An ordered sparse subspace clustering algorithm based on p-Norm

Liping Chen¹,² | Gongde Guo¹,²* | Hui Wang³

¹ School of Mathematics and Informatics, Fujian Normal University, Fuzhou, China
² Digital Fujian Internet-of-Things Laboratory of Environmental Monitoring, Fujian Normal University, Fuzhou, China
³ School of Computing and Mathematics, Faculty of Computing and Engineering, Coleraine, UK

Correspondence
Gongde Guo, School of Mathematics and Informatics, Fujian Normal University, Fujian Province, China.
Email: ggd@fjnu.edu.cn

Present Address
*School of Mathematics and Informatics, Fujian Normal University

Funding information
National Natural Science Foundation of China; Natural Science Foundation of Fujian Province; Leading Program in the Science and Technology Department of Fujian Province, Grant/Award Number: 2015Y0054; Natural Science Foundation, Grant/Award Number: 2018J01776; Chinese National Natural Science Foundation, Grant/Award Numbers: 41601477, 61772134 and 61672157

Abstract
Images in video may include both Gaussian noise and geometric rotation. Thus, it is challenging to represent an image sequence in its intrinsically low-dimensional space in a noise-robust and rotation-robust manner. In this paper, we propose a novel ordered sparse subspace clustering algorithm based on a p-norm to achieve an effective clustering of sequential data under heavy noise conditions. We also use the wavelet-histogram of oriented gradient (HOG) transform in the kernel view to extract both the global features (with the wavelet process) and the local features (with the HOG process) from the image. In addition, we assign different weights to different features to obtain a sparse coefficient matrix that helps to emphasize the global and local correlations in each sample. Similarly, the clustering algorithm based on the p-norm for sequential images emphasizes the within-class correlations amongst samples. Therefore, in this paper, we select additional denoising main components under a Laplacian constraint to achieve a better block-diagonal structure and highlight the independence of different clusters. Extensive experiments performed on various public datasets (including the ordered face dataset, handwritten recognition dataset, video scene segmentation dataset, and object recognition dataset) demonstrate that the proposed method is more resilient to noise and rotation than other representative sparse subspace clustering methods.

KEYWORDS
sparse subspace clustering, wavelet-HOG, block-diagonal transform, image sequence clustering

1 | INTRODUCTION

Rapid advances in digital technology have provided people with the ability to record their lives using their mobile phones. Additionally, many people connect with their friends by uploading pictures and videos to social media websites. According to statistics released on YouTube’s seventh anniversary, the equivalent of 72 hr of video is uploaded to their website every minute, and this number is constantly increasing. Although video analysis can reveal a considerable amount of useful information, it remains a challenging task.

Image sequence data have typical “big data” characteristics because they usually include large numbers of data samples (i.e., images). The image dimensions are often high; therefore, an enormous number of features exist without any preprocessing. Additionally, there is, inevitably, noise in image sequences, which degrades the image quality and may even blur object boundaries. Finally, an object’s appearance may change significantly due to changes in illumination, different poses, and unexpected rotations. We attempt to solve these problems by
considering the nature of an image's structure. One component of an image's structure lies in its graphic features; image structure can also be described by certain probabilistic and statistical properties, such as a histogram. A wavelet transform can effectively capture the graphical features of an image and effectively resist noise, even when the image is characterized by strong noise. The histogram of oriented gradient (HOG) is a common method for extracting features that is based on statistical classification. The combined wavelet-HOG detection method can extract features that reflect the global and spatial coherence constraints using a combination of wavelet high-low (HL) and low-high (LH) sub-bands.

Although computer vision has made significant progress since its inception, it still presents a significant gap compared with human abilities for understanding video content. One difficulty associated with video analysis is the high dimensionality in video data representations. Image sequence data have typical big data characteristics, and including large numbers of data samples (images) and image dimensions are typically relatively high. In real-world scenarios, images are produced by cameras or webcams with several independent physical parameters. Therefore, the lower dimensional spaces associated with these physical parameters exist in each high-dimensional image. Vageeswaran, Mitra, and Chellappa noted that a nine-dimensional linear space is sufficient to characterize all facial images under all possible illumination conditions (Vageeswaran, Mitra, & Chellappa, 2013). Theoretical studies have shown that natural scenes are represented in the brain in the primary visual cortex (area V1) of the visual system via a sparse structure (Vinje & Gallant, 2000). Therefore, we select subspace clustering algorithms as a useful approach (Elhamifar & Vidal, 2013; Elhamifar, & Vidal, 2009) for simultaneously reducing data dimensionality and clustering data.

