. K. Varshney, "Dimensionality reduction for registration of high-dimensional data sets," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3041–3049, Aug. 2013.

[25] C. C. Leng, H. Zhang, G. R. Cai, I. Cheng, and A. Basu, "Graph regularized Lp smooth non-negative matrix factorization for data representation," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 2, pp. 584–595, Mar. 2019.

[26] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[27] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[28] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. 14th Int. Conf. Neural Information Processing Systems*, Cambridge, USA, 2001, pp. 585–591.

[29] D. Cai, X. F. He, and J. W. Han, "Isometric projection," in *Proc. 22nd AAAI Conf. Artificial Intelligence*, Vancouver, Canada, 2007, pp. 528–533.

[30] S. P. Liu, Y. T. Xian, H. F. Li, and Z. T. Yu, "Text detection in natural scene images using morphological component analysis and Laplacian dictionary," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 1, pp. 214–222, Jan. 2020.

[31] R. He, T. N. Tan, and L. Wang, "Robust recovery of corrupted low-rank matrix by implicit regularizers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 770–783, Apr. 2014.

[32] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[33] W. Xu, X. Liu, and Y. H. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, Toronto, Canada, 2003, pp. 267–273.

[34] F. Shahnaz, M. W. Berry, V. Pauca, and R. J. Plemmons, "Document clustering using nonnegative matrix factorization," *Inf. Process. Manage.*, vol. 42, no. 2, pp. 373–386, 2006.

[35] H. F. Liu, Z. H. Wu, X. L. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1299–1311, Jul. 2012.

[36] D. Wang, X. B. Gao, and X. M. Wang, "Semi-supervised nonnegative matrix factorization via constraint propagation," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 233–244, Jan. 2016.

[37] D. Cai, X. F. He, J. W. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[38] X. Luo, M. C. Zhou, H. Leung, Y. N. Xia, Q. S. Zhu, Z. H. You, and S. Li, "An incremental-and-static-combined scheme for matrix-factorization- based collaborative filtering," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 333–343, Jan. 2016.

[39] X. L. Li, G. S. Cui, and Y. S. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3840–3853, Nov. 2017.

[40] R. H. Shang, W. B. Wang, R. Stolkin, and L. C. Jiao, "Non-negative spectral learning and sparse regression-based dual-graph regularized feature selection," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 793–806, Feb. 2018.

[41] A. Ghaffari and E. Fatemizadeh, "Image registration based on low rank matrix: Rank-regularized SSD," *IEEE Trans. Med. Imaging*, vol. 37, no. 1, pp. 138–150, Jan. 2018.

[42] M. S. Shang, X. Luo, Z. G. Liu, J. Chen, Y. Yuan, and M. C. Zhou, "Randomized latent factor model for high-dimensional and sparse matrices from industrial applications," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 1, pp. 131–141, Jan. 2019.

[43] X. Luo, Z. G. Liu, S. Li, M. S. Shang, and Z. D. Wang, "A fast non-negative latent factor model based on generalized momentum method," *IEEE Trans. Syst.*, *Man*, *Cybern.*: *Syst.*, vol. 51, no. 1, pp. 610–620, Jan. 2021.

[44] X. Luo, M. C. Zhou, S. Li, L. Hu, and M. S. Shang, "Non-negativity constrained missing data estimation for high-dimensional and sparse matrices from industrial applications," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1844–1855, May 2020.

[45] D. Wu, Q. He, X. Luo, M. S. Shang, Y. He, and G. Y. Wang, "A posterior-neighborhood-regularized latent factor model for highly accurate web service QoS prediction, " *IEEE Trans. Serv. Comput.*, to be published. DOI: 10.1109/TSC.2019.2961895

[46] X. Luo, Z. D. Wang, and M. S. Shang, "An instance-frequency-weighted regularization scheme for non-negative latent factor analysis on high-dimensional and sparse data," *IEEE Trans. Syst., Man, Cybern.: Syst.*, to be published. DOI: 10.1109/TSMC.2019.2930525

[47] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*: *Nonlinear Phenom.*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[48] W. He, H. Y. Zhang, and L. P. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

[49] L. Tong, B. Qian, J. Yu, and C. B. Xiao, "Spectral and spatial total-variation-regularized multilayer non-negative matrix factorization for hyperspectral unmixing," *J. Appl. Remote Sens.*, vol. 13, no. 3, Article No. 036510, Sep. 2019.

[50] Y. Yuan, Z. H. Zhang, and Q. Wang, "Improved collaborative non-negative matrix factorization and total variation for hyperspectral unmixing," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 998–1010, Mar. 2020.

