

# Adaptive Two-Dimensional Embedded Image Clustering

Zhihui Li,<sup>1,2</sup> Lina Yao,<sup>2</sup> Sen Wang,<sup>3</sup> Salil Kanhere,<sup>2</sup> Xue Li,<sup>3</sup> Huaxiang Zhang<sup>1</sup>

<sup>1</sup>School of Information Science and Engineering, Shandong Normal University

<sup>2</sup>School of Computer Science and Engineering, University of New South Wales

<sup>3</sup>School of Information Technology and Electric Engineering, The University of Queensland  
zhihuilics@gmail.com, {lina.yao, salil.kanhere}@unsw.edu.au, sen.wang@uq.edu.au, xueli@itee.uq.edu.au,  
huaxzhang@163.com

## Abstract

With the rapid development of mobile devices, people are generating huge volumes of images data every day for sharing on social media, which draws much research attention to understanding the contents of images. Image clustering plays an important role in image understanding systems. Often, most of the existing image clustering algorithms flatten digital images that are originally represented by matrices into 1D vectors as the image representation for the subsequent learning. The drawbacks of vector-based algorithms include limited consideration of spatial relationship between pixels and computational complexity, both of which blame to the simple vectorized representation. To overcome the drawbacks, we propose a novel image clustering framework that can work directly on matrices of images instead of flattened vectors. Specifically, the proposed algorithm simultaneously learn the clustering results and preserve the original correlation information within the image matrix. To solve the challenging objective function, we propose a fast iterative solution. Extensive experiments have been conducted on various benchmark datasets. The experimental results confirm the superiority of the proposed algorithm.

## Introduction

Clustering has attracted increasingly interests in the fields of computer vision and machine learning. Its goal is to partition data points into disjoint groups such that the data points in the same group close to each other and those in different groups far apart (Jain and Dubes 1988; Chang et al. 2015; Abassi and Boukhris 2019). To effectively index, retrieve and organize the explosively increasing multimedia data, researchers have employed many clustering algorithms for image clustering (Datta et al. 2008; Xia et al. 2016; Caron et al. 2018).

To improve the performance of automatic image annotation systems, image clustering algorithms have been adopted in various ways (Wang et al. 2018; Peng et al. 2018). For example, Li *et al.* propose to first cluster personal album images represented by bags of weighted vectors, learn the conceptual categories for the clusters using a web-based annotation approach, and then obtain the conceptual categories

of the whole dataset in a semi-supervised learning fashion. Researchers have also successfully applied image clustering algorithms in content based image retrieval (CBIR) systems. In (Gordon, Greenspan, and Goldberger 2003), Gordon *et al.* utilize the traditional clustering algorithm,  $k$ -means, as a preprocessing approach. To represent the diverse views of the landmarks, Kennedy *et al.* employ the clustering algorithms to select the most discriminative images (Kennedy and Naaman 2008).

Most of the existing clustering algorithms, however, only work well when the sample's dimensionality is low. Due to the curse of dimensionality, the clustering performance of these algorithms is not guaranteed. For example, classical  $k$ -means clustering algorithm iteratively assigns each sample to the cluster with the closest centroid based well-defined metric learning method and updates the centroid of the data points in each cluster. However, it is challenging for the metric learning method to be accurate, which tends to deteriorate subsequent clustering performance. Inspired by the fact that high-dimensional data may exhibit dense grouping in a low-dimensional subspace (Chang et al. 2016), researchers seek to project the high-dimensional data into a low-dimensional subspace using some subspace learning algorithms, and then perform clustering with the learned low-dimensional representation. An intuitive way is to employ principal components analysis (PCA) for dimension reduction, and classical  $k$ -means for clustering in the lower-dimensional space (Xu and II 2005). Some researchers, however, claim that linear discriminant analysis (LDA) is more suitable than PCA for clustering because LDA is capable of encoding discriminant information (Ding and Li 2007). They show that better performance is obtained when integrating  $k$ -means and LDA into a single framework.

Another direction of clustering is to consider the manifold structure of data. The representative algorithms are spectral clustering (SC) algorithm, normalized cut (NCut), and  $k$ -way NCut (Yu and Shi 2003), which have demonstrated their superiority in terms of image segmentation and other real-world applications. These clustering algorithms rely on the graph construction, which is computed with the similarities among the data points. Therefore, these algorithms are sensitive to the bandwidth parameter. To solve this prob-

lem, researchers have proposed different variants of spectral clustering algorithms. For example, Wu *et al.* propose to learn a Laplacian matrix using local learning based clustering (LLC), which is built on the assumption that the cluster label of each data point is close to the one predicted by the local regression model. Self-tuning spectral clustering (Zelnik-Manor and Perona 2004) is capable of learning the best parameters automatically in an unsupervised setting.

Although most existing clustering algorithms have achieved promising performance, they require that an image be represented by a vector. The researchers usually concatenate each row/column of an image matrix, resulting in a vectorized representation. This process has some inherent limitations. First, it erases the spatial relationship of the image. Second, the computational burden increases dramatically since the dimensionality increases when the matrix representation is transformed into the vector representation. An intuitive way to solve this problem is to work with data in a matrix representation, which is able to preserve the original spatial information and reduce the computational complexity. Although much progress has been made on two-dimensional dimensionality reduction (Chang et al. 2016; Hou et al. 2017; Ma et al. 2013), few attention has been paid to two-dimensional clustering (Luo, Huang, and Ding 2011).