In this paper, we propose a novel subspace clustering method that uses wavelet-HOG mapping to capture the inner structures of images and apply the block-diagonal transformation, after which we perform spectral clustering-based denoising on image sequences. In addition to constructing a special dictionary to separate the samples, we add a penalty term during the problem formulation to enhance the correlation between neighbouring images in the time domain and use the selected denoising method to maximize the within-class correlation and minimize the between-class correlation. Experimental comparisons conducted with handwriting databases, face datasets, and video scene segmentation databases clearly demonstrate that the proposed method is better at identifying within-image structure, enlarging the within-class correlation and reducing the between-class correlation; consequently, it achieves more robust results.

2 | SPARSE SUBSPACE CLUSTERING

Subspace clustering technology is a promising approach for finding the intrinsic low-dimensional structure of image sequences. Several methods exist to construct a proper affinity matrix (Elhamifar & Vidal, 2013; Chen & Yang, 2014; Vidal & Favaro, 2014; Tierney, Gao, & Guo, 2014; Tierney et al., 2014; Feng, Lin, Xu, & Yan, 2014; Pham, Budhaditya, Phung, & Venkatesh, 2012; Chen, Guo, & Chen, 2017). Some studies have used algorithms to search for each subspace via statistical learning methods (Rao, Yang, Sastry, & Ma, 2010) until they achieve an adequate number of inliers. Other studies have applied algebraic techniques, such as the generalized principal component analysis (PCA) method, to make the subspace fit each naturally occurring cluster. Some algorithms iteratively assign points to subspaces (Ma, Yang, Derksen, & Fossum, 2008); however, the results are sensitive to initialization (Bradley & Mangasarian, 2000; Ho, Yang, Lim, Lee, & Kriegman, 2006). Other algorithms employ matrix factorization to approach the orthogonal subspace decomposition (Boult & Brown, 1991; Wu, Zhang, Huang, & Lin, 2001).

The sparse subspace clustering (SSC) (Elhamifar & Vidal, 2013) algorithm is one of the most representative subspace clustering algorithms. It achieves good performance in a clean version of the Yale B face database by minimizing the 1-norm of the coefficient vector. The low rank property is another type of sparse representation (Wu et al., 2001; Wang, Xiao-Ping, Feng, & Wang, 2015); therefore, low rank representation subspace clustering (LRR; Chen & Yang, 2014; Vidal & Favaro, 2014) has received much attention. One study used the nuclear norm to substitute for the rank function in an attempt to find the low rank property of data samples, even with multiview and multimodal data (Xin, Wang, Gao, Wipf, & Xin, 2017). However, representing image sequences in intrinsically low-dimensional space in a noise-robust way remains a challenge. Recently, a novel sparsity-promoting algorithm was proposed (Xin et al., 2017) that adopts a data-dependent penalty by calculating the approximate coefficient matrix with expectation maximization updates for SSC. This data-dependent Subspace Clustering (DD-SSC) algorithm can compensate for the clustering dictionary structure. However, these methods are designed primarily for vectors; comparatively fewer studies have focused on image sequence clustering.

Compared with traditional data types, image sequence data have the following characteristics:

1. Image sequences usually contain large number of data samples.
2. The number of features is enormous prior to preprocessing because a large amount of redundancy exists.
3. Real data in image sequences often contain noise and geometric rotation.
4. Prior knowledge of the intrinsic low rank structure is lacking.
5. Adjacent samples in the time domain are similar.
Consider an image sequence: clearly, adjacent frames are correlated as time advances. Ordered sparse subspace clustering (OSC; Tierney et al., 2014) imposes an additional penalty to highlight the continuous relationship of the adjacent frames. However, because the penalty structure works on the original pixel vector, it is unsuitable for geometric structures such as 2D images without any transform. Applying the penalty to an image sequence requires some preprocessing to identify the global and local features within the frame. Other subspace clustering methods in recent studies have focused on the regularisation design. For example, reweighted sparse subspace clustering (Xu, Xu, Chen, & Ruan, 2015) involves iteratively weighting the 1-norm minimization to achieve a better sparse representation estimate. The contribution lies in applying different weights to different features in the norm calculation to guarantee the effects of subspace clustering. In paper (Feng et al., 2014), the Laplacian graph constraint was taken into consideration to obtain the coefficient matrix with a block-diagonal transformation prior. The results highlight the differences between different classes (Feng et al., 2014) using the block-diagonal structure.

