



Manifold Learning via the Principle Bundle Approach

Chen-Yun Lin^{1*}, Arin Minasian², Xin Jessica Qi³ and Hau-Tieng Wu^{1,4,5}

¹ Department of Mathematics, Duke University, Durham, NC, United States, ² Department of Electrical Engineering, University of Toronto, Toronto, ON, Canada, ³ Department of Mathematics, University of Toronto, Toronto, ON, Canada, ⁴ Department of Statistical Science, Duke University, Durham, NC, United States, ⁵ Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan

In this paper, we propose a novel principal bundle model and apply it to the image denoising problem. This model is based on the fact that the patch manifold admits canonical groups actions such as rotation. We introduce an image denoising algorithm, called the *diffusive vector non-local Euclidean median* (dvNLEM), by combining the traditional nonlocal Euclidean median (NLEM), the rotational structure in the patch space, and the diffusion distance. A theoretical analysis of dvNLEM, as well as the traditional nonlocal Euclidean median (NLEM), is provided to explain why these algorithms work. In particular, we show how accurate we could obtain the true neighbors associated with the rotationally invariant distance (RID) and Euclidean distance in the patch space when noise exists, and how we could apply the diffusion geometry to stabilize the selected metric. The dvNLEM is applied to an image database of 1,361 images and a comparison with the NLEM is provided. Different image quality assessments based on the error-sensitivity or the human visual system are applied to evaluate the performance.

Keywords: non-local means, diffusion geometry, principal bundles, patch size, image denoising

OPEN ACCESS

Edited by:

Lixin Shen,
Syracuse University, United States

Reviewed by:

Yiming Ying,
University at Albany, United States
Lei Shi,
Fudan University, China

*Correspondence:

Chen-Yun Lin
cylin@math.duke.edu

Specialty section:

This article was submitted to
Mathematics of Computation and
Data Science,
a section of the journal
Frontiers in Applied Mathematics and
Statistics

Received: 23 April 2018

Accepted: 31 May 2018

Published: 25 June 2018

Citation:

Lin C-Y, Minasian A, Qi XJ and
Wu H-T (2018) Manifold Learning via
the Principle Bundle Approach.
Front. Appl. Math. Stat. 4:21.
doi: 10.3389/fams.2018.00021

1. INTRODUCTION

A widely accepted approach to model the low dimension structure is by considering the manifold; that is, we assume that the collected dataset, while might be of high dimension, is located on a low dimensional manifold [1–3]. Under this assumption, several algorithms were proposed toward this goal, like isomap [1], locally linear embedding [4], eigenmaps (EM) [2], diffusion maps [3], Hessian LLE [5], vector diffusion maps [6, 7], non-linear independent component analysis or empirical intrinsic geometry [8, 9], alternating diffusion [10, 11], horizontal diffusion maps [12] etc.

One example of such model is the patch manifolds. It is well-known that the non-local mean (NLM) [13] algorithm leads to a better edge preservation [14], and this improvement is directly related to the diffusion process on the nonlinear geometric structure [15]. One can further consider the rotational structure of patch spaces to denoise images [16–21]. Two patches are viewed the same or called *rotationally invariant*, if they are the same up to rotation. While the patch space model and diffusion-based algorithms have been successfully applied to different fields, such as the inpainting problem [22–25] and the medical imaging problem [19, 26, 27], however, to the best of our knowledge, a companion theoretical study is lacking. It is not clear how we could correctly find the neighbors when noise exists, and why the patch size should be neither too big nor too small and thus why patch space approaches are better than pixel-based ones.

In this paper, we focus on the idea of manifold learning via the principal bundle approach and its associated theoretical analysis. We model the patch space by a principal bundle so that

rotationally invariant patches are modeled by fibers diffeomorphic to $SO(2)$. We further show that with high probability, which depends on the patch size and the noise level, one could accurately determine the neighborhood from the noisy patches (Theorem 3.2 and Remark 3.3). We apply it to the image denoising problem and propose an image denoising algorithm, dvNLEM. Although the results indicate that a suitable chosen patch size enhance the neighbor search accuracy, the error is inevitable. We then further stabilize the above theoretical results by applying the diffusion distance (DD), which has been theoretically proved to be robust to noise [28]. Based on the theoretical understanding, we consider the *diffusive vector nonlocal Euclidean median* (dvNLEM) algorithm that combines the rotational alignment, the DD, and the median. In the dvNLEM, we consider the RID where the RID of two rotationally invariant patches is 0. To reduce the noise influence, the RID is further stabilized by the DD. With the DD, we determine the neighboring patches, and evaluate the median value of the central pixels of all neighboring patches. To overcome this computational intensity, we use scale-invariant feature transform (SIFT) features as a relaxation step to estimate the neighbors.

2. MATHEMATICAL MODEL

For the i -th pixel on a given grayscale image $I \in \mathbb{R}^{N \times N}$ and an odd integer q , we associate it with a patch P_i of size $q \times q$ centered at the i -th pixel. If the i -th pixel is close to the boundary of I , we pad the image by 0 and generate the corresponding patch. Denote the set of all patches as $\mathcal{X}_I := \{P_i\}_{i=1}^{N^2} \subset \mathbb{R}^{q^2}$. To express the notation in a compact format, we stack the columns of the matrix I into a vector $I^\vee \in \mathbb{R}^{N^2}$, where the $((\ell - 1)N + 1)$ -th to the (ℓN) -th entries in I^\vee is the ℓ -th column of I , where $\ell = 1, \dots, N$. We represent the image I and patches P_i in the matrix form and the column form interchangeably.

For a clean image $I^{(c)}$, we represent its noisy image as

$$I^{(n)} := I^{(c)} + \sigma \xi, \text{ where } \xi \sim \mathcal{N}(0, I) \text{ is Gaussian noise,} \quad (1)$$

the i -th clean patch as $P_i^{(c)}$ and its associated noisy patch as

$$P_i^{(n)} = P_i^{(c)} + \sigma \xi_i, \quad (2)$$

where ξ_i is additive Gaussian white noise with mean 0 and standard deviation 1.

Definition 2.1. *The rotationally invariant distance (RID) between two patches, P_i and P_j is defined as*

$$d_{RID}(P_i, P_j) := \min_{O \in SO(2)} \|P_i - O.P_j\|, \quad (3)$$

where $\|\cdot\|$ denotes the ℓ^2 norm and $O.P_j$ means rotating the patch P_j by O .

By identifying patches up to rotation, we make the following assumption. First, recall the definition of a fiber bundle with group structure G [29].

