



# K-Means for noise-insensitive multi-dimensional feature learning

Nicholas Pellegrino\*, Paul W. Fieguth, Parsin Haji Reza

PhotoMedicine Labs, Department of Systems Design Engineering, University of Waterloo, 200 University Avenue West, Waterloo, N2L 3G1, Ontario, Canada

## ARTICLE INFO

### Article history:

Received 12 August 2022

Revised 29 January 2023

Accepted 20 April 2023

Available online 24 April 2023

Edited by: Jiwen Lu

### MSC:

41A05

41A10

65D05

65D17

### Keywords:

Feature learning

Clustering

Photoacoustic remote sensing

## ABSTRACT

Many measurement modalities which perform imaging by probing an object pixel-by-pixel, such as via Photoacoustic Microscopy, produce a multi-dimensional feature (typically a time-domain signal) at each pixel. In principle, the many degrees of freedom in the time-domain signal would admit the possibility of significant multi-modal information being implicitly present, much more than a single scalar “brightness”, regarding the underlying targets being observed. However, the measured signal is neither a weighted-sum of basis functions (such as principal components) nor one of a set of prototypes (K-means), which has motivated the novel clustering method proposed here. Signals are clustered based on their shape, but not amplitude, via angular distance, and centroids are calculated as the direction of maximal intra-cluster variance, resulting in a clustering algorithm capable of learning centroids (signal shapes) that are related to the underlying, albeit unknown, target characteristics in a scalable and noise-robust manner.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Medical imaging broadly serves as a powerful diagnostic tool, often non-invasively giving medical care providers valuable information that cannot be obtained otherwise [1,2]. Many biomedical sensing techniques operate by capturing time-domain (TD) signals from which diagnostically relevant information can be inferred. A list of some of these modalities includes sonography / ultrasound imaging [3,4], echocardiography [5], electrocardiography [6–8], electromyography [9,10], phonocardiography [11,12], phonomyography [13], etc. In many cases, objects are scanned, pixel-by-pixel, to produce a TD signal at each pixel. Two such modalities are Time-Domain Optical Coherence Tomography (TD-OCT) [14,15] and Photoacoustic Microscopy (PAM) [16–18].

In principle, the many degrees of freedom available within TD signals admit the possibility of there existing significant multi-modal information related to the imaged target, far beyond a single scalar value used to represent pixel “brightness”. However, extracting the information from these signals is not necessarily straightforward. Indeed, the work presented in this paper is motivated by the TD signals of Photoacoustic Remote Sensing (PARS) Mi-

croscopy [19–23], a novel all-optical variation of PAM. The underlying physics lead to TD signals having shapes specific to tissue type, but where the signals from a given target may vary in amplitude, be inverted (negative amplitudes), and suffer from noise. What is required is a set of time-domain *features*, that adequately capture information from the underlying target that is present in the TD signals.

## 2. Background

This work is motivated by imaging modalities that scan, pixel-by-pixel, leading to a measured signal,  $s_j(t)$ , over time,  $t$ , at each pixel,  $j$ :

$$s_j(t) = \sum_{\forall i} \alpha_{i,j} f_i(t) + v_j(t), \quad (1)$$

for weights,  $\alpha_{i,j}$ , applied to features,  $f_i$ , for target type  $i$ , where measurement noise,  $v_j(t)$ , is additive. In principle,  $t$  is continuous, however in practice, the TD signals are sampled discretely. For the purpose of this work, the same notation is used for both the discrete- and continuous-time representations, however all numerical computations involving the TD signals clearly refer to the measured discrete-time representation.

Conventionally, in PAM and PARS [2,19–21,24–26], only a scalar amplitude is extracted from each TD signal. Recently, other methods have been developed to extract additional information related

Abbreviation: PARS, Photoacoustic remote sensing.

\* Corresponding author.

E-mail addresses: [npellegr@uwaterloo.ca](mailto:npellegr@uwaterloo.ca) (N. Pellegrino), [paul.fieguth@uwaterloo.ca](mailto:paul.fieguth@uwaterloo.ca) (P.W. Fieguth), [phajireza@uwaterloo.ca](mailto:phajireza@uwaterloo.ca) (P. Haji Reza).

to the frequency content of the signals, as a means of inferring information related to the imaged target [27–29]. This work takes a different approach by proposing an unsupervised (clustering) approach to learn time-domain features that relate to the underlying target.

In principle, such a feature inference would seem to have been solved. Principal Components Analysis (PCA) and its variations [30–34] are capable of extracting features (the principal components) from TD data; the principal components yield a representation,

$$s_j(t) = \sum_i \alpha_{j,i} b_i(t), \quad (2)$$

based on a weighted sum of basis elements  $b_i$ . The principal component basis elements are those minimizing the variance of the residual error, but not necessarily those which *individually* effectively represent most of the signals, as in Eq. 1. That is, the principal components are unknown weighted combinations of the desired features, and therefore do not individually necessarily act as meaningful features.

In contrast, clustering methods such as K-Means [35–38] and K-Medoids [39,40] do seek to produce meaningful features, but assume that *each* measured data-point (TD signal) is *one* of a set of prototypes, and has variation only as a consequence of measurement error and noise. This assumption does not hold in Eq. 1, as measured TD signals may be scaled or inverted versions of what would be a prototype relating to a specific target. Furthermore, as will be discussed in Section 3, some fraction of the signals will be influenced by more than one tissue type, and therefore represent a mixture of classes.

A final clustering issue arises when computing centroids. In K-Means, cluster centroids are calculated as the mean of the points associated to each cluster, and in K-Medoids, the most centrally located data point is chosen to be the centroid. In both cases, high amounts of background noise, if included in any cluster, would strongly influence the mean, steering it away from a truer representation of the non-noise portion of the cluster.

Addressing these constraints has prompted the development of the method proposed in this paper, capable of learning features that relate characteristic signal shapes to individual components of the target in a scalable and noise-robust manner.