Inspired by the progress of clustering analysis and two-dimensional learning, we propose a novel clustering algorithm for image matrices. The proposed method is able to work directly on matrices of images instead of flattened vectors. It simultaneously learns the clustering results and preserves the original structure information within the image matrix. We name the new method Adaptive Two-Dimensional Embedded Image Clustering (A2DEIC). The main contributions of this work are summarized as follows

- To the best of our knowledge, we proposed the first clustering algorithm that can directly deal with matrix representations. In this way, we can preserve the spatial correlations within the original data.
- Since the objective function is non-smooth and difficult to solve, we propose a fast iterative solution to optimize it. The experimental results confirm that the objective function generally converges within 10 steps.
- We conduct extensive experiments on several benchmark datasets, and confirmed the superiority of the proposed algorithm, which demonstrates the benefits of two-dimensional embedding for improving the performance of image clustering.

## Related Work

In this section, we briefly review related works on vector-based image clustering approaches and two-dimensional feature analysis.

### Vector-based Image Clustering

Currently, vector-based clustering methods have been applied in many Content Based Image Retrieval (CBIR) applications to improve performance. Conventionally, 2D images are flattened into 1D vectors and packed together to

form a large matrix followed by learning procedure. Because of simplicity, k-means has been pervasively utilized for image clustering by the computer vision community. For images represented by 1D vectors, K-means iteratively calculates distances between image vectors and all the centroids, and assigns the closest cluster label to each image, followed by updating the new centroids for each cluster. However, in such a big data era, high-definition images have larger vector representations, which results in high computational cost. The performance downturn of k-means clustering algorithm and its variants that are performed in the original feature space are inevitable. To tackle this problem, one line of research introduce subspace learning that transforms the original features onto a lower dimensional subspace to exploit distinctive features without sacrificing the effectiveness and efficiency of learning performance. One typical solution is to apply principal component analysis (PCA) converting high dimensional data into lower dimensional data and then perform k-means clustering in the lower dimensional subspace. Another representative algorithms that learn a low dimensional subspace in a linear manner include linear discriminant analysis (LDA) (Belhumeur, Hespanha, and Kriegman 1997), have been studied in (Ding and Li 2007; Ye, Zhao, and Wu 2008) and proved be superior to a combination of PCA and k-means regarding data clustering in the learned subspace. To enhance of robustness of aforementioned subspace learning algorithms, non-Euclidian measures, such as  $\ell_1$ -norm and its generalized version,  $\ell_p$ -norm, have been well studied in (Zhong and Zhang 2013; Zhong, Zhang, and Li 2014; Lu, Zou, and Wang 2016; Kwak 2014; Oh and Kwak 2013) to deal with outliers effectively. Although the conventional algorithms are widely used for image clustering, the nature of two dimension within images can not be preserved because of the vector-based image representation.

### Two-Dimensional Feature Analysis

To capture the 2D nature of images, 2D-based subspace methods have aroused broad interests in the fields of pattern recognition and machine learning. Yang et al. (Yang et al. 2004) propose 2DPCA performing feature analysis on a 2D image matrix rather than a flattened 1D image vector, preserving the topology structures between pixels. As 2DPCA can only take advantage of the variations between rows, which ignores important information between columns, bilateral projection-based 2DPCA (B2DPCA) (Kong et al. 2005; Zhang and Zhou 2005; Yang and Liu 2007; Wang et al. 2017), 2DSVD (Ding and Ye 2005) and its high order version (HOSVD) (Luo, Huang, and Ding 2011) have been developed to analyze features both in rows and columns. (Zhang et al. 2015) have pointed that the squared F-norm based 2DPCA and related variants are not robust to outliers. To handle this problem, L1-norm has been studied in (Li, Pang, and Yuan 2010; Wang et al. 2015) as the distance metric in the criterion function, as well as the nuclear norm (N-2DPCA and N-B2DPCA) (Zhang et al. 2015). Comparing to the family of two-dimensional PCA feature analysis methods, another type of representative two-dimensional feature analysis algorithms, 2DLDA (Ye, Janardan, and Li

2005) and its variants (Yan et al. 2007; Li et al. 2008; Luo, Ding, and Huang 2009; Chang et al. 2016) aim to maximize the ratio of between-class to within-class scatter matrices as a matrix-based discriminant criterion. Inspired by the  $\ell_1$ -norm 2DPCA technique, L1-norm based 2DLDA algorithms (Chen, Yang, and Jin 2014; Li, Shao, and Deng 2015) have been developed for two-dimensional feature analysis in the sense that measurement of outliers can be obviously detrimental to the objective function with the squared F-norm metric.

## Clustering Image Matrices

In this section, we first describe the details of the proposed algorithm, followed by an efficient iterative approach to solve the objective function.

### Problem Formulation

Our work is built on the intuition that an image's 2D structure should be better represented by a matrix form (Chang et al. 2016; Hou et al. 2017; Ma et al. 2013). Inspired by recent progress on two-dimensional feature learning, we propose a two-dimensional embedded clustering algorithm for image matrices.

To facilitate the later presentation, we first give the notations that will be used in this work. Given an image dataset  $\mathcal{X} = \{X_1, X_2, \dots, X_n\}$ ,  $X_i \in \mathbb{R}^{w \times h}$  ( $1 \leq i \leq n$ ) is the  $i$ -th image,  $w$  and  $h$  are the width and height of an image, and  $n$  is the total number of images in this dataset. The goal of image clustering is to partition  $\mathcal{X}$  into  $c$  clusters. Let us define the cluster assignment matrix by  $Y = \{y_1, y_2, \dots, y_n\} \in \mathbb{R}^{n \times c}$ , where  $y_i \in \{0, 1\}^{c \times 1}$  ( $1 \leq i \leq n$ ) denotes the cluster indicator vector of the image  $X_i$ . The  $j$ -th element of  $y_i$  is 1 if the image  $X_i$  is partitioned into the  $j$ -th cluster, and 0 otherwise.