In high-dimensional space, the distance between two samples tends to be the same; however, the distance between one sample point and its nearest and farthest neighbours tends to decrease when the conditions are noisy. Therefore, the distance based on all global features is insufficient to distinguish between samples. Traditional SSC algorithms do not distinguish between the feature attributes well enough to calculate the affinity matrix, and there is no distinction between key and redundant features in the dictionary construction. DD-SSC compensates for the clustering dictionary structure with the expectation maximization process. In contrast to DD-SSC, the ordered sparse subspace clustering algorithm based on the p-norm (LpOSC) balances the clustering dictionary structure by virtue of the proposed wavelet-HOG transform on the image.

3 | KERNEL VIEW OF WAVELET AND HOG

3.1 | Kernel view of wavelet

Real-life data in image sequences often include noise and geometric attacks. When the image acquisition frequency differs from the working frequency of some lights, the mismatch can cause jitter in the image. When the camera-view angle changes, the object will rotate. Additionally, the appearances of objects may change due to different incident light levels.

As a "mathematical microscope" (Unser, 2015), the wavelet transform can mirror an image's structure. A wavelet transform can effectively distinguish edges in the image from the main elements. Using wavelet transforms (Mallat & Zhang, 1993; Deng, Li, & Fu, 2010), the image is divided into sub-bands: Xs, Xh, Xv, and Xvh. Setting s as the scaling factor and (a,b) as the shift value, the wavelet transform can be described as follows:

\[
X_{\text{warf}}(s, a, b) = \int \frac{1}{|s|} f(x_{\text{img}}, y_{\text{img}}) \varphi \left( \frac{x_{\text{img}} - a}{s}, \frac{y_{\text{img}} - b}{s} \right) dx_{\text{img}} dy_{\text{img}}.
\]  

Most of the energy in an image is concentrated in the low frequency sub-band Xs of the wavelet transform. Therefore, this sub-band is resistant to noise. The high-frequency sub-bands Xh, Xv, and Xvh reflect the information from the edges of the image. It is appropriate to calculate the direction and the amplitude of the local region gradient in the high-frequency sub-bands. It was shown that the continuous wavelet transform image space can reproduce the kernel Hilbert space in the paper (Deng et al., 2010). This property results from the ability of wavelets to capture short high-frequency phenomena such as transient signals when identifying nonlinear dynamic systems.

3.2 | The kernel view of HOG

The HOG operator is a local feature proposed by Dalal and Triggs (Dalal & Triggs, 2005). It provides an effective statistical analysis of the orientation and amplitude of the gradients in an image. After obtaining complete edge information image from the wavelet transform, the HOG function is implemented as follows (Dalal & Triggs, 2005; Bo, Ren, & Fox, 2010):

1. The input image is normalized. Given a pixel x, calculate the direction and intensity of each pixel gradient as \( \theta(x) \) and \( m(x) \), respectively.
2. The image is subdivided into an \( m \times n \) patch denoted as \( P \), and the gradient magnitude in each \( P \) is normalized as \( \hat{m}(x) = \frac{m(x)}{\sqrt{\sum_{x \in P} m^2(x)} + \epsilon} \)
   where \( \epsilon \) stands for a very small constant. Thus, we can easily obtain the unit histogram: \( H(x) = \frac{m(x)}{\sqrt{\sum_{x \in P} m^2(x)} + \epsilon} \).
3. After the unit histogram is normalized and grouped, the resulting descriptor \( \text{HOG}(P) \) resists contrast and adapts to changes in illumination. \( \text{HOG}(P) \) is calculated as follows:

\[
\text{HOG}(P) = \sum_{x \in P} \frac{m(x) \delta(x)}{\sqrt{\sum_{x \in P} m^2(x)} + \epsilon}
\]  

\( \text{HOG}(P) \) is a rotationally invariant feature, which is robust to illumination changes.
Here, $\delta(x) = [\delta_1(x), \cdots, \delta_d(x)]$ is the d-dimensional indicator function, and each $\delta_i(x) = \max(\cos(\theta(x) - c_i)^9, 0)$, where $c_i$ represents the centre of the $i$th bin. The similarity between two patches equates to the inner product of the linear kernel in kernel-space feature mapping (Bo et al., 2010).