### 3. Methods

We wish to cluster TD signals based on their signal *shape*, but not amplitude. Building on Eq. 1, a given pixel (and its corresponding TD signal) may be expressed in terms of characteristic signal shapes of one or more targets  $\{f_i\}$  and a residual term,  $r_j(t)$ ,

$$s_j(t) = \sum_{\forall i} \alpha_{i,j} f_i(t) + r_j(t), \quad (3)$$

but where *most* pixels come from only a single target, such that  $\{\alpha_{i,j}\}$  is sparse. The proposed method is based on K-Means, but varies in its definition of distance and its method for computing cluster centroids (the learned features). Our goal is to discover characteristic signals shapes, a set of  $K$  centroids,  $\mathcal{F} = \{f_i(t)\}$ ,  $i = 1, \dots, K$ , by creatively clustering the TD signals from a given image.

Note that TD signals are just vectors in space,  $\mathbb{R}^n$ , where the dimension,  $n$ , of the space is simply the number of discrete TD samples. Thus, the equivalences  $s(t) \equiv \vec{s}$ ,  $f(t) \equiv \vec{f}$ , etc. are understood. Because TD signals are treated as Cartesian vectors, the signal shape is then analogous to the vector *angle*.

TD signals associated with a given target may be arbitrarily scaled and be subject to noise, thus any proposed clustering algorithm must be tolerant to these effects. The assumptions for the proposed method are as follows:

1. Each underlying class (i.e., isolated component) is characterized by a single prototype signal.
2. Most pixels are a member of only one class. Relatively few pixels may be members of several classes (i.e., exhibit a response associated with a mixture of underlying components).
3. The noise level, both in background and in higher-amplitude signals, is significant.
4. A large fraction of the pixels (the background) may be a member of *no* class, thus containing only noise, and thus should *not* significantly influence or bias the centroid calculation.

#### 3.1. Distance metric

The clustering algorithm requires a distance metric that is:

1. Scale (amplitude) invariant, and
2. Polarity-agnostic (i.e., insensitive to signal inversions).

Consider an arbitrary TD signal,  $\vec{s} = m\vec{u}$ , for unit-vector  $\vec{u}$  that defines the characteristic signal. The *negative* of this signal,  $-\vec{s} = (-m)\vec{u}$ , shares the same direction,  $\vec{u}$ , and therefore has the exact same signal shape and thus is associated with the same underlying imaged target. Therefore, a polarity-agnostic distance metric must assign a distance of zero between  $\vec{s}$  and  $-\vec{s}$ .

For simplicity, and to achieve symmetry in the distance metric, all angles will be considered to be positive. The angle,  $\vartheta$ , between two vectors,  $\vec{v}_1$  and  $\vec{v}_2$ , is defined as

$$\vartheta := \angle(\vec{v}_1, \vec{v}_2) = \arccos\left(\frac{\langle \vec{v}_1, \vec{v}_2 \rangle}{\|\vec{v}_1\| \|\vec{v}_2\|}\right), \quad (4)$$

for  $\arccos(\cdot)$  defined on  $[-1, 1]$  and mapping to  $[0, \pi]$ . The proposed distance metric, satisfying polarity-agnosticism, is

$$d(\vec{v}_1, \vec{v}_2) = \sin(\vartheta), \quad (5)$$

for  $\sin(\cdot)$  defined on the interval  $[0, \pi]$ , constraining the range of  $d(\cdot)$  to  $[0, 1]$ . Note that the function composition  $\sin(\arccos(x)) = \sqrt{1 - x^2}$ , therefore the distance metric simplifies as

$$d(\vec{v}_1, \vec{v}_2) = \sqrt{1 - \left(\frac{\langle \vec{v}_1, \vec{v}_2 \rangle}{\|\vec{v}_1\| \|\vec{v}_2\|}\right)^2}. \quad (6)$$

#### 3.2. Calculation of cluster centroids

As was mentioned at the end of Section 2, an alternative method for computing cluster centroids is required. Although angle-based variations of K-Means do exist – such as Spherical K-Means [41–43] whereby data-points are projected onto the unit-hypersphere via normalization and distance is defined by cosine *dissimilarity* – centroids are still computed by taking the mean of all points within a given cluster, an approach which does not apply to the representation given in Eq. 3:

- We have a requirement for polarity-agnosticism, whereby antipodal TD signals are clustered together, and these signals will largely cancel (negate each other) if averaged.
- A large fraction of pixels are background (composed of zero-mean noise), having a random angle, and therefore are associated to clusters at random, leading to significant biases if included in the sample mean, and it is highly undesirable for background noise to dominate (or affect at all) the learned centroids.

Similarly, while angular distance metrics can be used with K-Medoids, centroids are selected to minimize intra-cluster distance and would be *strongly* influenced by large fractions of background noise (random angles). Instead, we desire a unit-vector pointing in

the direction of the non-noise portion of the given cluster to define the centroid. The method needs to be polarity-agnostic, however we do not actually know, a priori, whether both positive and negative examples are present. Given the set of points,  $\mathcal{S}_i$ , associated with cluster,  $i$ , we construct the union set,

$$\mathcal{S}_i^\pm = \mathcal{S}_i \cup (-\mathcal{S}_i), \quad (7)$$

made up of the cluster,  $\mathcal{S}_i$ , and its negated points,  $-\mathcal{S}_i$ . From  $\mathcal{S}_i^\pm$ , and because of the sparsity of the weights  $\alpha_{i,j}$ , the centroid can be found as the direction of greatest variance (the first principal component from the sample covariance of  $\mathcal{S}_i^\pm$ ), allowing higher amplitude signals (having a greater signal to noise ratio) to have greater influence, effectively eliminating the influence of background noise (those random-angle data-points near the origin).

### 3.3. Algorithm

With the distance metric and the method for calculating centroids now defined, the clustering algorithm, detailed in [Algorithm 1](#), follows fairly naturally from conventional K-Means. The calculation of cluster centroids is detailed in Centroid Update ([line 16](#)), and on [line 19](#) an SVD (Singular Value Decomposition) is used to extract the first principal component.

Similar to conventional K-Means, we need to define convergence criteria:

1. Sufficiently few data-points change clusters between iterations (algorithm parameter *MinMoves*), or
2. The difference in the mean residual is sufficiently small between iterations (algorithm parameter *MinDifference*).