Given the image dataset  $\mathcal{X}$ , two-dimensional clustering algorithm minimizes the following objective function:

$$\begin{aligned} \min_{U, V, C} \sum_{i=1}^n \sum_{j=1}^c \|U^T X_i V - C_j\|_F^2, \quad (1) \\ \text{s.t. } U^T U = I, V^T V = I \end{aligned}$$

where  $U \in \mathbb{R}^{w \times w_1}$  and  $V \in \mathbb{R}^{h \times h_1}$  are projection matrices which map the original data points into a lower dimensional subspace  $\mathbb{R}^{w_1 \times h_1}$ .  $C_j$  is the centroid matrix of the  $j$ -th cluster in the lower-dimensional subspace.

By incorporating the cluster assignment matrix  $Y$ , the objective function arrives at:

$$\begin{aligned} \min_{U, V, C, Y} \sum_{i=1}^n \sum_{j=1}^c y_{ij} \|U^T X_i V - C_j\|_F^2, \quad (2) \\ \text{s.t. } U^T U = I, V^T V = I, Y \in \text{Ind} \end{aligned}$$

where  $Y$  is used to weight the overall loss.

Meanwhile, we maximize the covariance in the projected space by the following objective function:

$$\begin{aligned} \max_{U, V} \sum_{i=1}^n \|U^T (X_i - \bar{X}) V\|_F^2 \quad (3) \\ \text{s.t. } U^T U = I, V^T V = I \end{aligned}$$

Combining Equation (2) and Equation (3), the objective function arrives at:

$$\begin{aligned} \min_{U, V, Y, C} \sum_{i=1}^n \sum_{j=1}^c y_{ij} \|U^T X_i V - C_j\|_F^2 \\ - \lambda \sum_{i=1}^n \|U^T (X_i - \bar{X}) V\|_F^2 \quad (4) \\ \text{s.t. } U^T U = I, V^T V = I, Y \in \text{Ind} \end{aligned}$$

where  $\lambda$  is a parameter to balance the two terms.

### Optimization

In this section, we propose an iterative approach to optimize the proposed objective function in Equation (4).

#### (1) Fixing $Y$ , and optimizing $U$ , $V$ and $C$ :

Setting the derivative of the objective function *w.r.t.*  $C$  to zero, we get:

$$C_j = \frac{\sum_{i=1}^n y_{ij} U^T X_i V}{\sum_{i=1}^n y_{ij}} \quad (5)$$

By incorporating Equation (5) into the original objective function, we arrive at:

$$\begin{aligned} \min_{U, V} \sum_{i=1}^n \sum_{j=1}^c y_{ij} \|U^T (X_i - \bar{X}_j) V\|_F^2 \\ - \lambda \sum_{i=1}^n \|U^T (X_i - \bar{X}) V\|_F^2 \quad (6) \\ \text{s.t. } U^T U = I, V^T V = I \end{aligned}$$

where

$$\bar{X}_j = \frac{\sum_{i=1}^n y_{ij} X_i}{\sum_{i=1}^n y_{ij}} \quad (7)$$

We rewrite the objective function as follows:

$$\min_{U, V} \|U^T M V\|_F^2, \quad \text{s.t. } U^T U = I, V^T V = I \quad (8)$$

where

$$M = \sum_{i=1}^n \sum_{j=1}^c y_{ij} (X_i - \bar{X}_j) - \lambda \sum_{i=1}^n (X_i - \bar{X}) \quad (9)$$

By fixing  $U$ , the objective function becomes:

$$\min_V \text{tr}(V^T M^T U U^T M V), \quad \text{s.t. } V^T V = I, \quad (10)$$

---

**Algorithm 1:** Optimization Algorithm for A2DEIC

---

- 1 Input:  $X_i \in \mathbb{R}, \lambda$
  - 2 Output:  $U, V, Y, C$
  - 3 Compute the image mean  $\bar{X}$
  - 4 Initialize the indicator matrix  $Y \in \mathbb{R}^{n \times n}$
  - 5 **repeat**
  - 6     Compute  $\bar{X}_j$  according to Equation (7)
  - 7     Compute  $C_j$  according to Equation (5)
  - 8     Compute  $M$  according to Equation (9)
  - 9     Compute  $V$  by eigen-decomposition of  $M^T U U^T M$
  - 10    Compute  $U$  by eigen-decomposition of  $M V V^T M^T$
  - 11    Compute  $Y$  by solving Equation (12)
  - 12 **until** convergence;
- 

where  $tr(\cdot)$  denotes the trace operator. The objective function of Equation (10) can be solved by eigen-decomposition of  $M^T U U^T M$ .

Similarly, by fixing  $V$ , the objective function becomes:

$$\min_U tr(U^T M V V^T M^T U), \quad s.t. U^T U = I \quad (11)$$

The objective function of Equation (11) can be solved by eigen-decomposition of  $M V V^T M^T$ .

**(2) Fixing U, V, C, and optimizing Y:** The objective function arrives at:

$$\min_{Y \in \text{Ind}} \sum_{i=1}^n \sum_{j=1}^c y_{ij} \|U^T X_i V - C_j\|_F^2 \quad (12)$$

This equation can be easily solved and result in the optimal  $Y$ .

The optimization of  $C, U, V$ , and  $Y$  is iterated until convergence. We summarize the detailed iteration process in Algorithm 1.