4 | A NOISE-RESILIENT SSC ALGORITHM

4.1 | LpOSC based on wavelet-HOG processing

In our method, no prior knowledge of the dimension or cardinal information of the subspaces exists at the time that subspace clustering is used to evaluate the subspace to which each sample belongs. Traditional SSC often uses the original pixel vector to calculate this affinity, but that approach ignores the local image structure that we want to reveal.

The pixels of images in a sequence have strong spatial similarities, and therefore, large redundancies. However, general SSC algorithms regard images as vectors in which each pixel is assumed to be independent during the clustering process. These clustering methods are vulnerable to severe noises partly because dictionaries only copy the 1-D structural features of the data. Pixels are locally correlated within an image. Here, it is important to note that we apply the multiscale wavelet transform to reflect the spatial features of the image before constructing the dictionary. As shown in Figure 1, the wavelet transform preprocessing (Mallat & Zhang, 1993; Deng et al., 2010) divides the image into sublevel images, thus revealing the within-image structure of each vector. Wavelet analysis describes the data from different angles; here, $X_s$ represents the low frequency image, which includes most of the energy. Next, $X_h$, $X_v$, and $X_{vh}$ represent the sub-bands containing the details of the image structure, such as the horizontal, vertical, and diagonal directions. The details in the $X_h$, $X_v$, and $X_{vh}$ sub-bands reflect edge information and most of the high-frequency noise that is not captured in the $X_s$ sub-band of the wavelet. Consequently, the wavelet transform reflects the within-image neighbour features in both the time and frequency domain during signal processing.

The wavelet transform performs well at separating signals from noise; it also reflects the image structure information in both the space and the frequency domain. We construct a multidirectional local dictionary using the wavelet transform in the proposed algorithm that combines the spatial detail and structural information of the image at different scales. When $M = \begin{bmatrix} -1 & 1 & \cdots & 1 \\ 1 & -1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & -1 \end{bmatrix} \in \mathbb{R}^{N(N-1)}$, we use the penalty $\|MZ\|_p$, which reflects the similarity between adjacent frames in the time domain; $\|\|_p$ represents Norm 1 when $p = 1$ and Norm 2 when $p = 2$.

The resulting optimization function is defined as follows:

$$\min_{Z \in \mathbb{R}^{N(N-1)}} \|Z\|_1 + \beta \|MZ\|_p,$$

subject to $\|X_s - X_s Z\|_F^2 < \epsilon_1$, $\|X_h - X_h Z\|_F^2 < \epsilon_2$, $\|X_v - X_v Z\|_F^2 < \epsilon_3$, $\|X_{vh} - X_{vh} Z\|_F^2 < \epsilon_4$, diag$(Z) = 0$.

The wavelet-HOG transform requires preprocessing to eliminate noise. Further studies of wavelet coefficients have shown that for pixels that lie along a horizontal edge, the corresponding wavelet coefficients of those pixels are large in the $X_h$ sub-band but nearly zero in the $X_v$ sub-band. Similarly, for pixels that lie along a vertical edge, the matching wavelet coefficients are large in the $X_v$ sub-band but close to zero in the $X_h$ sub-band. For pixels lying along an oblique edge, the wavelet coefficients are almost identical in both the $X_h$ and $X_v$ sub-bands. Thus,
the coefficients of the $Xh$ and the $Xv$ sub-bands define the contours of the image. Next, HOG is used to extract the shallow features for the constructed contours.

The HOG (Dalal & Triggs, 2005) is a common statistical classification method. It performs well under conditions with light variations, and it effectively characterizes edge characteristics. However, because HOG involves a large amount of overlap and requires calculating histogram statistics, it has a high feature dimension and involves large calculations. During the process of directional fusion, we combine each pixel from the $Xh$ and $Xv$ sub-bands and then apply the HOG method; for example, $\Psi(Xv, Xh) = HOG((Xv + Xh)/2)$. This method not only considers the information in the important direction, but also reduces the computational cost. A good data representation with better mapping from the original space into the feature space can greatly improve the subspace clustering performance. Compared with the original one-dimensional vector, the processed vector reflects some of the detailed structural characteristics of the two-dimensional image in the space domain. Consequently, we use the detected HOG feature set of the image to generate visual dictionaries. We denote $\Psi(Xv, Xh)$ as the vectorization of the HOG feature. Consequently, the objective function in Equation 1 can be rewritten as follows:

$$\begin{align*}
\min_{Z} & \frac{1}{2\lambda} \|Xs - Z\|_F^2 + \gamma \|\Psi(Xv, Xh) - \Psi(Xv, Xh)Z\|_F^2 \\
+ & a \|Z\|_1 + \beta \|MZ\|_F \\
\text{s.t.} \ & \text{diag}(Z) = 0.
\end{align*}$$