Definition 2.2 (Fiber bundle with group structure G). *Let F and M be manifolds. A fiber bundle E with fiber F over M consists of a topological space E together with a map $\pi : E \rightarrow M$ satisfying the local triviality condition. Let G be a Lie group, for example the rotation group $SO(2)$. Let the map $\cdot : G \times F \rightarrow F$ be a smooth left action of F . That is, the map (the action of G) $\cdot : (x, g) \mapsto g.x$ from $G \times F$ to F satisfies*

$$e.g = g \text{ for all } g \in G, \quad (4)$$

where e is the neutral element of G and

$$(gh).x = g.(h.x) \text{ for all } x \in M \text{ and } g, h \in G. \quad (5)$$

This group G is often called the gauge group or the transition group. If the fiber is equal to the structure group, then (π, E, M, G) is called a principal G bundle. See Definition 5.7 in Gallier [30]. This definition is equivalent to the definition of principal bundle requiring G as a right action on E . See, Propositions 5.5 and 5.6 in Gallier [30]. Suppose $P : \mathbb{R}^2 \rightarrow \mathbb{R}$ represents a patch. For $O \in SO(2)$ and $\mathbf{s} \in \mathbb{R}^2$ expressed by a column vector, we define the action on P as

$$(O.P)(\mathbf{s}) = P(O^{-1}\mathbf{s}). \quad (6)$$

This is a left group action since for any $O_1, O_2 \in SO(2)$,

$$\begin{aligned} O_2.(O_1.P)(\mathbf{s}) &= O_1.P(O_2^{-1}\mathbf{s}) \\ &= P(O_1^{-1}O_2^{-1}\mathbf{s}) = (O_2O_1).P(\mathbf{s}) \end{aligned} \quad (7)$$

Since each fiber, including a patch and its rotated patches, can be identified as S^1 , and $SO(2)$ is diffeomorphic to S^1 , the patch space \mathcal{X}_I could be viewed as a principal $SO(2)$ bundle.

Assumption 2.3 ($SO(2)$ -principal bundle). *The patch space \mathcal{X}_I is a finite subset of the fiber bundle E with $\pi : E \rightarrow M$ and a left $SO(2)$ group action so that the quotient space $\mathcal{X}_I/SO(2)$ is a finite subset of the corresponding base manifold.*

Note that the $SO(2)$ action preserves the fiber that is diffeomorphic to $SO(2)$. In Supplementary Material, we further discuss how patch spaces naturally carry principal bundle structures.

3. THEORETICAL ANALYSIS

The theoretical findings below show that with high probability, the clean patch neighbors could be “accurately” evaluated from the noisy patch neighbors with respect to RID. Lemma 3.1 gives that the square distance between non-overlapping patches is distributed according to a non central chi-squared distribution. We approximate the square distance between overlapping patches by a chi-squared distribution. Theorem 3.2 gives the large deviation bound for the RID, which allows us to take the RID into account to denoise the image. If we omit the rotation action, then we obtain the large deviation bound for the Euclidean distance (Corollary 3.4), which explains why the traditional NLM/NLEM algorithms work.

We introduce the following sets associated with $P_i^{(n)}$ and $P_j^{(n)}$, and $O \in SO(2)$:

$$K_O(O) := \{(a, b) \mid \xi_i(a) = [O.\xi_j](b)\} \tag{8}$$

whose cardinality is the size of the overlapping pixels;

$$K_S(O) := \{(a, b) \mid \xi_i(a) = [O.\xi_j](b); \xi_i(b) = [O.\xi_j](a)\} \tag{9}$$

which is associated with the “swapped” pixel indices after rotation; and

$$K_I(O) := \{(a, a) \mid \xi_i(a) = [O.\xi_j](a)\} \tag{10}$$

which is associated with the overlapped pixels with “identical indices” after rotation and depends on O . By definition, $K_I(O) \subseteq K_S(O) \subseteq K_O(O)$. Note that $K_I(O)$ has only at most one element. For the NLEM, $K_I(O)$ and $K_S(O)$ are both empty.

Lemma 3.1. For any two patches $P_i^{(n)}$ and $P_j^{(n)}$ and a fixed O , denote $P = P_i^{(c)} - O.P_j^{(c)}$ and $X = \frac{\|P_i^{(n)} - O.P_j^{(n)}\|^2}{2\sigma^2}$.

1. If $P_i^{(n)}$ and $P_j^{(n)}$ are disjoint, then $X \sim \chi_{q^2}^2(\frac{\|P\|^2}{2\sigma^2})$ is a noncentral χ^2 random variable with q^2 degrees of freedom and non-centrality parameter $\frac{\|P\|^2}{2\sigma^2}$.
2. If $P_i^{(n)}$ and $P_j^{(n)}$ overlap, then

$$\mathbb{E}(X) = \frac{\|P\|^2}{2\sigma^2} + q^2 - |K_I|, \text{ and} \tag{11}$$

$$\text{Var}(X) = \frac{2}{\sigma^2}(\|P\|^2 - \sum_{(a,b) \in K_O \setminus K_I} P(a)P(b) + 2q^2 + |K_S| + |K_O| - 4|K_I|) \tag{12}$$

Moreover, X can be approximated by $Y \sim c\chi_f^2$, where

$$c = \frac{\text{Var}(X)}{2\mathbb{E}(X)} \text{ and } f = \frac{2\mathbb{E}^2(X)}{\text{Var}(X)}. \tag{13}$$

Since $2\sigma^2(q^2 - |K_I|) > 0$, $\|P_i^{(n)} - O.P_j^{(n)}\|^2$ is a biased estimator of $\|P_i^{(c)} - O.P_j^{(c)}\|^2$. When two patches overlap, depending on the clean patches’ structure, the RID estimator from the noisy patches might be biased toward overlapped patches. Thus, if the overlapping patches are included in the denoising process, the search of the nearest neighboring patches might be biased to the “local patches” that have overlaps.