### Convergence Analysis

In this subsection, we prove the convergence of Algorithm 1. By the following theorem, we can verify that the proposed iterative approach in Algorithm 1 converges and the optimal solution of  $U, V, C$ , and  $Y$  are achieved.

**Theorem 1.** *The value of the proposed objective function in Equation (4) monotonically decreases in each iteration until convergence using the iterative approach in Algorithm 1.*

**Lemma 1.** *By fixing  $Y$ , we get the optimal solutions for  $U, V$  and  $C$ . Similarly, by fixing  $U, V$ , and  $C$ , we get the optimal solutions for  $Y$ .*

*Proof.* By fixing  $Y$ , we have obtained the optimal  $C$  by setting the derivative of the objective function w.r.t.  $C$  to zero. Then the optimal  $U$  and  $V$  are obtained by eigen-decomposition, respectively. Finally, the optimal  $Y$  is obtained by solving the objective function Equation (12).  $\square$

Next, we prove Theorem 1 as follows:

*Proof.* Suppose after the  $t$ -th iteration, we get  $Y^{(t)}, U^{(t)}, V^{(t)}$ , and  $C^{(t)}$ . In the next iteration, we fix  $Y$  as  $Y^{(t)}$ , and solve for  $U^{(t+1)}, V^{(t+1)}$ , and  $C^{(t+1)}$ . According to Lemma 1, we obtain:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^c y_{ij}^{(t)} \|U^{(t+1)T} X_i V^{(t+1)} - C_j^{(t+1)}\|_F^2 \\ & \quad - \lambda \sum_{i=1}^n \|U^{(t+1)T} (X_i - \bar{X}) V^{(t+1)}\|_F^2 \\ & \leq \sum_{i=1}^n \sum_{j=1}^c y_{ij}^{(t)} \|U^{(t)T} X_i V^{(t)} - C_j^{(t)}\|_F^2 \\ & \quad - \lambda \sum_{i=1}^n \|U^{(t)T} (X_i - \bar{X}) V^{(t)}\|_F^2 \end{aligned} \quad (13)$$

Similarly, when we fix  $U$  as  $U^{(t)}$ ,  $V$  as  $V^{(t)}$  and  $C$  as  $C^{(t)}$ , the following inequality holds:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^c y_{ij}^{(t+1)} \|U^{(t)T} X_i V^{(t)} - C_j^{(t)}\|_F^2 \\ & \quad - \lambda \sum_{i=1}^n \|U^{(t+1)T} (X_i - \bar{X}) V^{(t+1)}\|_F^2 \\ & \leq \sum_{i=1}^n \sum_{j=1}^c y_{ij}^{(t)} \|U^{(t)T} X_i V^{(t)} - C_j^{(t)}\|_F^2 \\ & \quad - \lambda \sum_{i=1}^n \|U^{(t)T} (X_i - \bar{X}) V^{(t)}\|_F^2 \end{aligned} \quad (14)$$

By integrating Equation (13) and Equation (14), we arrive at:

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=1}^c y_{ij}^{(t+1)} \|U^{(t+1)T} X_i V^{(t+1)} - C_j^{(t+1)}\|_F^2 \\ & \quad - \lambda \sum_{i=1}^n \|U^{(t+1)T} (X_i - \bar{X}) V^{(t+1)}\|_F^2 \\ & \leq \sum_{i=1}^n \sum_{j=1}^c y_{ij}^{(t)} \|U^{(t)T} X_i V^{(t)} - C_j^{(t)}\|_F^2 \\ & \quad - \lambda \sum_{i=1}^n \|U^{(t)T} (X_i - \bar{X}) V^{(t)}\|_F^2 \end{aligned} \quad (15)$$

Equation (15) demonstrates that after each iteration, the value of the proposed objective function decreases. Thus, Theorem 1 has been proved.  $\square$

## Experiment

In this section, we perform extensive experiments on several widely used benchmark datasets to confirm the superiority of the proposed algorithm.

Table 1: We evaluate the performance of all the compared algorithms in terms of clustering accuracy (ACC±Standard Deviation). Performance is reported in percentages.

	ACC					
	UUIIm	CVL	Pointing'04 (tilt)	Pointing'04 (pan)	USPS	Coil20
<i>k</i> -means	84.49±1.46	73.64±2.16	41.85±1.85	26.38±1.69	90.04±1.62	60.68±1.58
DKM	87.12±1.84	75.82±1.68	43.26±1.83	27.85±2.17	91.52±2.16	63.24±1.63
SC	86.45±1.82	76.06±2.41	42.72±1.54	28.16±2.08	92.37±1.85	62.83±1.84
SSC	88.83±1.69	77.85±1.66	44.36±1.83	29.44±1.84	92.24±2.23	64.19±1.36
BKM	87.01±1.13	77.68±1.45	45.01±1.94	28.63±1.66	93.48±2.07	65.88±1.69
BCLR	88.96±1.75	76.59±1.82	46.80±1.88	30.15±1.95	94.43±1.78	<b>67.34±1.62</b>
SSC [a]	86.15±1.60	76.94±1.85	44.85±1.98	28.68±2.16	92.85±1.88	64.92±1.68
DGSC [b]	88.54±1.92	78.25±1.72	46.82±2.17	30.52±1.85	94.26±1.60	66.26±1.74
Sep. 2D proj. & Cluster.	86.82±1.95	77.21±2.08	44.82±1.92	27.96±2.21	92.64±1.95	63.16±1.88
A2DEIC	<b>90.14±1.88</b>	<b>79.74±1.56</b>	<b>48.76±2.05</b>	<b>32.68±1.86</b>	<b>96.15±1.97</b>	67.16±1.95

## Dataset Description

We evaluate the performance of the proposed algorithm on the following datasets.