Feature amplitude extraction is achieved by performing a change of basis on the TD signals. Given the set of feature vectors,  $\mathcal{F} = \{\vec{f}_i\}$ , resulting from [Algorithm 1](#), the TD signals can be expressed as a weighted sum of the feature vectors plus a residual term, as in [Eq. 3](#). Arranging the feature vectors in columns, forming a matrix of features,  $F = [\vec{f}_1 \mid \vec{f}_2 \mid \dots]$ , the TD signals can then be represented concisely as  $s_j(t) = F\vec{\alpha}_j + \vec{r}_j$ . To solve for the vector of feature weights / amplitudes,  $\vec{\alpha}$ , the pseudo-inverse [44] of  $F$  is used, thus,  $\vec{\alpha}_j = F^+ s_j(t)$ .

## 4. Results

Because ground truth is not available for PARS images of tissue, synthetic data are used in [Subsections 4.1](#) and [4.2](#) and annotated hyperspectral data in [Subsection 4.3](#) to quantitatively evaluate the performance of the proposed method. However, usage on PARS data is also demonstrated in [Subsection 4.4](#) to illustrate the proposed method's promise.

### 4.1. Visual comparison of methods on synthetic data

We begin by comparing the proposed method to standard Principal Components Analysis (PCA) and Angular-Distance K-Means on the basis of synthetic data.

Given that none of the methods being compared are sensitive to correlations between specific indices / samples of the TD signals, randomly generated synthetic data are used. Ground truth prototype signals are first generated, then scaled (including negatives) versions with added noise are used. Additionally, background noise (zero-signal with added noise) is included. To improve visual interpretability, temporal correlations are introduced to the prototypes using a first-order auto-regressive model [45]. In total, 300 points are generated for this analysis: 100 for each of two classes and 100 for background noise.

---

### Algorithm 1 Proposed Clustering Algorithm.

---

#### Input:

Set of TD signals,  $\mathcal{S} = \{s_j(t)\}$ , to be clustered.  
 Number of desired clusters,  $K$ .  
 Minimum allowable moves criterion, *MinMoves*.  
 Difference in mean residual criterion, *MinDifference*.

#### Output:

Set of cluster labels,  $\mathcal{L} = \{\ell\}$ , associated with each TD signal.

Set of cluster centroids,  $\mathcal{F} = \{\vec{f}_i\}$ , for  $i = 1, \dots, K$ .

#### Initialization:

Randomly select  $K$  data-points as initial centroids.

1: **for**  $i = 1, \dots, K$  **do**

2:      $\vec{f}_i \xleftarrow{\text{Random}} s(t) \in \mathcal{S}$

3: **end for**

   Set previous value of mean residual to 0.

4:  $\mu_r^{\text{prev}} \leftarrow 0$

5: **repeat**

   Set number of changed cluster labels to 0.

6:  $moves \leftarrow 0$

**Membership Update:** Finding nearest centroid to each point.

7: **for all**  $s_j(t) \in \mathcal{S}$  **do**

8:      $\ell_j \leftarrow \arg \min_{i \in \{1, \dots, K\}} \left\{ d(s_j(t), \vec{f}_i) \right\}$

      Increment *moves* if cluster membership changes.

9:     **if**  $\ell_j$  changed this iteration **then**

10:          $moves \leftarrow moves + 1$

11:     **end if**

12: **end for**

   Evaluate mean residual (objective).

13:  $\mu_r \leftarrow \frac{1}{\|\mathcal{S}\|} \sum_{s_j(t) \in \mathcal{S}} d(s_j(t), \vec{f}_{\ell_j})$

14:  $\Delta\mu_r \leftarrow \mu_r - \mu_r^{\text{prev}}$

15:  $\mu_r^{\text{prev}} \leftarrow \mu_r$

**Centroid Update:** Use data-points within clusters to update.

16: **for**  $i = 1, \dots, K$  **do**

   Get set of data-points within cluster.

17:      $\mathcal{S}_i \leftarrow \{s_j(t) \mid \ell_j = i\}$

      Take union of set with its negative.

18:      $\mathcal{S}_i^\pm \leftarrow \mathcal{S}_i \cup (-\mathcal{S}_i)$

      Compute first principal component via SVD. Assign to centroid.

19:      $\vec{f}_i \leftarrow \text{PC}_1(\mathcal{S}_i^\pm)$

      Normalize centroid to fall on unit-hypersphere.

20:      $\vec{f}_i \leftarrow \vec{f}_i / \|\vec{f}_i\|$

21: **end for**

22: **until**  $\Delta\mu_r \leq \text{MinDifference}$  **OR**  $moves \leq \text{MinMoves}$

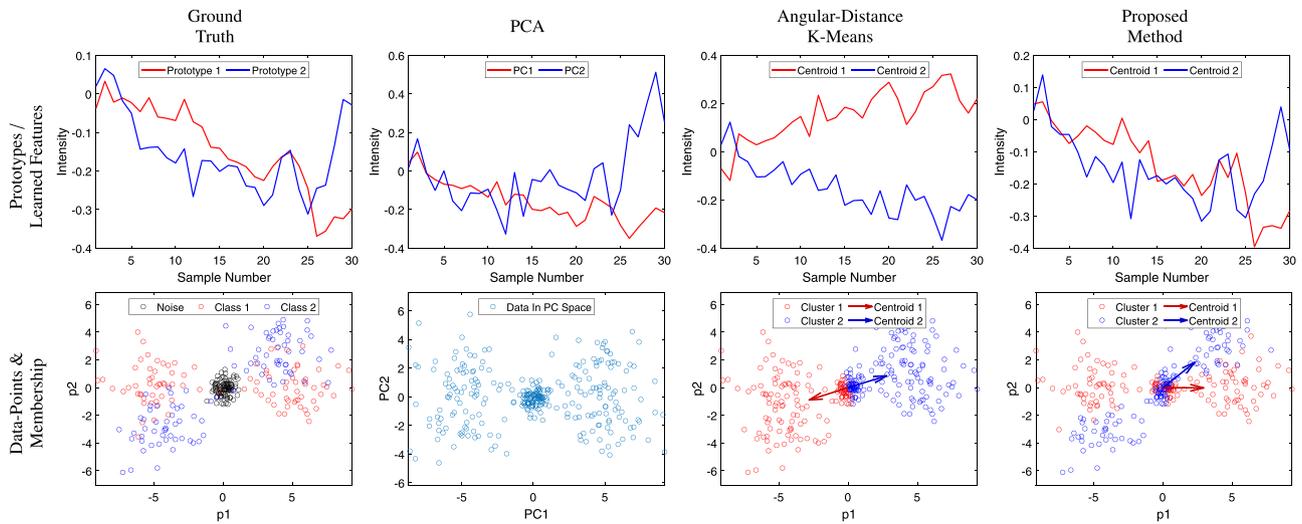
---

The synthetic data tests are illustrated in [Fig. 1](#). For visualization purposes in the bottom row, the data-points, which are 30-dimensional, are projected onto a 2D plane defined by the ground truth prototypes (except for the case where PCA is used, where the data-points are projected into the principal component space).