Theorem 3.2. Suppose $P_i^{(c)}$ and $P_j^{(c)}$ are disjoint patches and $\sigma q < \epsilon$. If $d_{RID}(P_i^{(c)}, P_j^{(c)}) < \epsilon$, then

$$\Pr(d_{RID}(P_i^{(n)}, P_j^{(n)}) \geq 2\epsilon) \leq \exp\left(-\left(\sqrt{\frac{\epsilon^2}{\sigma^2} - \frac{q^2}{4}} - \sqrt{\frac{q^2}{4} + \frac{d_{RID}^2(P_i^{(c)}, P_j^{(c)})}{4\sigma^2}}\right)^2\right) \tag{14}$$

which decreases as q decreases. If $d_{RID}(P_i^{(c)}, P_j^{(c)}) > 2\epsilon$, then

$$\Pr(d_{RID}(P_i^{(n)}, P_j^{(n)}) < \epsilon) \leq \exp\left(-\left(\frac{2\sigma^2 q^2 + d_{RID}^2(P_i^{(c)}, P_j^{(c)}) - \epsilon^2}{4\sigma\sqrt{\sigma^2 q^2 + d_{RID}^2(P_i^{(c)}, P_j^{(c)})}}\right)^2\right) \tag{15}$$

which decreases as q increases.

Remark 3.3. For overlapping patches, $\frac{\|P_i^{(n)} - O.P_j^{(n)}\|^2}{2\sigma^2}$ can be approximated by $Y \sim c\chi_f^2$ where c and f are given by (13). By direct computation, one can see c decreases and the degree of freedom f increases as q increases. This suggests similar conclusion as in Theorem 3.2 due to the behaviour of chi-squared distributions. That is, if $d_{RID}(P_i^{(c)}, P_j^{(c)}) < \epsilon$, the tail-area probability $\Pr(d_{RID}(P_i^{(n)}, P_j^{(n)}) \geq 2\epsilon)$ decreases as q increases; if $d_{RID}(P_i^{(c)}, P_j^{(c)}) > 2\epsilon$, then $\Pr(d_{RID}(P_i^{(n)}, P_j^{(n)}) < \epsilon)$ decreases as q increases.

The above observations suggest that we need to choose a suitable patch size q to rule out unwanted patches and keep desired patches. In practice, we found that an odd value q between 7 and 15 leads to a good performance, but the optimal q depends on the image. If O is the identity, the same argument explains why the ℓ^2 distance between two clean patches could be well approximated by noisy patches.

Corollary 3.4. Suppose $P_i^{(c)}$ and $P_j^{(c)}$ are disjoint patches and $\sigma q < \epsilon$. If $\|P_i^{(c)} - P_j^{(c)}\| < \epsilon$, then

$$\Pr(\|P_i^{(n)} - P_j^{(n)}\| \geq 2\epsilon) \leq \exp\left(-\left(\sqrt{\frac{\epsilon^2}{\sigma^2} - \frac{q^2}{4}} - \sqrt{\frac{q^2}{4} + \frac{\|P_i^{(c)} - P_j^{(c)}\|^2}{4\sigma^2}}\right)^2\right) \tag{16}$$

which decreases as q decreases. If $\|P_i^{(c)} - P_j^{(c)}\| > 2\epsilon$, then

$$\Pr(\|P_i^{(n)} - P_j^{(n)}\| < \epsilon) \leq \exp\left(-\left(\frac{2\sigma^2 q^2 + \|P_i^{(c)} - P_j^{(c)}\|^2 - \epsilon^2}{4\sigma\sqrt{\sigma^2 q^2 + \|P_i^{(c)} - P_j^{(c)}\|^2}}\right)^2\right) \tag{17}$$

which decreases as q increases.

The quantity $\sigma^2 q^2$ could be understood as the “total energy” of the added noise, and the condition $\sigma^2 q^2 < \epsilon^2$ means that the RID estimated from two noisy patches is controlled by the square root of the energy of the noise. With this energy viewpoint, we could apply the technique developed in El Karoui and Wu [28]. However, we carried out the proof in the above way to emphasize the main purpose of the RID, and to find the true neighbors and the dependence on the patch size.

Before closing this section, we mention that even if by choosing suitable patch size we could more accurately find the neighbors, the error is inevitable. Thus, to further find correct neighbourhoods, we need more tools to stabilize the found neighbors. In this work, we apply the diffusion geometry framework to stabilize the RID for the image denoising purpose, which is detailed in the next section.

4. DIFFUSIVE VECTOR NON-LOCAL EUCLIDEAN MEDIAN ALGORITHM

Take a clean grayscale image denoted as $I^{(c)} \in \mathbb{R}^{N \times N}$ normalized to be of mean 0 and standard deviation 1. Take the associated noisy image $I^{(n)}$, the noisy patches $P_i^{(n)}$ with $\sigma > 0$. The goal of the *denoising problem* is finding an algorithm that will recover $I^{(c)}$ from $I^{(n)}$ as accurately as possible. Based on the theoretical analysis shown in section 3, we now discuss the dvNLEM algorithm.

4.1. dvNLEM Algorithm

Theorem 3.2 essentially says that even when noise exists, we could still find the true neighbors with the RID with high probability; in other words, it is possible to find wrong neighbors. Based on the principal bundle model, we apply the diffusion map and the DD to “filter out” wrong neighbors and improve the overall accuracy. The idea is to combine the fiber bundle structure underlying the patch space with the diffusion map (DM). With the RID, define the affinity matrix $W \in \mathbb{R}^{N^2 \times N^2}$ by

$$W_{ij} = \exp(-d_{\text{RID}}^2(P_i^{(n)}, P_j^{(n)})/\epsilon), \quad (18)$$

where $i, j = 1, \dots, N^2$ and $\epsilon > 0$ is the pre-determined bandwidth. With the affinity matrix W , the graph Laplacian and the associated transition matrix. $A = D^{-1}W$ where D is the *degree matrix* which is diagonal and defined as $D_{ii} = \sum_{j=1}^{N^2} W_{ij}$. By taking the top m eigenvalues and eigenvectors of the transition matrix, we then embed each patch into a low-dimensional space via the DM (24) and calculate the DD to evaluate the true neighbors of each patch. For each patch $P_i^{(n)}$, we identify its $N_1 \in \mathbb{N}$ nearest neighbors in the sense of DD and denote this set as $N_{\text{DD}}(i)$. Based on the robustness property of the DM [28, 31], this step acts as an additional filtering procedure to dismiss the patches in the initial nearest neighbors set determined by the RID. The denoised image, denoted as $\tilde{I}^{(\text{dvNLEM})} \in \mathbb{R}^{N \times N}$.

If we identify a patch $P_i^{(n)}$ its $N_1 \in \mathbb{N}$ nearest neighbors associated with the RID instead of DD, denoted as $N_{\text{RID}}(i)$, and denoise the noisy image by the Euclidean median, we obtain an *intermittent denoising algorithm*, which we call the *vector nonlocal Euclidean median* (vNLEM) and denote the denoised image as $\tilde{I}^{(\text{vNLEM})}$.