- UUIIm Head Post and Gaze dataset (Weidenbacher et al. 2007): This dataset has been widely used to test the performance in terms of head pose and gaze recognition. There are 2,220 images in this dataset, which can be grouped to 10 different people. Each image has been resized to  $24 \times 32$ .
- CVL handwritten dataset (Mouchère et al. 2013): We use this dataset to evaluate the performance in terms of handwritten digit recognition. This dataset has 21,780 handwritten digital images. All the images in this dataset are resized to  $32 \times 32$ .
- Pointing'04 Head Pose dataset (Gourier, Hall, and Crowley 2004): This dataset has been widely used for head pose estimation. The dataset contains 2,790 images, which is captured from 15 different persons. In this work, each image was resized to  $40 \times 30$ . We report the experimental results in terms of tilt and pan angles.
- USPS handwritten digit dataset: This dataset is used to evaluate the performance of handwritten digit recognition of the proposed algorithm. This dataset constitutes of 9,298 images. We resize each image to  $16 \times 16$ .
- Coil20 object dataset (Nene, Nayar, and Murase 1996): There are 1,440 images in this dataset, which can be grouped into 20 objects. We resize all the images to  $32 \times 32$ .

## Experimental Setup

We compare the proposed algorithm with *k*-means (KM), Discriminative *k*-means (DKM) (Ye, Zhao, and Wu 2007), Spectral Clustering (SC) (Chen et al. 2011), Spectral Shrunken Clustering (SSC) (Chang et al. 2015), Balanced *k*-means (BKM) (Malinen and Fränti 2014) and Balanced Clustering with Least Square Regression (BCLSR) (Liu et al. 2017). For fair comparison, we tune the parameters of all the compared algorithms by grid search, from the range of  $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$ .

Following previous works (Chang et al. 2015; Liu et al. 2017), we employ clustering accuracy (ACC) and normalized mutual information (NMI) as the evaluation metrics.

Let  $q_i$  represent the clustering label result from a clustering algorithm and  $p_i$  represent the corresponding ground truth label of an arbitrary data point  $x_i$ . Then *ACC* is defined as follows:

$$ACC = \frac{\sum_{i=1}^n \delta(p_i, \text{map}(q_i))}{n}, \quad (16)$$

where  $\delta(x, y) = 1$  if  $x = y$  and  $\delta(x, y) = 0$  otherwise.  $\text{map}(q_i)$  is the best mapping function that permutes clustering labels to match the ground truth labels using the Kuhn-Munkres algorithm. A larger ACC indicates better clustering performance.

For any two arbitrary variables  $P$  and  $Q$ , NMI is defined as follows:

$$NMI = \frac{I(P, Q)}{\sqrt{H(P)H(Q)}}, \quad (17)$$

where  $I(P, Q)$  computes the mutual information between  $P$  and  $Q$ , and  $H(P)$  and  $H(Q)$  are the entropies of  $P$  and  $Q$ . Let  $t_l$  represent the number of data in the cluster  $\mathcal{C}_l$  ( $1 \leq l \leq c$ ) generated by a clustering algorithm and  $\tilde{t}_h$  represent the number of data points from the  $h$ -th ground truth class. NMI metric is then computed as follows:

$$NMI = \frac{\sum_{l=1}^c \sum_{h=1}^c t_{l,h} \log\left(\frac{n \times t_{l,h}}{t_l \tilde{t}_h}\right)}{\sqrt{(\sum_{l=1}^c t_l \log \frac{t_l}{n})(\sum_{h=1}^c \tilde{t}_h \log \frac{\tilde{t}_h}{n})}}, \quad (18)$$

where  $t_{l,h}$  is the number of data samples that lie in the intersection between  $\mathcal{C}_l$  and  $h$ -th ground truth class. Similarly, a larger NMI indicates better clustering performance.

## Experimental Results

We report the experimental results in Tables 1 and 2. The performance is reported in percentages. From the experimental results, we can see that the proposed method consistently performs better than the other compared algorithms

Table 2: We evaluate the performance of all the compared algorithms in terms of normalized mutual information (NMI±Standard Deviation). Performance is reported in percentages.

	NMI					
	UUIIm	CVL	Pointing'04 (tilt)	Pointing'04 (pan)	USPS	Coil20
<i>k</i> -means	63.58±2.36	58.72±1.95	33.83±1.75	18.78±2.28	83.95±1.86	48.74±1.64
DKM	66.91±2.52	60.24±1.58	35.52±2.05	19.85±1.82	84.37±2.18	50.18±1.87
SC	66.54±2.16	62.41±2.05	34.96±1.82	21.49±2.07	84.96±2.35	52.27±1.82
SSC	68.02±1.95	64.83±1.89	36.28±1.96	20.88±2.25	88.01±1.98	51.86±1.96
BKM	69.38±1.87	63.98±2.06	38.04±2.12	22.29±1.98	89.67±2.16	53.95±1.82
BCLR	70.40±1.92	65.12±2.68	37.72±2.36	23.02±1.87	88.75±1.79	54.82±2.14
SSC [a]	68.25±1.82	64.96±1.82	36.42±2.05	21.12±1.99	88.24±1.86	52.04±2.04
DGSC [b]	70.14±1.97	64.86±2.04	38.08±1.86	23.04±2.06	89.17±2.21	53.75±1.86
Sep. 2D proj. & Cluster.	67.28±2.06	62.97±2.18	36.46±1.96	21.42±1.85	87.87±1.92	52.09±1.84
A2DEIC	<b>72.86±1.88</b>	<b>67.46±2.44</b>	<b>40.18±2.29</b>	<b>25.16±2.14</b>	<b>91.42±1.78</b>	<b>56.78±2.06</b>