(4)

Here, the low-frequency band $Xs$ reveals the global structure of the image. Additionally, the HOG is a local information descriptor that can locally characterize the image based on local gradients. Using this method, we can apply various features to augment the geometric structure of the image; thus, the dictionary in this algorithm is composed of a set of multidirection wavelet and HOG coefficients with different weights. We can expect the sparse representation method along with the dictionary described above to reduce the noise. That is, using prior knowledge of the image structure can help restore the latent relationships of the corrupted image structure. As a result, the impact of noise is decreased, and there is greater similarity in the spatial structure. Thus, this approach is beneficial for evaluating the similarities of scenes in video frames.

4.2 The denoising block-diagonal transformation based on the structure of spectral clustering

The global similarity measure is often used in spectral clustering methods because it helps to ignore local differences. In contrast, the wavelet-HOG transform method converts an image into special vectors and explicitly considers local differences.

From a statistical point of view, the regularization process is actually based on an a priori image information constraint that requires sparse distribution. Generally, in noise-free cases, the coefficient matrix for coding the coefficients is directly related to each specific subspace under a given sparsity constraint. Therefore, most elements in the coefficient matrix are zero, and the number of non-zero coefficients is small. As the noise interference increases, the coefficients in the matrix tend to become non-zero. This change strongly affects the block-diagonal structure of the coefficient matrix and leads to a decline in the subspace clustering performance. For a single image, each column corresponds to the same noise in the sequence. To find the inner correlation between the representations, we use the following calculation:

$$z_{ij} = \begin{cases} 
0, & \text{if } z_{ij} < \frac{\sum_{j=1}^{n} z_{ij}}{n} \\
\frac{\sum_{j=1}^{n} z_{ij}}{n}, & \text{else} 
\end{cases}$$

(5)

This approach causes the coefficient matrix to become sparser because the negative coefficients are deleted as the noise interference is reduced. Next, we select the $n$ largest elements to strengthen the internal consistency of the matrix. These steps improve the clustering accuracy. In addition to minimizing the differences between different clusters, we select the major components based on a threshold to suppress the noise interference, emphasizing consistency within the class. As Figure 2 shows, the values of the coefficient matrix $Z$ are directly related to each specific subspace in the noise-free case. When $z_{ij}$ does not equal 0, then samples $i$ and $j$ belong to the same cluster. In contrast, when $z_{ij}$ equals 0, samples $i$ and $j$ originate from different clusters. Therefore, the coefficient matrix $Z$ in SSC always has a block-diagonal structure, which reflects the differences between different types of data. Specifically, when a mismatch occurs between the actual noise and the model, it is unlikely that the resulting affinity matrix will have a block-diagonal structure. Traditional spectral clustering methods have difficulty in achieving robust segmentation. We apply Equation 6 (Feng et al., 2014) to get the main components of the denoising data to ensure the block-diagonal structure.
\[ Z \in \kappa = \{ Z | \text{rank}(Lw) = n - k, \ W = \frac{1}{2}(|Z| + |Z|^T) \}. \] (6)

### 4.3 Implementation of the optimization algorithm

The subspace clustering method cannot be directly applied to nonlinear image data. In the kernel view of the wavelet-HOG, we apply the wavelet-HOG transform to general image vectors in an effort to measure the similarity between image patches in the feature mapping kernel space. Because the wavelet transform reflects the global image structure and the HOG reveals local characteristics, it is reasonable to combine the global and local structural information to construct a multiscale local dictionary. In the spatial domain, the data from different perspectives are preassigned different weights to emphasize the global and local characteristics of different spatial data. In the time domain, the design of regular items is used to imitate the continuity of adjacent sample data to reflect the association between local and global information.