4.1.1. Numerical Techniques to Speed Up the Computation–Search Window

While the proposed algorithm seems a straightforward generalization of the NLM/NLEM/NLPR, there are numerical issues we have to handle.

Obtaining the pairwise distances between all pairs of patches is computationally intense and time-consuming. One

practical solution is to only consider a pre-assigned number of nearest neighbors. By doing so, we simultaneously reduce the computational time and the memory required to save W .

We further limit our algorithm to consider only patches within a given search window centered at the reference patch. For a patch $P_i^{(n)}$, we consider the search window of size $(2N_2 + 1) \times (2N_2 + 1)$ centered at the patch

$$S_i = \{P_j^{(n)} \mid \text{the difference of } i \text{ and } j \text{ is bounded by } N_2 \},$$

where $N_2 \in \mathbb{N}$ so that $(2N_2 + 1) \times (2N_2 + 1) > N_1$. With this search window, we form the affinity matrix by the following:

$$W_{ij} = \begin{cases} \exp(-d_{\text{RID}}^2(P_i^{(n)}, P_j^{(n)})/\epsilon), & \text{for } P_j^{(n)} \in S_i \\ 0, & \text{otherwise} \end{cases}. \quad (19)$$

4.1.2. Numerical Techniques to Improve the RID Evaluation–SIFT

Finding the RID between two given patches incurs huge computational costs. Also, performing a direct numerical rotation might lead to a non-negligible error and deviate the estimated RID. To alleviate these two troubles and facilitate the derivation of the affinity matrix, we use the scale invariant feature transform (SIFT) [32] to approximate the RID. We could consider the central moments [16, 17, 19], the curvelet transform [21], or the Fast Affine Template Match [33] algorithm, to capture the rotational feature. To simplify the discussion, we focus on the SIFT algorithm.

The particular feature we extract by SIFT is the orientation feature. For each pixel, based on the local image gradient direction, an orientation angle is calculated and assigned. We use this orientation to approximate the RID distances between the patches. Denote the orientation for the center of $P_i^{(n)}$ as $\theta_i^{(n)}$. The relative angle between $P_i^{(n)}$ and $P_j^{(n)}$ achieving the RID is approximated by $\theta_i^{(n)} - \theta_j^{(n)}$. We then rotate $P_j^{(n)}$ by $R_{\theta_j^{(n)}} \cdot R_{\theta_j^{(n)}}^{-1} \cdot P_j^{(n)}$, where $R_\theta \in SO(2)$ means the rotation by θ degrees. Next, we perform the exhaustive search over a small range centered around the estimated angular relationship $\theta_i^{(n)} - \theta_j^{(n)}$:

$$\theta_{ij} := \underset{\theta \in \{\theta : |\theta - (\theta_i^{(n)} - \theta_j^{(n)})| < \theta_1\}}{\text{argmin}} \|U_k P_i^{(n)} - R_\theta U_k P_j^{(n)}\|, \quad (20)$$

where $U_k : \mathbb{R}^{q^2} \rightarrow \mathbb{R}^{k^2 q^2}$ is the chosen upsampling operator that increases the sampling rate of the patch $P_i^{(n)}$ by $k \in \mathbb{N}$ times and $\theta_1 > 0$ is the parameter chosen by the user. The use of U_k is to improve the accuracy of numerical rotation. Hence the estimated RID is

$$\tilde{d}_{\text{RID}}(P_i^{(n)}, P_j^{(n)}) := \|P_i^{(n)} - R_{\theta_{ij}} P_j^{(n)}\|. \quad (21)$$

We use the following affinity weights in (25):

$$W'_{ij} = \exp(-\tilde{d}_{\text{RID}}^2(P_i^{(n)}, P_j^{(n)})/\epsilon), \text{ where } j \in S_i. \quad (22)$$

4.1.3. The Proposed dvNLEM Algorithm

The proposed algorithms after the above modifications are summarized in Algorithm 1. We slightly abuse the terminology and also call the modified denoising scheme in (25) based on the estimated DD the dvNLEM and denote $\tilde{I}^{(dvNLEM)}$ as the denoised image by dvNLEM. Similarly, we call the modified denoising

Algorithm 1 Diffusive vector non-local Euclidean median algorithm.

Input : Noisy image $I^{(n)}$, patch size $q \in \mathbb{N}$, the number of nearest neighbors $N_1 \in \mathbb{N}$, the search window size $N_2 \in \mathbb{N}$, the kernel bandwidth $\epsilon > 0$, the DM embedding dimension $m \in \mathbb{N}$, the diffusion time $t > 0$, and the power $0 < \gamma \leq 1$.

Output : Denoised image \tilde{I} .

[pre-1] Pad the image array with a border of $\lceil q/2 \rceil$ pixels.

[pre-2] Create the patch space $\mathcal{X}^{(n)} := \{P_i^{(n)}\}_{i=1}^{N^2} \subset \mathbb{R}^{q^2}$, where the center of $P_i^{(n)}$ is $I^{(n)}(i)$.

[pre-3] Find SIFT orientation feature for each patch and form an affinity matrix W using these orientations from the search window S_i of size $(N_2 + 1) \times (N_2 + 1)$ according to equation (19).

[vNLEM. Step 1] For each i , find N_1 nearest neighbours from S_i according to W .

[vNLEM. Step 2] Find the more accurate estimation of RID, \tilde{d}_{RID} in (21), and form $N_{RID}(i)$ that contains $\lceil N_1/2 \rceil$ patches that are closer to patch i according to \tilde{d}_{RID} , where $\lceil x \rceil$ means the smallest integer greater than or equal to $x \in \mathbb{R}$.

[vNLEM. Step 3] For each i , set $\tilde{I}^{(vNLEM)}(i)$ to be the center point of

$$\operatorname{argmin}_{P \in \mathbb{R}^{q \times q}} \sum_{P_j^{(n)} \in N_{RID}(i)} W'_{ij} \|P - O_{ij} \cdot P_j^{(n)}\|^\gamma. \quad (23)$$

[dvNLEM. Step 1] Form the eigenvalue decomposition of $D^{-1}W$, where $D \in \mathbb{R}^{N^2 \times N^2}$ is the diagonal matrix determined by $D_{ii} = \sum_{j=1}^{N^2} W_{ij}$.