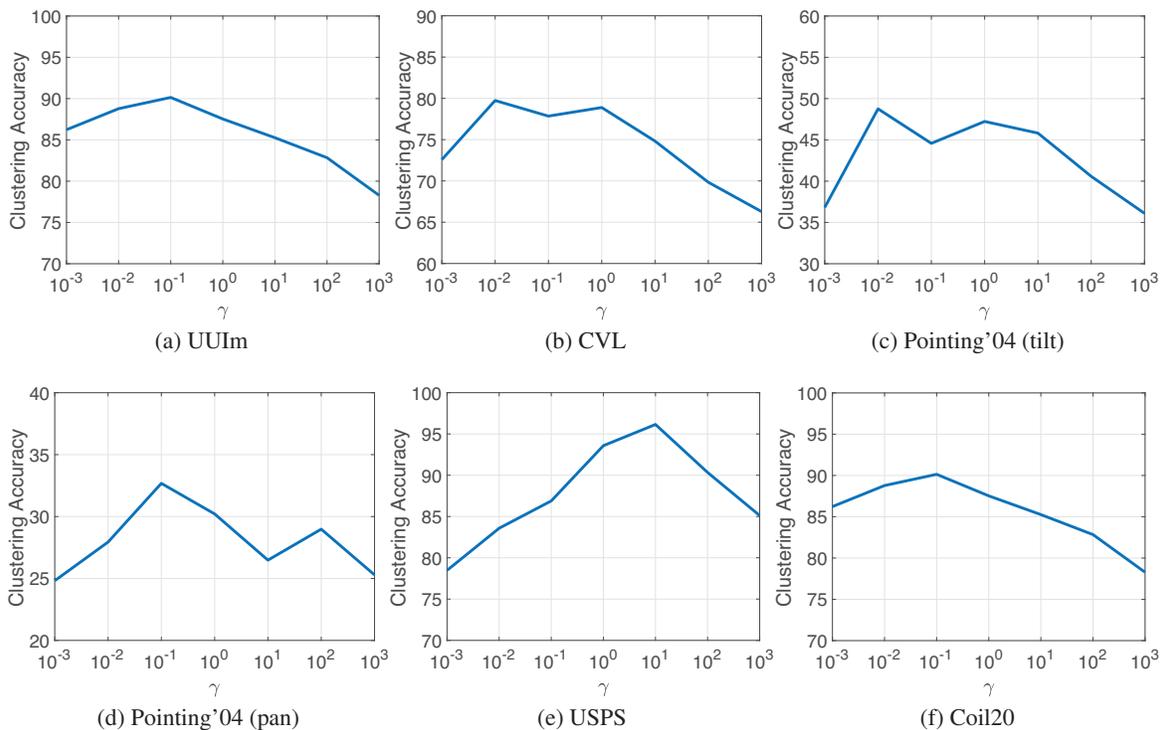


Figure 1: We show the clustering performance variations of the proposed algorithm w.r.t. various parameters. We use clustering accuracy as the evaluation metric. From the experimental results, we observe that the proposed algorithm has a consistent preference on parameter setting, which makes it uncomplicated to get optimal parameter value in practice.

in terms of both evaluation metrics. We have the following observations:

- The spectral clustering performs better than the classical *k*-means and its variants. From this observation we can see the advantage of exploring the pairwise similarity information between all data points from a weighted graph adjacency matrix.
- BCLR is the second best algorithm. This is because it employs a balance constraint to regularize the clustering model, and gets a balanced clustering result.

- The proposed algorithm generally achieves the best results of clustering in terms of both clustering accuracy and normalized mutual information, which demonstrates the superiority of two-dimensional embedding for image clustering.

### Sensitivity Analysis

In this subsection, we conduct extensive experiments to analyze the sensitivity of the proposed algorithm w.r.t. the parameter  $\lambda$  in the objective function, and report the results in