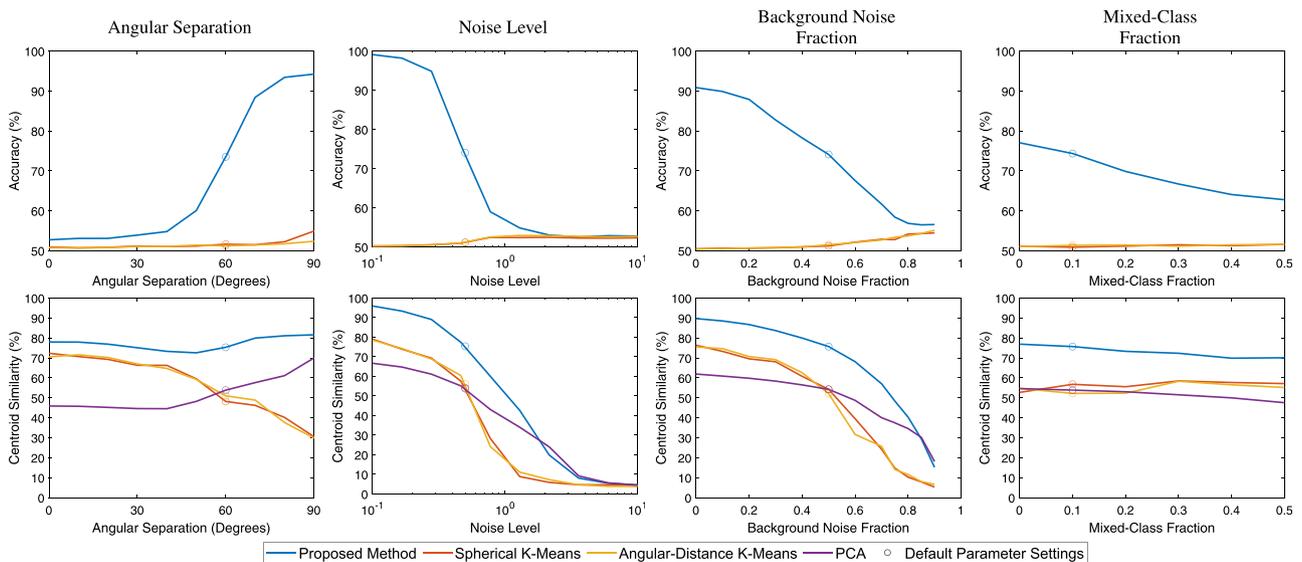
It can clearly be seen that neither PCA nor Angular-Distance K-Means produce learned features that even resemble the ground truth prototypes, since neither approach is able to produce the desired sparse-weighted features, whereas the proposed method is in fact highly effective at doing so.

### 4.2. Quantitative evaluation of proposed method

To more comprehensively test the proposed approach, the synthetic test of [Subsection 4.1](#) is generalized by evaluating performance as a function of key problem parameters, including the



**Fig. 1.** A comparison of methods to learn features from synthetic data. The first column shows the ground truth classes present in the data and the associated true cluster prototypes. Note that each data point, shown in the bottom row, is actually 30-dimensional but, for ease of viewing, is projected onto a plane defined by the true prototypes, with coordinates  $p_1$  (aligned with the true class 1 prototype) and  $p_2$  (orthogonal). In the PCA case, a projection into the principal component space is used instead. In addition to the two well-defined classes, there is low amplitude background noise present in the data set. The remaining columns demonstrate the ability of PCA, Angular-Distance K-Means, and the proposed method to recover the true clusters and centroids. In the case of PCA, no such clusters are identified; however, the principal components (learned features) can be compared to the true prototypes.



**Fig. 2.** For synthetically generated data in 1024-dimensions, the proposed method, Spherical K-Means, Angular-Distance K-Means, and PCA are used to cluster the data and learn features / centroids. In all cases,  $K = 2$  features are learned. By comparing to the ground truth, the clustering accuracy with respect to the true classes (top row) and centroid similarity to the true prototypes (bottom row) are evaluated for each method. Since PCA does not cluster points, its performance (purple line) is shown only in the bottom row, comparing the learned principal components and the true prototypes. In each column, the performance of the methods is evaluated while varying only one data parameter, holding the other parameters at their default values (circled). The results are averaged over 50 trials. In nearly every case, our proposed method outperforms the competing approaches.

nominal angular separation of the underlying classes, the noise level, and the fraction of points of low amplitude (background noise), and the fraction of points that are a mixture of multiple classes. Synthetic data (sets of 500 points) comprising two classes are generated in 1024-dimensions, based on ground truth class prototypes with a specified angular separation. To form a comparison, the proposed method, Spherical K-Means, Angular-Distance K-Means, and PCA are evaluated using this input data. Problem parameters are varied individually, holding all others fixed at a given nominal

value, and results are averaged over 50 trials to mitigate random sampling error.

Results are shown in Fig. 2. In all cases, two ( $K = 2$ ) features are learned. With PCA, points are not clustered; however, the principal components are learned. The task of selecting the number of clusters,  $K$ , is not addressed in this paper, and in principle the same issues apply here as in selecting  $K$  in regular K-Means or the dimensionality in PCA, and so  $K$  is assumed to be known based on external information.

**Table 1**

Tests performed on the hyperspectral data, summarized here as averages taken over five runs. Four tests are performed, as described in [Subsection 4.3](#). In each case, the proposed method is compared against PCA over a range of requested features specified by parameter  $K$ , where it is seen that results are generally strongest for  $K = 16$ , matching the number of ground truth classes.