[dvNLEM. Step 2] Embed $P_i^{(n)}$ into \mathbb{R}^m by

$$\Phi_t^{(m)}(P_i^{(n)}) = (\lambda_2^t \phi_2(i), \dots, \lambda_{m+1}^t \phi_{m+1}(i)) \quad (24)$$

and evaluate the DD between patches.

[dvNLEM. Step 3] For each $P_i^{(n)}$, find N_1 nearest neighbours in terms of DD.

[dvNLEM. Step 4] Find $\lceil N_1/2 \rceil$ closest patches with respect to \tilde{d}_{RID} among the N_1 patches from the previous step to form $N_{DD}(i)$.

[dvNLEM. Step 5] For each i , set $\tilde{I}^{(dvNLEM)}(i)$ to be the center point of

$$\operatorname{argmin}_{P \in \mathbb{R}^{q \times q}} \sum_{P_j^{(n)} \in N_{DD}(i)} W'_{ij} \|P - O_{ij} \cdot P_j^{(n)}\|^\gamma. \quad (25)$$

scheme in (23) based on the approximated RID (22) the vNLEM algorithm and denote $\tilde{I}^{(vNLEM)}$ as the denoised image by vNLEM.

5. IMAGE QUALITY ASSESSMENT

In the literature, the denoising performance is commonly evaluated by the “error-based” measurements, such as the signal-to-noise ratio (SNR) or the peak SNR. However, it is well known that those error-based quantities might capture only partial information of the image quality. In this study, we consider different image quality assessment (IQA) methods that additionally take different aspects of the image into account, like the human visual system, to evaluate the performance of different denoising algorithms. IQA is an important subfield in image processing. The goal is to find an index quantifying “how good” an image is that is suitable for different scenarios. We consider measures of two major. The first category consists of objective measures based on a chosen theoretical model without taking the human visual system (HVS) into account. The second category consists of objective measures based on models taking the HVS into account. Below we summarize these measures. Denote the clean image as $I \in \mathbb{R}^{N \times N}$. We are concerned with how close the noisy observation $I + \sigma\xi$ is to I , or the denoised image \tilde{I} is to I .

The SNR and peak-signal-to-noise ratio (PSNR) belong to the first category.

While SNR and PSNR are widely applied IQA’s in the field, they do not necessarily tell us all aspects of how well the denoising methods performed. For example, they do not readily capture the edge preserving capability of an algorithm.

To capture the edge preservation performance, we consider the third measurement, the Sobolev index [34], which also belongs to the first category. Let \hat{I} and $\hat{\tilde{I}}$ denote the discrete Fourier transforms of I and \tilde{I} , respectively. The Sobolev index of order κ is then defined by the Sobolev norm, and is given by

$$\text{SOB} = \left[\frac{1}{|\Omega|^2} \sum_{\omega \in \Omega} (1 + |\eta_\omega|^2)^\kappa |\hat{I}(\omega) - \hat{\tilde{I}}(\omega)|^2 \right]^{1/2}, \quad (26)$$

where Ω is the lattice of the frequency domain and η_ω is the two-dimensional frequency vector associated with $\omega \in \Omega$.

We further consider the earth mover’s distance (EMD) [35]. The EMD between two probability distributions μ and ν on \mathbb{R} is defined as

$$d_{OT}(\mu, \nu) := \int_{\mathbb{R}} |f_\mu(x) - f_\nu(x)| dx, \quad (27)$$

where $f_\mu(x) := \int_{-\infty}^x d\mu$ is the cumulative distribution function of μ and similarly for f_ν . We will evaluate the EMD to compare how close the distribution of the estimated noise is to the added noise.

While the above measurements have been widely applied in different problems and provide useful information, it has been well-accepted that they do not capture all aspects from the perspective of image quality. In generally, it is not statistically consistent with human observers [36].

Several metrics have been designed in the past decades to faithfully take the HVS into account. These metrics emphasize the importance of luminance, the contrast, and the frequency/phase content. We consider the state-of-art measurement the Feature SIMilarity (FSIM) index [36] in this category. The FSIM is based on the model that the HVS perceives an image mainly based on its low-level features, such as edges and zero crossings, and it separates the similarity measurement task into phase congruency and gradient magnitude. Here we summarize the FSIM index. Suppose the dynamical range of the image is \mathcal{R} . The definition of FSIM depends on the definition of the phase congruency and gradient magnitude. The phase congruent of I at i , denoted as $P_I(i)$, and the gradient magnitude of I at i , denoted as $G_I(i)$, are defined in Zhang et al. [36]. Similarly, we can define $P_{\tilde{I}}(i)$ and $G_{\tilde{I}}(i)$. The FSIM between I and \tilde{I} is defined as

$$FSIM(I, \tilde{I}) := \frac{\sum_{i=1}^{N^2} S_L(i) P_m(i)}{\sum_{i=1}^{N^2} P_m(i)}, \quad (28)$$

where

$$P_m(i) = \max\{P_I(i), P_{\tilde{I}}(i)\}, \quad S_L(i) := S_p(i) S_G(i), \quad (29)$$

$$S_p(i) := \frac{2P_I(i)P_{\tilde{I}}(i) + T_1}{P_I(i)^2 + P_{\tilde{I}}(i)^2 + T_1}, \quad (30)$$

and

$$S_G(i) := \frac{2G_I(i)G_{\tilde{I}}(i) + T_2}{G_I(i)^2 + G_{\tilde{I}}(i)^2 + T_2}. \quad (31)$$

Here, we follow [36] and choose $T_1 = 0.85$ and $T_2 = 160$. There are several other measures of this kind in the field, and we refer interested readers to Zhang et al. [36] and Wang et al. [37] for a review of these indices.

6. NUMERICAL RESULT

We fix the following parameters for NLEM, vNLEM, and dvNLEM for a fair comparison. Fix $q = 13$. We build 13×13 patches around each pixel of the noisy image. We chose $\epsilon = (16.5)^2$, the number of nearest neighbors as $N_1 = 100$, the size of the search window for creating the initial affinity matrix is determined by $N_2 = 10$; that is, 21×21 neighbours in each the search window. The θ_l in (20) is set to 30 degrees, the upsampling operator U_k is implemented by the bicubic interpolation, and k is set to 2. After building the transition matrix, we choose $m = 30$ to evaluate the DM and DD. We select $\gamma = 0.1$ for the final denoising step. The Matlab code is available via request.

To compare our results with those of the NLEM algorithm, we also preformed the NLEM denoising with $\epsilon = (6.5)^{21}$ with the same patch size $q = 13$ and search window 21×21 . The kernel bandwidth is chosen to give the best performance for the NLEM algorithm in terms of SNR and PSNR.