[51] C. C. Leng, G. R. Cai, D. D. Yu, and Z. Y. Wang, "Adaptive total-variation for non-negative matrix factorization on manifold," *Pattern Recognit. Lett.*, vol. 98, pp. 68–74, Oct. 2017.

[52] H. L. Zhang, L. M. Tang, Z. Fang, C. C. Xiang, and C. Y. Li, "Nonconvex and nonsmooth total generalized variation model for image restoration," *Signal Process.*, vol. 143, pp. 69–85, Feb. 2018.

[53] W. D. Zhao, H. M. Lu, and D. Wang, "Multisensor image fusion and enhancement in spectral total variation domain," *IEEE Trans. Multimed.*, vol. 20, no. 4, pp. 866–879, Apr. 2018.

[54] F. X. Yang, F. Ma, Z. L. Ping, and G. X. Xu, "Total variation and signature-based regularizations on coupled nonnegative matrix factorization for data fusion," *IEEE Access*, vol. 7, pp. 2695–2706, Nov. 2018.

[55] T. P. Zhang, B. Fang, Y. Y. Tang, G. H. He, and J. Wen, "Topology preserving non-negative matrix factorization for face recognition," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 574–584, Apr. 2008.

[56] H. Q. Yin and H. W. Liu, "Nonnegative matrix factorization with bounded total variational regularization for face recognition," *Pattern Recognit. Lett.*, vol. 31, no. 16, pp. 2468–2473, Dec. 2010.

[57] C. C. Leng, Z. Chen, G. R. Cai, I. Cheng, Z. H. Xiong, J. Tian, and A. Basu, "Total variation constrained graph regularized NMF for medical image registration," in *Proc. the 12th IEEE Image, Video, and Multidimensional Signal Processing Workshop*, Bordeaux, France, 2016, pp. 1–5.

[58] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, Jun. 1994.

[59] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. 13th Int. Conf. Neural Information Processing Systems*, Denver, USA, 2000, pp. 556–562.

[60] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Prob.*, vol. 19, no. 6, pp. S165–S187, Nov. 2003.

[61] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*: *Ser. B* (*Methodol.*), vol. 39, no. 1, pp. 1–22, 1977.

[62] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005.

[63] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Anchorage, USA, 2008, pp. 1221–1228.

[64] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, Manchester, UK, 1988, pp. 147–151.

[65] A. Myronenko and X. B. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.

[66] C. C. Leng, H. Zhang, G. R. Cai, Z. Chen, and A. Basu. "Total Variation Constrained Non-Negative Matrix Factorization for Medical Image Registration," [Online]. Available: http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Datasets, Accessed on: May 26, 2020.

Chengcai Leng received the Ph.D. degree in applied mathematics from Northwestern Polytechnical University, Xi'an, China, in 2012. From 2010 to 2011 and 2017 to 2018, he was a Visiting Student and Visiting Scholar with the Department of Computing Science, University of Alberta, Edmonton, Canada, respectively. He served as a Postdoctoral in the Institute of Automation, Chinese Academy of Sciences, Beijing, China from 2013 to 2016. He is currently an Associate Professor with the School of Mathematics, Northwest University, Xi'an, China. His current research interests include image processing, statistical analysis, computer vision and optical molecular imaging.

Hai Zhang received the Ph.D. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, in 2012. He was a Visiting Scholar with the Department of Statistics, University of California, Berkeley, USA from 2011 to 2012. He is currently a Professor and Chair in the Department of Financial Mathematics and Statistics at Northwest University, China. His research interests include statistical machine learning, high-dimensional statistics, and social network analysis.

Guorong Cai received Ph.D. degree in computer science from Xiamen University, Fujian, China, in 2013. Currently, he is an Associate Professor with Computer Engineering College, Jimei University, Xiamen, China. His research interests include 3D reconstruction, machine learning, object detection/recognition, and image/video retrieval.

Zhen Chen received the Ph.D. degree in mechanical design and theory from Northwestern Polytechnical University, Xi'an, China, in 2003. From 2006 to 2007, he was a Visiting Scholar with the Department of Biomedical Engineering, the University of Kansas, USA. He is currently a Professor with the School of Measuring and Optical Engineering, Nanchang Hangkong University, China. His current research interests include image understanding and measurement.

Anup Basu (M'90−SM'02) received the Ph.D. degree in computer science from the University of Maryland, College Park, USA. He was a Visiting Professor with the University of California, Riverside, USA, a Guest Professor with the Technical University of Austria, Graz, Austria, and the Director with the Hewlett-Packard Imaging Systems Instructional Laboratory, University of Alberta, Edmonton, Canada, since 1999, where he has been a Professor with the Department of Computing Science, and is currently an iCORE-NSERC Industry Research Chair. His current research interests include 3D/4D image processing and visualization especially for medical applications, multimedia in education and games, and wireless 3D multimedia transmission.