Data Pre-processing	Evaluation Metric		# Learned Features $K$				
			2	4	8	12	16
None / Raw	SVM: Accuracy (%)	Proposed	17.8	19.0	24.8	28.9	29.0
		PCA	25.8	37.6	41.3	41.2	49.9
	One-hot: Accuracy (%)	Proposed	8.3	19.8	22.8	31.8	33.7
		PCA	2.2	6.2	6.2	6.2	6.2
	One-hot: Number of Unique Output Labels	Proposed	2.0	4.0	8.0	10.8	11.8
		PCA	2.0	4.0	6.4	6.6	6.6
Randomized Amplitudes	One-hot: Accuracy (%)	Proposed	13.0	18.1	26.7	31.5	34.8
		PCA	0.5	0.5	0.5	0.5	0.5
	One-hot: Number of Unique Output Labels	Proposed	2.0	4.0	7.8	10.2	11.2
		PCA	1.8	2.0	2.0	2.0	2.0
	Feature Similarity (%)	Proposed	99.0	98.8	98.8	98.9	98.9
		PCA	53.2	26.8	13.5	9.1	6.8

By comparing to the ground truth classes and prototypes, the clustering accuracy (top row) and centroid similarity (bottom row) are evaluated for each method. In each column, the performance of the methods is evaluated independently for each parameter. It can clearly be observed that the proposed method significantly outperforms the other three tested methods in nearly all cases. Further discussion is given in [Section 5](#).

#### 4.3. Application study – Hyperspectral data

The annotated *Indian Pines* [46] hyperspectral data set contains 16 ground truth classes and is used here to quantitatively evaluate and compare the performance of the proposed method to PCA. Note that while individual signals at each pixel are in the *spectral-domain* (i.e., across discrete wavelength bands) rather than the *time-domain*, this does not preclude the use of the analysis and processing techniques discussed here.

Firstly, we learn and extract features using the proposed method and PCA and evaluate the classification performance of a Support Vector Machine (SVM) [47,48] and a simpler ‘One-hot’ classifier. Because SVM (especially in the ‘one-vs-one’ ensemble scheme, used here) is capable of learning complex classification boundaries, it is relatively insensitive to whether the features provided to it are individually meaningful; rather, its performance depends mainly on the degree to which the provided features explain variation between ground truth classes. As such, we introduce a simple ‘One-hot’ classifier that associates each learned feature to a ground truth class, and classifies points based on whichever extracted feature amplitude is greatest (i.e., ‘hottest’). Note that the association is made based on whichever ground truth class is most relatively abundant for each feature, and multiple learned features can be associated to the same class label. The number of unique output labels indicates how many of the ground truth classes have associated features. This classification scheme has the advantage that its performance depends entirely on how well the learned features relate to ground truth classes.

Because the hyperspectral data are ‘one-sided’, (only positive), we repeat the ‘One-hot’ classification test but with randomized amplitudes, accomplished by first normalizing signals to unit length, and then multiplying by samples from a normal distribution. Lastly, because the proposed method is *not* designed for use as a dimensionality reducer / pre-processing step for a subsequent classifier (as is PCA), but is designed to learn independently meaningful features that relate to ground truth classes, we evaluate the *similarity* of the learned features to the mean signal of each ground truth class.

Note that each test is run across a range of requested numbers of features, from  $K = 2$  to 16 (matching the number of ground truth classes), and results are averaged over five trials to mitigate random error. In each trial, the data set is randomly split such that 20% is used for training and 80% for testing. The results are summarized in [Table 1](#).

#### 4.4. Application study – PARS data

We wish to apply the proposed method to real-world PARS microscopy data. Centroids (representative features) are learned and then are used to extract feature amplitudes from the 1024-dimensional TD signals of the PARS image.

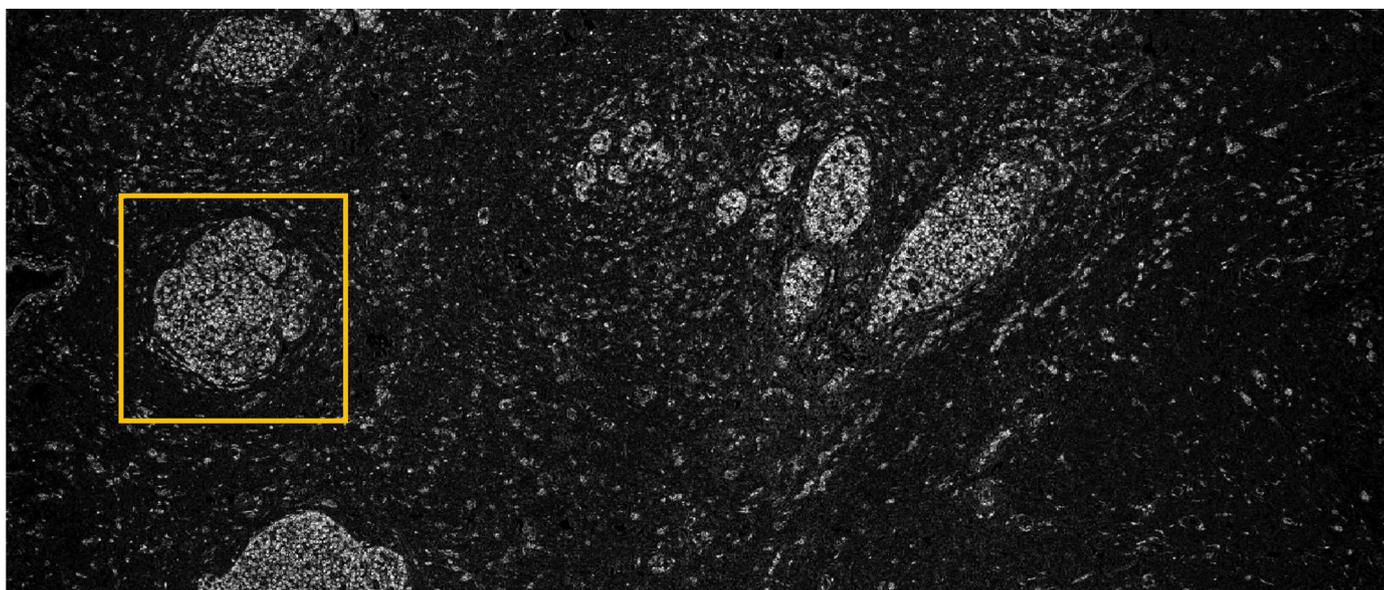
The use of the proposed algorithm for performing feature extraction is demonstrated here on an unstained, formalin-fixed paraffin-embedded (FFPE) human breast tissue slide. The standard projection PARS image of the slide is shown in [Fig. 3](#). This image was captured and provided by Benjamin R. Ecclestone with gratitude from the authors. A boxed-in region indicates the selection of TD signals that were used for feature learning, and the number of desired clusters was set to  $K = 6$ .