¹The code is available in <https://www.mathworks.com/matlabcentral/fileexchange/40624-non-local-patch-regression>

In **Table 1** we report the different IQA metrics, including SNR, PSNR, RMS, SOB, OT, and FSIM discussed previously as well as the computational time, by running the three denoising algorithms on 1,361 sample images of size 512×512^2 . There are 98 images for animals, 143 images for flowers, 52 images for fruits, 115 images for landscapes, 450 images for faces, 419 images for manmade structures, and 44 miscellaneous images. The SOB metric is applied to the image recovery error. This measure particularly reflects the amount of edge information wiped out due to the denoising process. Therefore, the scheme with a lower SOB index performs the better. For the other indices, the higher the index is, the better the performance is. Under the null hypothesis that the performance of two algorithms is the same, we reject the hypothesis by the Mann–Whitney U -test with the p value less than 10^{-4} . Note that based on the overall statistics, vNLEM and dvNLEM outperform NLEM statistically significantly on all IQA metrics. On the other hand, we cannot distinguish the performance of vNLEM and dvNLEM statistically, except on the FSIM index. This result suggests that dvNLEM could better recover features sensitive to HVS.

The execution times based on 17 images are 501.8 ± 203.3 , 1489.8 ± 26.1 , and 1619 ± 34.8 s for NLEM, vNLEM, and dvNLEM, respectively. This execution time is obtained on a PC with 8 Gb of RAM using a single core from Intel Corei7 CPU with a clock speed of 3.7 GHz running on Microsoft Windows 7.

6.1. Application to the Cytometry Problem

In this last example, we apply the vNLEM and dvNLEM to the third-harmonic-generation (THG) microscopy image. The goal of the THG microscopy-based imaging cytometry is to automatically differentiate and count different types of blood cells with less blood *ex vivo*, or even *in vivo* [38]. One of the many strengths of THG is reflecting the granularity of leukocytes, which allows us to apply image processing techniques for the automatic classification. However, the raw data is noisy most of time, and a denoising technique is needed. We now apply the

TABLE 1 | Summary statistics over 1,361 images of different denoising algorithms evaluated by different image quality assessment metrics.

	NLEM	vNLEM	dvNLEM
PSNR (dB)	18.78 ± 2.92* [#]	19.49 ± 2.72*	19.62 ± 2.81 [#]
SNR (dB)	13.33 ± 2.78* [#]	14.04 ± 2.36*	14.18 ± 2.49 [#]
RMS×100 (a.u.)	5.77 ± 1.58* [#]	5.35 ± 1.33*	5.24 ± 1.36 [#]
SOB×100 (a.u.)	5.9 ± 1.63* [#]	5.45 ± 1.37*	5.35 ± 1.4 [#]
OT×100 (a.u.)	0.59 ± 0.4* [#]	0.32 ± 0.16*	0.35 ± 0.22 [#]
FSIM×100 (a.u.)	88.33 ± 2.98* [#]	89.64 ± 2.13* [†]	90.03 ± 2.13 ^{#†}

*[#] $p < 10^{-8}$. [†] $p < 10^{-6}$. au, the arbitrary unit.

²The images are collected from :

- Caltech-UCSD Birds-200-2011 collection at : <http://www.vision.caltech.edu/visipedia/CUB-200-2011.html>
- Digital Image Processing, 3rd ed, by Gonzalez and Woods at : http://www.imageprocessingplace.com/DIP-3E/dip3e_book_images_downloads.htm
- USC-SIPI image database at: <http://sipi.usc.edu/database/>

NLEM, vNLEM, and dvNLEM to the THG sectioning image of the whole blood smear at 1 hour post blood sampling. The data is provided by Professor Tzu-Ming Liu, Faculty of Health Sciences, University of Macau. The result is shown in **Figure 1**. Note that since we do not have the “ground truth” for a comparison, we only show the FSM for the quality evaluation purpose. While the result is encouraging, a systematical study of the problem, and a systematic comparison of the proposed algorithm with existing algorithms is needed. The result will be reported in a future work.

7. CONCLUSION AND DISCUSSION

In this paper, a principal bundle structure is applied to model the patch space and describe the rotational structure, based on which a theoretical analysis of the RID is provided to explain how accurate we could obtain the true neighbors when noise exists. Based on the well studied DD, the RID could be further stabilized when noise exists; that is, we nonlinearly filter out the wrongly found nearest neighbors. The theoretical result also serves as an explanation of why NLM and NLEM work well. The theoretical results lead to an image denoising algorithm, vNLEM and dvNLEM, by combining the traditional NLEM, the rotational structure in the patch space, and the DD. Numerically, we apply

the commonly applied SIFT as a relaxation step to estimate the rotational alignment. The numerical simulation provides positive evidence of the potential of the proposed algorithm. The potential of the proposed model, algorithms, and the associated theory are statistically supported by a large image database composed of 1,361 images. Below, we discuss the limitations of the current work and several future works.

First, the computational complexity needs to be further improved. Note that the main difference between the vNLEM/dvNLEM and the NLEM algorithms is the chosen metric. In the vNLEM, there is no fast algorithm yet to determine the nearest neighbors with respect to the RID, to the best of our knowledge. Although we have delegated the problem of evaluating the RID distance to that of evaluating the SIFT distance, the numerical performance still has a significant room for improvement.

Second, although the manifold model has been widely accepted in the field, and the DD applied in dvNLEM is based on the manifold structure, it is certainly arguable if in general a patch space could be well approximated by a manifold. On one hand, we might need different models, and hence different metrics, for different kinds of images. For example, while the RID helps reduce the dimension of the patch space of a “structured” image,