The following equation (6) can be solved as a linear regression problem of L1 norm regularization. If we set \( X = [X_s, \sqrt{2} \Psi(X_v, X_h)]^T \), then Equation 2 becomes:

\[
\begin{align*}
\min_{Z, E} & \frac{1}{2} ||X - EZ||^2_F + \alpha |Z|_1 + \beta |MZ|_p, \\
\text{s.t.} & \text{ diag}(Z) = 0.
\end{align*}
\] (7)

For ease of implementation, we employ the alternating direction multiplier method (Boyd, Parikh, Chu, Peleato, & Eckstein, 2010). As the algorithm shows, \( X_s \) is the low frequency part of the image, which includes most of the image's energy, and \( \Psi(X_v, X_h) \) expresses the separate structural details of the local image. To balance the sparsity and similarity of the coefficient matrix of self-representation, different scale representation modules must set \( \sqrt{\gamma} \) to. Thus, Equation 4 is theoretically equivalent to the following:

\[
\begin{align*}
\min_{Z, U} & \frac{1}{2} ||X - EZ||^2_F + \alpha |Z|_1 + \beta |U|_p, \\
\text{s.t.} & Z - \tilde{Z} = 0, U = \tilde{Z}M, \text{ diag}(Z) = 0, \quad (8)
\end{align*}
\]

where \( M = \begin{bmatrix} -1 & 1 & \cdots & 1 \\ 1 & -1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & -1 \end{bmatrix} \in \mathbb{Z}^{N \times (N-1)}. \)

The constraint is guaranteed dynamically during program execution. The original augmented Lagrange function is expressed as follows:

\[
L(Z, \tilde{Z}, U) = \frac{1}{2} ||X - EZ||^2_F + \alpha |Z|_1 + \beta |U|_p + \langle G, Z - \tilde{Z} \rangle + \frac{Y_1}{2} ||Z - \tilde{Z}||^2_F + \langle F, U - \tilde{Z}M \rangle + \frac{Y_2}{2} ||U - \tilde{Z}M||^2_F,
\] (9)

and the proposed optimization algorithm is presented below:
Algorithm

Input: the image sequence $I_1$, $I_2$,$\ldots$, $I_N$
Output: coefficient matrix

Step 1: Initialize $U$, $G$, and $F$; determine the maximum number of iterations; and assign matrix $M$ as shown in Equation 6.
Step 2: Each image in the image sequence is decomposed using the multiscale wavelet transform method; each scale is set with different weights; and the combination is changed into a one-dimension vector. As an example, the $N$th image is composed of the following data: matrix $X = [X_i, \sqrt{\gamma} \Psi(X_i, X_0)]^T$
Step 3: Solve the following expression with ADMM:

\[
\text{while } (\frac{1}{2} \|X - Z^k\|_F^2 + \alpha \|Z\|_1 + \beta \|U\|_p > \epsilon \text{ and } k < k_{\text{max}}) \]

1. Others are set as follows: $Z^{k+1} = \arg\min_Z \|Z\|_1 + < G, Z^k - Z > + \frac{\gamma_1}{2} \|Z^k - Z\|_F^2$
2. The diagonal elements of the matrix are obtained as 0s: $\text{diag}(Z^k) = 0$
3. Setting others as follows, update the variable: $\tilde{Z}^{k+1} = \arg\min_Z \frac{1}{2} \|X - XZ^{k+1}\|_F^2 + < G, Z^{k+1} - Z > + \frac{\gamma_1}{2} \|Z^{k+1} - Z\|_F^2 + < F, U^k - Z^{k+1} M >$
4. Others are set as follows: $U^{k+1} = \arg\min_U \|U\|_p + < F, U^k - Z^{k+1} M > + \frac{\gamma_2}{2} \|U^k - Z^{k+1} M\|_F^2$
5. Update $F$ and $G$ with the following formula:
\[
F = F + \gamma_2 \left( U^{k+1} - Z^{k+1} M \right) \quad G = G + \gamma_1 \left( Z^{k+1} - Z^{k+1} \right)
\]
6. Update the following variables: $\gamma_1 = \rho \gamma_1$, $\gamma_2 = \rho \gamma_2$, $k = k + 1$
end while
Step 4: Denoise the coefficient matrix using Equation 5, sort the coefficients of $Z$ from large to small, and select the first several coefficients.
Step 5: Strengthen the block-diagonal structure of $Z$ using Equation 6:

5 | EXPERIMENTS AND ANALYSIS

To evaluate the effectiveness of the proposed algorithm, we used the proposed OSC algorithm based on a $p$-norm (LpOSC) to cluster images of faces, USPS handwritten digits and video datasets. The different datasets reflect different varieties of noise: the United States Postal Service (USPS) dataset includes shape transformation; the video dataset reflects the dither transform; the Yale B dataset shows the luminosity transformation; and the Columbia University Image Library (COIL) dataset exhibits changes in viewpoint. Some example images are shown in Figure 3. The LpOSC model has two regular parameters, $\alpha$ and $\beta$. When $\alpha$ and $\beta$ are too large, the affine matrix obtained by LpOSC will be a 0 matrix. Therefore, we set $\alpha$ and $\beta$ between 0.01 and 0.1. As shown in Figure 4 and Figure 5 in the Yale B database without noise, the accuracy

![Some Images Samples](image-url)
of LpOSC is not sensitive to changes in $\alpha$ and $\beta$. Therefore, we fixed the parameters of each experiment to the settings suggested by the original authors. In our experiment, we set $\alpha = 0.1$, $\beta = 0.01$, and $\gamma = \frac{1}{2}$, which statistical experiments suggested would optimize the results. In most cases, selecting 8%–50% of the primary elements will decrease the accuracy. We compared the performance of our algorithm with the performances of other representative subspace clustering algorithms, including SSC (Elhamifar & Vidal, 2013), OSC (Tierney et al., 2014), LRR (Chen & Yang, 2014), and the spatial domain subspace clustering algorithm (SpatSC) (Pham et al., 2012). Most studies (Elhamifar & Vidal, 2013; Pham et al., 2012; Tierney et al., 2014) have used the clustering error rate as an evaluation metric; we also followed this practice in our experiments. Specifically, we used the mean of the clustering error rate to ensure a fair comparison.

Under heavy noise conditions, the SSC, OSC, SpatSC, and LRR algorithms perform even more poorly than expected. At the moment, the correct block-diagonal structure is significantly affected when Gaussian noise is added at a magnitude of 0.83 to the Yale B face dataset. The block diagonalization of the affinity matrix obtained by our LpOSC algorithm is superior to the results obtained by other algorithms. A poorly structured affinity matrix (due to lack of diagonalization) significantly affects the estimation results of the SSC, OSC, SpatSC, and LRR methods. As shown in Figure 6, the OSC, SpatSC, and LRR algorithms have error rates of 0.52, 0.74, and 0.69, respectively. In contrast, LpOSC ($p = 1$) and SSC have lower error rates of 0.22 and 0.28, respectively. As expected, LpOSC ($p = 2$) achieves the lowest error rate, 0.12.

5.1 Digital handwriting recognition experiment

The digital handwriting recognition data were sourced from the Mixed National Institute of Standards and Technology (MNIST) dataset (Lécun, Bottou, Bengio, & Haffner, 1998) and the USPS (Hull, 2002) handwriting dataset. The USPS dataset contains 9,298 greyscale images of
FIGURE 6  Final Clustering Results of different algorithms from heavy attacks

FIGURE 7  Error rates under different Gaussian Attacks on MNIST dataset
handwritten digits (0 to 9). Each image has a resolution of $16 \times 16$ pixels. We selected the first 100 images of the first seven types for clustering analysis. Differences in handwriting result in large differences in the shapes, sizes, and line widths of the handwritten numbers. Thus, the original features of the 700 images do not always reflect the features critical to recognition. In this experiment, the original features were first vectorized to a 60-dimensional vector using PCA (Tipping & Bishop, 1999). To conduct a fair comparison, we used feature vectors of the images transformed by wavelet-HOG to 60 dimensions using PCA. Of the handwritten numerals in the MNIST dataset, half came from high school students and half from the Census Bureau. The size of each image is $32 \times 32$ pixels. Because different people have very different writing styles, and there are breaks, blurring, and gaps in the written numerals, handwritten numbers can easily be confused. All the SSC algorithms yielded poor results when we used the original pixels. Therefore, prior to applying any of the algorithms, we applied the wavelet-HOG transform to the pixels and vectorized the data to 60 dimensions with PCA. To add difficulty to the experiments, we also added varying amounts of Gaussian noise. The results of 20 repeated experiments on the MNIST and USPS datasets with the same parameter settings show that the LpOSC algorithm achieves a lower error rate than do the SSC, OSC, SpatSC, and LRR algorithms on MNIST dataset (as Figure 7 shows) and on USPS dataset (as Figure 8 shows).