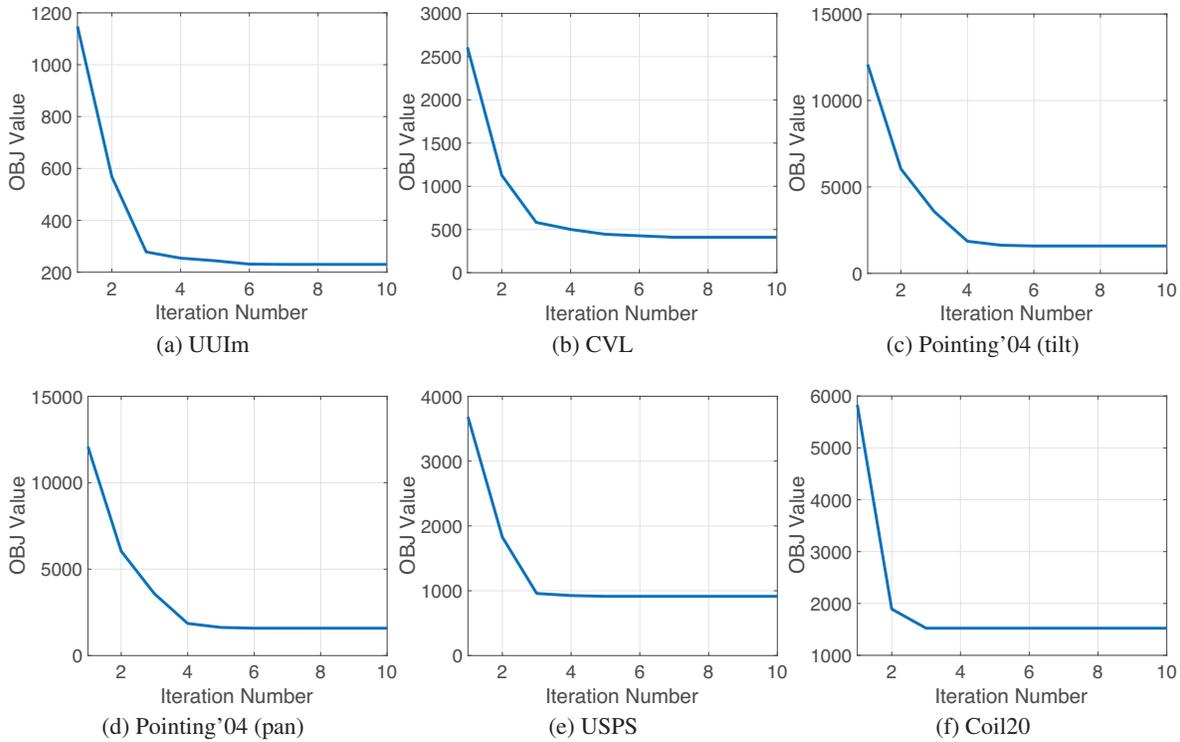


Figure 2: The convergence curves of the proposed algorithm on all the used datasets. From these figures, we can observe that the objective function generally converges within 10 steps, which is very efficient.

Figure 1. Clustering accuracy is used as an example in this subsection. From the experimental results we can see that the performance of the proposed algorithm varies when we tune the parameter  $\lambda$ . However, we also observe that the proposed algorithm consistently has a better performance when  $\lambda$  is in the range of  $\{10^{-2}, 10^{-1}, 10^0\}$ , which makes it not complicated to get the optimal parameter in the practical applications. We also have similar observations when normalized mutual information (NMI) is used.

### Convergence Study

In the previous section, we have proved the convergence of the proposed algorithm. In this subsection, we conduct extensive experiments to demonstrate the convergence of the proposed algorithm. Following (Chang et al. 2015), we fix the parameter  $\lambda$  at 1. The experimental results are shown in Figure 2. From the experimental results we can observe that the objective function value converges very quickly, which is generally within 10 steps. The convergence speed demonstrates the effectiveness of the proposed algorithm.

### Conclusion

In this paper, we have proposed a novel adaptive two-dimensional embedded image clustering algorithm. Our work is built on the intuition that an image’s 2D structure should be better represented by a matrix form. To solve the challenging objective function, we propose a fast iterative solution. We have conducted extensive experiments

to validate the superiority of the proposed algorithm in terms of clustering accuracy using five different datasets. In the future, we plan to apply the proposed algorithm for other related applications, *i.e.* image segmentation, co-segmentation, *etc.*

### Acknowledgements

This paper is partially supported by Taishan Scholar Project of Shandong, China, partially supported by National Natural Science Foundation of China under grant no. U1836216,61906109, and partially supported by Australian Research Council under Discovery Program (DP160104075).

### References

Abassi, L., and Boukhris, I. 2019. A worker clustering-based approach of label aggregation under the belief function theory. *Appl. Intell.* 49(1):53–62.

Belhumeur, P. N.; Hespanha, J. P.; and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* 19(7):711–720.

Caron, M.; Bojanowski, P.; Joulin, A.; and Douze, M. 2018. Deep clustering for unsupervised learning of visual features. In *ECCV*.

Chang, X.; Nie, F.; Ma, Z.; Yang, Y.; and Zhou, X. 2015. A convex formulation for spectral shrunk clustering. In *AAAI*.

Chang, X.; Nie, F.; Wang, S.; Yang, Y.; Zhou, X.; and Zhang, C. 2016. Compound rank-k projections for bilinear analysis. *IEEE Trans. Neural Netw. Learning Syst.* 27(7):1502–1513.