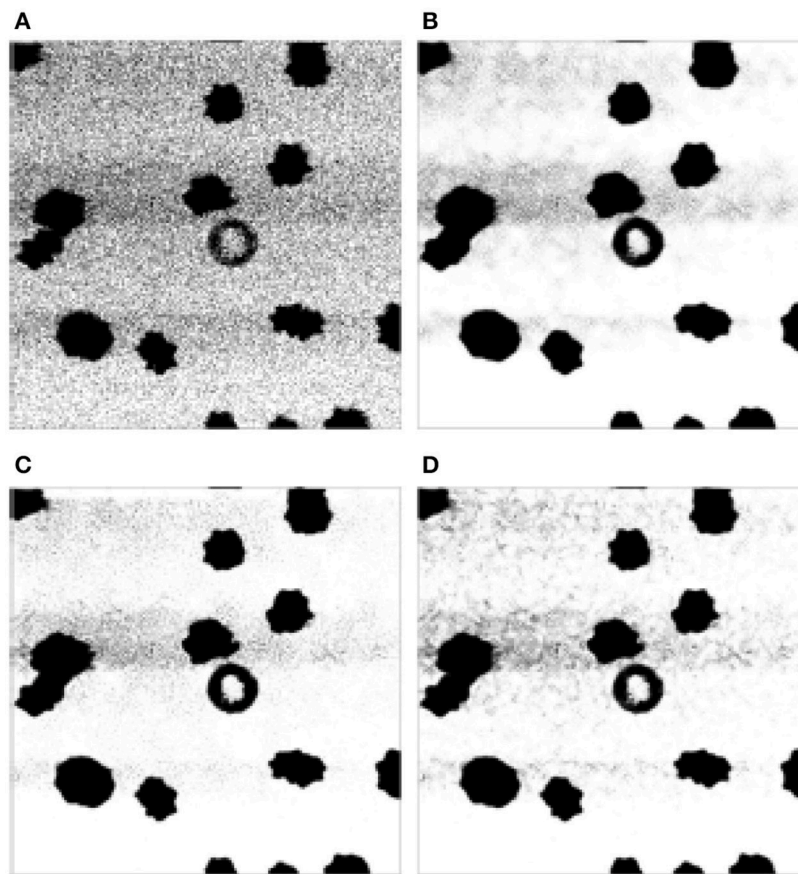


FIGURE 1 | The cytometry image. Since the “ground truth” is not available for a comparison, we only show the FSM for the quality evaluation purpose. **(A)** Original image. **(B)** NLEM, FSIM = 0.947. **(C)** vNLEM, FSIM = 0.967. **(D)** dvNLEM, FSIM = 0.97.

its deterministic nature might render it unsuitable for analyzing a “texture” image, since the texture features are stochastic in nature. In short, it might be beneficial to take the metrics designed for the texture analysis into account. On the other hand, we could consider to segment the given noisy image into different categories, and run the vNLEM on each category. This segmentation step is related to the “multi-manifold model” considered in the literature [39, 40], and could be understood as a generalization of the search window method used in this paper.

Third, we should consider different structures in the denoising procedure. In addition to taking the rotation group to fibrate the patch space, it is an intuitive generalization to further consider other groups, like the dilation group or even the general linear group. Also, while the current work focuses on grayscale images, the proposed algorithm has the potential to be generalized to colored images. In colored images, more structures, like the color space, will be taken into account. Furthermore, in practice we would expect to have more than one image from the practical problem. Under the assumption that the noise behavior is similar, it is of great interest to see if we could further improve the denoise performance by denoising multiple available images simultaneously.

Fourth, note that the proposed algorithm could be understood as aiming to reduce the error introduced to the clean image. However, it has been widely argued in the IQA society that simply reducing the error might not lead to the optimal result in all scenarios. It might be more important to take the human perception into account, if the images are meant to be watched by a human being. While the proposed vNLEM provides a satisfactory result by the FSIM evaluation, note that the “features” considered in the FSIM are not used in the algorithm. It is reasonable to expect that by taking these features into account, we could further improve the result.

Fifth, in this paper we focus only on comparing our algorithm with the NLEM and study the corresponding theoretical properties and the geometric structure of the underlying patch space model. For the image denoising purpose, there are several other image denoising algorithms available in the field, and we will do a systematic comparison in a upcoming report. For example, while not specifically indicated, the widely used algorithm BM3D [41] and its generalizations, for example [42], and BPFA are also based on the patch space model. We could view the sparsity structure used in BM3D and the Bayesian approach

in BPFA as a different way to design a “metric” to compare different patches.

Last but not least, although we compared the algorithm on a big image database and reported the statistical significance, note that statistical significance does not imply practical significance. Particularly, the included images are not exhaustive. A more systematic comparison is thus needed. Furthermore, we do not claim that the proposed denoise algorithm is better than any of the state-of-the-art image denoising algorithm, like the BM3D. In practice, the overall performance depends on the problems encountered, and the specific applications, like the cytometry problem, and a systematic comparison with other algorithms will be discussed in a upcoming research report.

AUTHOR CONTRIBUTIONS

AM and XQ contribute to the code implementation. C-YL and H-TW contribute to the theoretical development and manuscript preparation.

FUNDING

H-TW’s research is partially supported by Sloan Research Fellow FR-2015-65363 and partially by Connaught New Researcher grant 498992.

ACKNOWLEDGMENTS

H-TW would like to thank the valuable discussions with Professor Ingrid Daubechies and Professor Amit Singer. C-YL would like to thank Professor Chiahui Huang for her helpful discussions. The authors would like to thank Professor Tzu-Ming Liu for sharing the cytometry image. H-TW’s research is partially supported by Ministry of Science and Technology 107-2119-M-002-016-, NCTS Taiwan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fams.2018.00021/full#supplementary-material>

REFERENCES

1. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* (2000) **290**:2319–23. doi: 10.1126/science.290.5500.2319
2. Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* (2003) **15**:1373–96. doi: 10.1162/089976603321780317
3. Coifman RR, Lafon S. Diffusion maps. *Appl Comput Harmon Anal.* (2006) **21**:5–30. doi: 10.1016/j.acha.2006.04.006
4. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science* (2000) **290**:2323–6. doi: 10.1126/science.290.5500.2323
5. Donoho DL, Grimes C. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci USA.* (2003) **100**:5591–6. doi: 10.1073/pnas.1031596100
6. Singer A, Wu HT. Vector diffusion maps and the connection laplacian. *Commun Pure Appl Math.* (2012) **65**:1067–144. doi: 10.1002/cpa.21395
7. Singer A, Wu HT. Spectral convergence of the connection Laplacian from random samples. *Inf Infer.* (2017) **6**:58–123. doi: 10.1093/imaiai/iaw016
8. Singer A, Coifman RR. Non-linear independent component analysis with diffusion maps. *Appl Comput Harmon Anal.* (2008) **25**:226–39. doi: 10.1016/j.acha.2007.11.001
9. Talmon R, Coifman RR. Empirical intrinsic geometry for intrinsic modeling and nonlinear filtering. *Proc Nat Acad Sci USA.* (2013) **110**:12535–40. doi: 10.1073/pnas.1307298110