The extracted feature amplitudes (in absolute value, thus ignoring polarity) are combined as an RGB image in [Fig. 4](#): Feature 1 maps to red, feature 4 to green, and feature 5 to blue. The same is done for features extracted via Spherical and Angular-Distance K-Means (but with different feature-to-colour pairings), shown in [Fig. 4](#), and discussed in [Section 5](#).

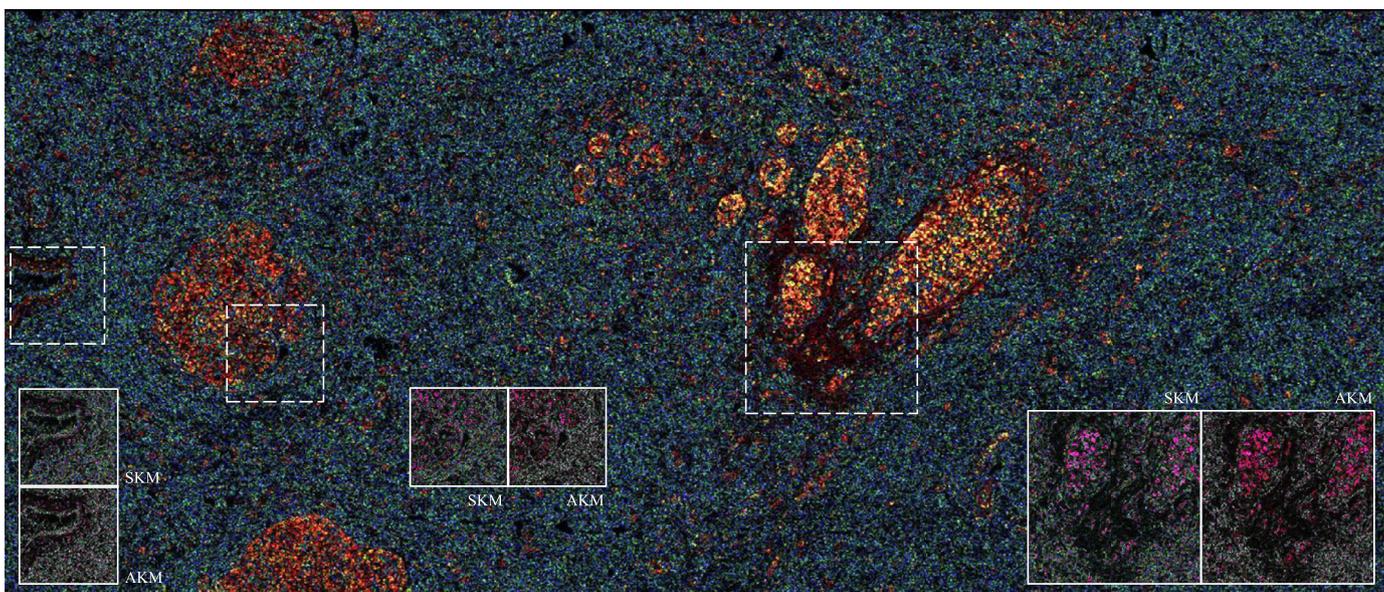
## 5. Discussion

The comparison between methods, shown in [Fig. 1](#), illustrates qualitatively that the learned features of PCA (the principal components) and Angular-Distance K-Means do not adequately match the ground truth. This can be explained by considering the underlying objective of each method:

- PCA seeks to learn a minimal basis (set of principal components) that preserves maximal variation in the data, but makes no assumption about the underlying basis nor the associated sparsity of its weights. The principal components could, in principle, be a linear combination of the true prototypes, however information of the true prototypes cannot directly be inferred.
- Regular K-Means, based on a Euclidean distance in the native-dimensional space, fails to assert the appropriate distance metric, and so is unable to accommodate variable signal amplitudes and would, in any event, most likely create a separate class for low-amplitude noise.



**Fig. 3.** A standard scalar signal projection of PARS imagery of unstained human breast tissue on a slide. The time-domain signals from the yellow boxed region are used for feature learning. Image captured by Benjamin R. Ecclestone.



**Fig. 4.** Absolute values of three feature amplitudes, extracted from the time domain signals underlying the imagery in Fig. 3 using the proposed method. The three features are combined to produce a colour image: Feature 1 maps to red, Feature 4 maps to green, Feature 5 maps to blue. Quite remarkable structures and tissue differentiation are clearly visible in this image, relative to that of Fig. 3. Insets of images produced via the same process (but with different feature-to-colour pairings) for features extracted using Spherical (SKM) and Angular-Distance (AKM) K-Means are included for comparison.

- Angular-Distance K-Means creates clusters such that minimal intra-cluster variance and maximal inter-cluster distance are achieved, as is intended; however, due to the bi-polar nature of the TD signals explored in this report, Angular-Distance K-Means cannot group antipodal components of the same class together. Even if the distance metric were changed to accommodate such bi-polar samples, K-Means still remains highly susceptible to the influence of background noise since its centroids are not weighted by signal amplitude.

The quantitative evaluations shown in Fig. 2 remains consistent with the preceding discussion, and broadly matches our expectations – performance decreases with higher noise levels or more

mixed-class points, and increases with greater separation angles. Notably, the proposed method is able to learn centroids well at low separation angles, even if the clustering accuracy is low, demonstrating its ability to learn and discern very similar underlying features. Generally, Spherical and Angular-Distance K-Means perform very similarly, and are outperformed in nearly every circumstance by our proposed approach.

Somewhat counter-intuitively, Spherical and Angular-Distance K-Means achieve greater centroid similarity at lower angular separation or greater fractions of mixed points, and improved clustering accuracy at greater fractions of background points. This is because the methods tend to learn two antipodal centroids, halfway between (equidistant to) the two ground truth centroids, as was

seen in Fig. 1. Thus at low angular separation, the learned centroids have a very high similarity to ground truth prototypes; however, clustering accuracy is minimal (~50%) since both classes are equidistant. High fractions of mixed-class points aid in learning equidistant centroids, and high fractions of background noise pull the learned centroids further away from lying directly between the ground truth ones, leading to a slight rise in clustering accuracy.