- Chen, W.; Song, Y.; Bai, H.; Lin, C.; and Chang, E. Y. 2011. Parallel spectral clustering in distributed systems. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(3):568–586.
- Chen, X.; Yang, J.; and Jin, Z. 2014. An improved linear discriminant analysis with l1-norm for robust feature extraction. In *ICPR*.
- Datta, R.; Joshi, D.; Li, J.; and Wang, J. Z. 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.* 40(2):5:1–5:60.
- Ding, C. H. Q., and Li, T. 2007. Adaptive dimension reduction using discriminant analysis and  $K$ -means clustering. In *ICML*.
- Ding, C., and Ye, J. 2005. 2-dimensional singular value decomposition for 2d maps and images. In *ICDM*.
- Gordon, S.; Greenspan, H.; and Goldberger, J. 2003. Applying the information bottleneck principle to unsupervised clustering of discrete and continuous image representations. In *ICCV*.
- Gourier, N.; Hall, D.; and Crowley, J. L. 2004. Estimating face orientation from robust detection of salient facial structures. In *ICPR*.
- Hou, C.; Jiao, Y.; Nie, F.; Luo, T.; and Zhou, Z. 2017. 2d feature selection by sparse matrix regression. *IEEE Trans. Image Processing* 26(9):4255–4268.
- Jain, A. K., and Dubes, R. C. 1988. *Algorithms for Clustering Data*. Prentice-Hall.
- Kennedy, L. S., and Naaman, M. 2008. Generating diverse and representative image search results for landmarks. In *WWW*.
- Kong, H.; Wang, L.; Teoh, E. K.; Li, X.; Wang, J.-G.; and Venkateswarlu, R. 2005. Generalized 2d principal component analysis for face image representation and recognition. *Neural Networks* 18(5):585–594.
- Kwak, N. 2014. Principal component analysis by  $l_{\{p\}}$ -norm maximization. *IEEE Transactions on Cybernetics* 44(5):594–609.
- Li, X.; Lin, S.; Yan, S.; and Xu, D. 2008. Discriminant locally linear embedding with high-order tensor data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 38(2):342–352.
- Li, X.; Pang, Y.; and Yuan, Y. 2010. L1-norm-based 2dpc. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 40(4):1170–1175.
- Li, C.-N.; Shao, Y.-H.; and Deng, N.-Y. 2015. Robust l1-norm two-dimensional linear discriminant analysis. *Neural Networks* 65:92–104.
- Liu, H.; Han, J.; Nie, F.; and Li, X. 2017. Balanced clustering with least square regression. In *AAAI*.
- Lu, G.-F.; Zou, J.; and Wang, Y. 2016. L1-norm and maximum margin criterion based discriminant locality preserving projections via trace lasso. *Pattern Recognition* 55:207–214.
- Luo, D.; Ding, C.; and Huang, H. 2009. Symmetric two dimensional linear discriminant analysis (2ldda). In *CVPR*.
- Luo, D.; Huang, H.; and Ding, C. H. Q. 2011. Discriminative high order SVD: adaptive tensor subspace selection for image classification, clustering, and retrieval. In *ICCV*.
- Ma, Z.; Yang, Y.; Nie, F.; and Sebe, N. 2013. Thinking of images as what they are: Compound matrix regression for image classification. In *IJCAI*.
- Malinen, M. I., and Fränti, P. 2014. Balanced k-means for clustering. In *IAPR S+SSPR*.
- Mouchère, H.; Viard-Gaudin, C.; Zanibbi, R.; Garain, U.; and Kim, D. H. 2013. ICDAR 2013 CROHME: third international competition on recognition of online handwritten mathematical expressions. In *ICDAR*.
- Nene, S. A.; Nayar, S. K.; and Murase, H. 1996. Columbia object image library (coil-20). Technical report, Technical Report CUCS-005-96.
- Oh, J. H., and Kwak, N. 2013. Generalization of linear discriminant analysis using lp-norm. *Pattern Recognition Letters* 34(6):679–685.
- Peng, X.; Feng, J.; Xiao, S.; Yau, W.; Zhou, J. T.; and Yang, S. 2018. Structured autoencoders for subspace clustering. *IEEE Trans. Image Processing* 27(10):5076–5086.
- Wang, R.; Nie, F.; Yang, X.; Gao, F.; and Yao, M. 2015. Robust 2dpc with non-greedy l1-norm maximization for image analysis. *IEEE transactions on cybernetics* 45(5):1108–1112.
- Wang, Q.; Gao, Q.; Gao, X.; and Nie, F. 2017. Optimal mean two-dimensional principal component analysis with f-norm minimization. *Pattern Recognition* 68:286–294.
- Wang, Y.; Wu, L.; Lin, X.; and Gao, J. 2018. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE Trans. Neural Netw. Learning Syst.* 29(10):4833–4843.
- Weidenbacher, U.; Layher, G.; Strauss, P.-M.; and Neumann, H. 2007. A comprehensive head pose and gaze database. In *Proc. 3rd IET Int. Conf. Intell. Environ.*
- Xia, Y.; Nie, L.; Zhang, L.; Yang, Y.; Hong, R.; and Li, X. 2016. Weakly supervised multilabel clustering and its applications in computer vision. *IEEE Trans. Cybernetics* 46(12):3220–3232.
- Xu, R., and II, D. C. W. 2005. Survey of clustering algorithms. *IEEE Trans. Neural Networks* 16(3):645–678.
- Yan, S.; Xu, D.; Yang, Q.; Zhang, L.; Tang, X.; and Zhang, H.-J. 2007. Multilinear discriminant analysis for face recognition. *IEEE Transactions on Image Processing* 16(1):212–220.
- Yang, J., and Liu, C. 2007. Horizontal and vertical 2dpc-based discriminant analysis for face verification on a large-scale database. *IEEE Transactions on Information Forensics and Security* 2(4):781–792.
- Yang, J.; Zhang, D.; Frangi, A. F.; and Yang, J.-y. 2004. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence* 26(1):131–137.
- Ye, J.; Janardan, R.; and Li, Q. 2005. Two-dimensional linear discriminant analysis. In *NIPS*.
- Ye, J.; Zhao, Z.; and Wu, M. 2007. Discriminative k-means for clustering. In *NIPS*.
- Ye, J.; Zhao, Z.; and Wu, M. 2008. Discriminative k-means for clustering. In *NIPS*.
- Yu, S. X., and Shi, J. 2003. Multiclass spectral clustering. In *ICCV*.
- Zelnik-Manor, L., and Perona, P. 2004. Self-tuning spectral clustering. In *NIPS*.
- Zhang, D., and Zhou, Z.-H. 2005. (2d) 2pca: Two-directional two-dimensional pca for efficient face representation and recognition. *Neurocomputing* 69(1):224–231.
- Zhang, F.; Yang, J.; Qian, J.; and Xu, Y. 2015. Nuclear norm-based 2-dpca for extracting features from images. *IEEE transactions on neural networks and learning systems* 26(10):2247–2260.
- Zhong, F., and Zhang, J. 2013. Linear discriminant analysis based on l1-norm maximization. *IEEE Transactions on Image Processing* 22(8):3018–3027.
- Zhong, F.; Zhang, J.; and Li, D. 2014. Discriminant locality preserving projections based on l1-norm maximization. *IEEE transactions on neural networks and learning systems* 25(11):2065–2074.