10. Lederman RR, Talmon R. Learning the geometry of common latent variables using alternating-diffusion. *Appl Comput Harmon Anal.* (2015) **44**:509–36. doi: 10.1016/j.acha.2015.09.002
11. Talmon R, Wu HT. Discovering a latent common manifold with alternating diffusion for multimodal sensor data analysis. *Appl Comput Harmon Anal* 2018.
12. Gao T. *The Patch Manifolds of Many Natural Images Have Low Dimensional Structure.* Available online at: <https://arxiv.org/pdf/1602.02330.pdf> (2016).
13. Buades A, Coll B. A non-local algorithm for image denoising. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* San Diego, CA (2005) p. 60–5.
14. Jain P, Taygi V. A survey of edge-preserving image denoising methods. *Inf Syst Front.* (2016) **18**:159–70. doi: 10.1007/s10796-014-9527-0
15. Singer A, Shkolnisky Y, Nadler B. Diffusion interpretation of nonlocal neighborhood filters for signal denoising. *SIAM J Imaging Sci.* (2009) **2**:118–39. doi: 10.1137/070712146
16. Zimmer S, Didas S, Weickert J. A rotationally invariant block matching strategy improving image denoising with non-local means. In: *Proc 2008 Int Workshop on Local and Non-local Approximation in Image Processing, LNLA 2008 Lausanne* (2008).
17. Grewenig S, Zimmer S, Weickert J. Rotationally invariant similarity measures for nonlocal image denoising. *J Vis Commun Image Represent.* (2011) **22**:117–30. doi: 10.1016/j.jvcir.2010.11.001
18. Qi X. *Vector Nonlocal Mean Filter.* Toronto, ON: University of Toronto (2015). Available online at: <Http://hdl.handle.net/1807/70528>
19. Guizard N, Coupe P, Fonov VS, Manjon JV, Arnold DL, Collins DL. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. *Neuroimage* (2015) **8**:376–89. doi: 10.1016/j.nicl.2015.05.001
20. Sreehari S, Venkatakrishnan SV, Drummy LF, Simmons JP, Bouman CA. Rotationally-invariant non-local means for image denoising and tomography. In: *International Conference on Image Processing (ICIP)* Quebec City, QC (2015).
21. Zhang D, He J, Du M. Image restoration via patch orientation-based low-rank matrix approximation and nonlocal means. *J Electron Imaging* (2016) **25**:023021. doi: 10.1117/1.JEI.25.2.023021
22. Gepshtein S, Keller Y. Image completion by diffusion maps and spectral relaxation. In: *IEEE Transactions on Image Processing : A Publication of the IEEE Signal Processing Society* (2013) p. 2983–94. doi: 10.1109/TIP.2013.2237916
23. Osher S, Shi Z, Zhu W. *Low Dimensional Manifold Model for Image Processing.* Tech report, UCLA, Tech Rep CAM report 1604. (2016).
24. Yin R, Gao T, Lu Y, Daubechies I. A tale of two bases: local-nonlocal regularization on image patches with convolution framelets. *SIAM J Imaging Sci.* (2017) **10**:711–50. doi: 10.1137/16M1091447
25. Zhou M, Chen H, Paisley J, Ren L, Li L, Xing Z, et al. Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images. *IEEE Trans Image Process.* (2012) **21**:130–44. doi: 10.1109/TIP.2011.2160072
26. Chan C, Fulton R, Feng DD, Meikle S. Median non-local means filtering for low SNR image denoising: application to PET with anatomical knowledge. In: *IEEE Nuclear Science Symposium Conference Record.* (2010) p. 3613–8. doi: 10.1109/NSSMIC.2010.5874485
27. Manjón JV, Coupé P, Buades A, Louis Collins D, Robles M. New methods for MRI denoising based on sparseness and self-similarity. *Med Image Anal.* (2012) **16**:18–27. doi: 10.1016/j.media.2011.04.003
28. El Karoui N, Wu HT. Connection graph Laplacian methods can be made robust to noise. *Ann Stat.* (2016) **44**:346–72. doi: 10.1214/14-AOS1275
29. Eguchi T, Gilkey PB, Hanson AJ. *Gravitation, gauge theories and differential geometry.* *Phys Rep.* (1980) **66**:213–393.
30. Gallier J. *Notes On Group Actions Manifolds, Lie Groups and Lie Algebras* (2005). Available online at: <http://www.cis.upenn.edu/~cis610/lie1.pdf> (Accessed November 21 2016).
31. El Karoui N. On information plus noise kernel random matrices. *Ann Stat.* (2010) **38**:3191–216. doi: 10.1214/10-AOS801
32. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vis.* (2004) **60**:91–110. doi: 10.1023/B:VISI.0000029664.99615.94
33. Korman S, Reichman D, Tsur G, Avidan S. Fast-match: fast affine template matching. *Int J Comput Vis.* (2017) **121**:111–25. doi: 10.1007/s11263-016-0926-1
34. Wilson DL, Baddeley AJ, Owens RA. A new metric for grey-scale image comparison. *Int J Comput Vis.* (1997) **24**:5–17.
35. Villanic C. *Topics in Optimal Transportation.* Graduate Studies in Mathematics; American Mathematical Society (2003).
36. Zhang L, Zhang L, Mou X, Zhang D. FSIM: a feature similarity index for image quality assessment. *IEEE Trans Image Process.* (2011) **20**:2378–86. doi: 10.1109/TIP.2011.2109730
37. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* (2004) **13**:600–12. doi: 10.1109/TIP.2003.819861
38. Wu CH, Wang TD, Hsieh CH, Huang SH, Lin JW, Hsu SC, et al. Imaging cytometry of human leukocytes with third harmonic generation microscopy. *Sci Rep.* (2016) **6**:37210. doi: 10.1038/srep37210
39. Yang W, Sun C, Zhang L. A multi-manifold discriminant analysis method for image feature extraction. *Pattern Recogn.* (2011) **44**:1649–57. doi: 10.1016/j.patcog.2011.01.019
40. Wang X, Slavakis K, Lerman G. Riemannian multi-manifold modeling. *ArXiv e-prints* (2014).
41. Dabov K, Foi A, Katkovnik V. Image denoising by sparse 3D transformation-domain collaborative filtering. *IEEE Trans Image Process.* (2007) **16**:1–16. doi: 10.1109/TIP.2007.901238
42. Katkovnik V, Foi A, Egiazarian K, Astola J. From local kernel to nonlocal multiple-model image denoising. *Int J Comput Vis.* (2010) **86**:1–32. doi: 10.1007/s11263-009-0272-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Lin, Minasian, Qi and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.