Results of the tests on the hyperspectral data are summarized in Table 1. On the raw hyperspectral data, PCA strongly outperforms the proposed method when used alongside the SVM classifier as a result of the PCA features explaining greater variation in the data. Note that the hyperspectral data are *highly correlated* such that the majority of ground truth classes are separated by 20° or less, and because of this, features extracted by the proposed method are also highly correlated and explain less variation than PCA features, leading to the poorer performance in the SVM test. However, with the simpler ‘One-hot’ classifier, the proposed method is able to consistently outperform PCA, indicating that the learned features are more meaningful in their correspondence to ground truth classes. The same conclusion is also confirmed by observing the number of unique output labels for each method. Up to  $K = 8$ , the number of unique output labels for the proposed algorithm is maximal (and continues to rise for larger  $K$ ) whereas for PCA, the number quickly reaches an upper limit of roughly 6 or 7.

On tests based on randomized signal amplitudes, PCA again suffers, resulting in only 0.5% classification accuracy in contrast to the proposed method which does not face a significant decline in performance, indicating, as designed, that the proposed method is polarity-agnostic. The number of unique output labels is minimally affected for the proposed method, however the PCA features are generally associated to only two classes. Lastly, the feature similarity scores indicate that the proposed method learns features that are highly similar to the ground truth data, whereas PCA does not (and cannot, given that all principal components must be orthogonal whereas the data are *highly correlated*).

Although ground truth is not available for the PARS imagery, the proposed method yielded biologically meaningful, spatially grouped, highly compelling results. The learned features correspond to specific tissue structures present in Fig. 4, and are distinctly separated in a way that is not at all the case in standard PARS imagery, as in Fig. 3. The highlighted details and ability to discern tissue structures in the inset images, produced from Spherical and Angular-Distance K-Means features, are far less striking than with the proposed method, indicating that the proposed method is able to learn features that more closely relate to underlying tissue components. A further detailed study involving commentary from histology experts would be necessary to validate the inferred features, and is the subject of ongoing research.

## 6. Conclusion

The method proposed in this paper is capable of learning features from high-dimensional signals that relate to individual components of the underlying data, in spite of signal amplitude variations, inversions, and noise. When tested on synthetic data, the proposed method learned features that closely match the ground truth prototypes, whereas the other compared methods could not.

The proposed method consistently performed as well or better than all other methods, across all four problem parameters, offering an attractive, intuitive, amplitude-flexible and class-mixing alternative to PCA or K-Means, due to its definition of distance, mixing weight sparsity, and strategy for computing centroids.

Tests on hyperspectral data resoundingly indicate that the proposed method learns features that correspond well to ground truth

classes, on the basis of having both high feature similarity to mean ground truth class signals, and high numbers of unique output labels in the ‘One-hot’ classifier.

When applied to real data from a PARS microscope, the proposed method yielded features that showed correspondence with tissue structures, showing significant promise. Clearly the proposed method is not in any way specific to PARS data, and is far more broadly applicable to the analysis of multi-dimensional / multi-modal signals.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Parsin Haji Reza reports a relationship with illumiSonics Inc. that includes: board membership, employment, and equity or stocks.

## Data availability

The authors are unable or have chosen not to specify which data has been used.

## Acknowledgment

The authors would like to thank Benjamin R. Ecclestone, a PhD student at PhotoMedicine Labs at the University of Waterloo, for collecting the PARS image analyzed here.

## References

- [1] B. Burbridge, E. Mah, Undergraduate Diagnostic Imaging Fundamentals, University of Saskatchewan, 2017. <https://openpress.usask.ca/undergradimaging>.
- [2] L.V. Wang, H.i. Wu, Biomedical Optics: Principles and Imaging, John Wiley & Sons, 2012.
- [3] P.L. Allan, G.M. Baxter, M.J. Weston, Clinical Ultrasound, 2-Volume Set E-Book: Expert Consult: Online and Print, Elsevier Health Sciences, 2011.
- [4] J.C. Hobbins, Obstetric Ultrasound: Artistry in Practice, John Wiley & Sons, 2008.
- [5] J. Laurenceau, M. Malergue, The Essentials in Echocardiography, Springer Science & Business Media, 2012, 4.
- [6] D.B. Foster, Twelve-Lead Electrocardiography: Theory and Interpretation, Springer Science & Business Media, 2007.
- [7] K. Wang, Atlas of Electrocardiography, JP Medical Ltd, 2012.
- [8] S. Chugh, Textbook of Clinical Electrocardiography, JAYPEE BROTHERS PUBLISHERS, 2014.
- [9] A.A. Leis, M.P. Schenk, Atlas of Nerve Conduction Studies and Electromyography, Oxford University Press, 2013.
- [10] G. Kamen, D.A. Gabriel, Essentials of Electromyography, Human Kinetics Publishers, 2009.
- [11] M.E. Tavel, Clinical phonocardiography and external pulse recording, in: Clinical phonocardiography and external pulse recording, 1972, p. 322.
- [12] A. Leatham, Auscultation of the Heart and Phonocardiography, Churchill London, 1970.
- [13] T. Fuchs-Buder, Neuromuscular Monitoring, Springer Science & Business Media, 2011.
- [14] U. Haberland, P. Jansen, V. Blazek, H.J. Schmitt, Optical Coherence Tomography of Scattering Media Using Frequency-Modulated Continuous-Wave Techniques with Tunable Near-Infrared Laser, Coherence Domain Optical Methods in Biomedical Science and Clinical Applications, International Society for Optics and Photonics, 1997, 2981, 20–28.
- [15] Y. Watanabe, F. Sajima, T. Itagaki, K. Watanabe, Y. Shuto, High-speed linear detection time domain optical coherence tomography with reflective grating-generated spatial reference delay, Appl. Opt. 48 (2009) 3401–3406.
- [16] P. Beard, Biomedical photoacoustic imaging, Interface Focus 1 (2011) 602–631.
- [17] G. Li, K.I. Maslov, L.V. Wang, Reflection-mode multifocal optical-resolution photoacoustic microscopy, J. Biomed. Opt. 18 (2013) 030501.
- [18] J. Yao, L.V. Wang, Photoacoustic microscopy, Laser Photon. Rev. 7 (2013) 758–778.
- [19] P. Hajireza, W. Shi, K. Bell, R.J. Paproski, R.J. Zemp, Non-interferometric photoacoustic remote sensing microscopy, Light: Sci. Appl. 6 (2017) e16278.
- [20] S. Abbasi, M. Le, B. Sonier, D. Dinakaran, G. Bigras, K. Bell, J.R. Mackey, P.H. Reza, All-optical reflection-mode microscopic histology of unstained human tissues, Sci. Rep. 9 (2019) 1–11.

- [21] S. Abbasi, K. Bell, P.H. Reza, Rapid high-resolution mosaic acquisition for photoacoustic remote sensing, *Sensors* 20 (2020) 1027.
- [22] B.R. Ecclestone, K. Bell, S. Abbasi, D. Dinakaran, F.K. van Landeghem, J.R. Mackey, P. Fieguth, P.H. Reza, Improving maximal safe brain tumor resection with photoacoustic remote sensing microscopy, *Sci. Rep.* 10 (2020) 1–7.
- [23] B.R. Ecclestone, K. Bell, S. Abbasi, D. Dinakaran, M. Taher, J.R. Mackey, P.H. Reza, Histopathology for mohs micrographic surgery with photoacoustic remote sensing microscopy, *Biomed. Opt. Express* 12 (2021) 654–665.
- [24] W. Bost, F. Stracke, M. Fournelle, R. Lemor, Developing a High-Resolution Photoacoustic Microscopy Platform, in: 4th European Conference of the International Federation for Medical and Biological Engineering, Springer, 2009. 448–451
- [25] G. Ku, K.I. Maslov, L. Li, L.V. Wang, Photoacoustic microscopy with 2- $\mu$ m transverse resolution, *J. Biomed. Opt.* 15 (2010) 021302.
- [26] B. Rao, L. Li, K. Maslov, L. Wang, Hybrid-scanning optical-resolution photoacoustic microscopy for in vivo vasculature imaging, *Opt. Lett.* 35 (2010) 1521–1523.
- [27] M.J. Moore, E. Hysi, M.N. Fadhel, S. El-Rass, Y. Xiao, X.-Y. Wen, M.C. Kolios, Photoacoustic f-mode imaging for scale specific contrast in biological systems, *Commun. Phys.* 2 (2019) 1–10.
- [28] P. Kedariseti, N.J. Haven, B.S. Restall, M.T. Martell, R.J. Zemp, Label-Free Virtual h&e Histopathology Using F-Mode Ultraviolet Photoacoustic Remote Sensing, *Photons Plus Ultrasound: Imaging and Sensing 2021*, International Society for Optics and Photonics, 2021. 11642, 116422S
- [29] P. Kedariseti, B.S. Restall, N.J. Haven, M.T. Martell, B.D. Cikaluk, J. Deschenes, R.J. Zemp, F-Mode ultraviolet photoacoustic remote sensing for label-free virtual h&e histopathology using a single excitation wavelength, *Opt. Lett.* 46 (2021) 3500–3503.
- [30] K. Pearson, Liii. on lines and planes of closest fit to systems of points in space, *London Edinbur. Dublin Philos. Mag. J. Sci.* 2 (1901) 559–572.
- [31] H. Hotelling, Analysis of a complex of statistical variables into principal components, *J. Educ. Psychol.* 24 (1933) 417.
- [32] R. Bro, A.K. Smilde, Principal component analysis, *Anal. Methods* 6 (2014) 2812–2831.
- [33] J. Shlens, A tutorial on principal component analysis, *arXiv preprint arXiv:1404.1100* (2014).
- [34] B.M.S. Hasan, A.M. Abdulazeez, A review of principal component analysis algorithm for dimensionality reduction, *J. Soft Comput. Data Mining* 2 (2021) 20–30.
- [35] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA, 1967. 1, 281–297
- [36] H. Steinhaus, et al., Sur la division des corps matériels en parties, *Bull. Acad. Polon. Sci.* 1 (1956) 801.
- [37] J.A. Hartigan, M.A. Wong, Algorithm as 136: a k-means clustering algorithm, *J. R. Stat. Soc. Ser. c (Appl. Stat.)* 28 (1979) 100–108.
- [38] S. Lloyd, Least squares quantization in pcm, *IEEE Trans. Inf. Theory* 28 (1982) 129–137.
- [39] L. Kaufman, P.J. Rousseeuw, Partitioning around medoids (program pam), *Find. Group Data: Introd. Cluster Anal.* 344 (1990) 68–125.
- [40] E. Schubert, P.J. Rousseeuw, Fast and eager k-medoids clustering: o(k) runtime improvement of the pam, clara, and clarans algorithms, *Inf. Syst.* 101 (2021) 101804.
- [41] I.S. Dhillon, D.S. Modha, Concept decompositions for large sparse text data using clustering, *Mach. Learn.* 42 (2001) 143–175.
- [42] V. Nguyen, Gene Clustering on the Unit Hypersphere with the Spherical k-Means Algorithm: Coping with Extremely Large Number of Local Optima, *CSREA Press, 2008. World Congress in Computer Science, Computer Engineering, and Applied Computing*, 226–233.
- [43] K. Hornik, I. Feinerer, M. Kober, C. Buchta, Spherical k-means clustering, *J. Stat. Softw.* 50 (2012) 1–22.
- [44] P. Fieguth, *Statistical Image Processing and Multidimensional Modeling*, Springer Science & Business Media, 2010.
- [45] G. Kirchgässner, J. Wolters, U. Hassler, *Introduction to Modern Time Series Analysis*, Springer Science & Business Media, 2012.
- [46] M.F. Baumgardner, L.L. Biehl, D.A. Landgrebe, 220 Band aviris hyperspectral image data set: june 12, 1992 indian pine test site 3, *Purdue Univ. Res. Reposit.* 10 (2015) 991.
- [47] C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* 20 (1995) 273–297.
- [48] W.S. Noble, What is a support vector machine? *Nat. Biotechnol.* 24 (2006) 1565–1567.