**FIGURE 8** Error rates under different Gaussian Attacks on USPS dataset

**FIGURE 9** Error rates under different Gaussian Attacks on Yale B dataset
5.2 | Face database experiment

We choose the Yale B (Georghiades, Belhumeur, & Kriegman, 2002) dataset as face clustering dataset; the facial images in this dataset include different poses and illumination types, and each image is 48 × 42 pixels. In each experiment, we randomly selected 20 pictures from the five categories of facial images and added noise; pictures from the same category were grouped together. From the extended Yale B face dataset, we selected five of the most frequent classes and added different amounts of Gaussian noise. Most of the algorithms, including SSC, OSC, and our algorithm, achieved good performances. However, our algorithm generally obtained the lowest clustering error (Figure 9). When the noise level is relatively low, our algorithm, SSC, and OSC all achieved error rates close to zero; when the noise level is greater than 0.8, SpatSC and OSC select sample representations from incorrect subspaces. Similarly, the block-diagonal structure of the matrix becomes poor, which reduces the segmentation precision. However, our algorithm and the SSC algorithm still achieve low error rates. When the noise level increases to 0.9, the error rate of our algorithm remains small whereas the other algorithms present much higher error rates.

**FIGURE 10** Error rates under different Gaussian Attacks on a video

**FIGURE 11** Error rates under different Gaussian Attacks on COIL-20 dataset
FIGURE 12  Average running time, in second, under different Gaussian Attacks on Yale B dataset

FIGURE 13  Average running time, in second, under different Gaussian Attacks on USPS dataset

FIGURE 14  Average running time, in second, under different Gaussian Attacks on MNIST dataset
5.3 Video segmentation experiments

Similar results were obtained in video segmentation experiments (Figure 10). A short video was obtained from the Internet open-video website (https://openvideo.org/). In general, our algorithm achieved the lowest error rate. As the noise levels increased, the error rates increased for all the algorithms. At Gaussian noise levels above 0.5, the SpatSC algorithm does not function well; when the noise level exceeds 0.6, the OSC and SSC algorithms begin to function poorly. In contrast, our algorithm maintains a low error rate.

5.4 Object recognition

To test the robustness of our algorithm at resisting rotation, we used COIL-20 (Nene, Nayar, & Murase, 1996). This challenging task usually requires a large amount of training, and unsupervised clustering methods often fail. The COIL-20 dataset contains 20 objects that are rotated by 5° horizontally each time. Therefore, each object is depicted in 72 images and is rotated in a complete circle. We selected the first seven objects and used all the images for those objects, which include toy models of ducklings, cars, and cats. For the self-rotation deformation of the images, the error rate of LpOSC is approximately 20%, whereas the error rates of SSC range from 40%–60%. These error rates mainly occur because of the preprocessing of the image dictionaries with good global and local properties. Figure 11 shows that our method successfully extracted the wavelet-HOG features from samples to form image dictionaries, which enables it to resist rotation of a certain object. Relative to the rotation,

![Figure 15: Average running time, in second, under different Gaussian Attacks on a video](image1)

**FIGURE 15** Average running time, in second, under different Gaussian Attacks on a video

![Figure 16: Average running time, in second, under different Gaussian Attacks on COIL-20 dataset](image2)

**FIGURE 16** Average running time, in second, under different Gaussian Attacks on COIL-20 dataset
Gaussian noise does not strongly affect the image. When the Gaussian noise variance increases from 0.1 to 0.3, the error rate of LpOSC increases by approximately 10%, but the algorithm still works.

5.5 Description of the running time

We use MATLAB to compare the average runtimes of all the tested algorithms (Figures 12–16). The algorithm runtimes increase with the number of samples and the feature dimension of the sample. The SSC method has the shortest runtime. The LpOSC requires the same amount of time as the OSC and the SptSC methods to calculate the similarity matrix, and it requires an average amount of time to complete a single subspace clustering because of the extra regulation item in the MNIST and COIL datasets. However, LpOSC's computational time for a single subspace clustering is slightly less than that of OSC and SptSC on the Yale B segmentation and USPS digit recognition experiments under different Gaussian noise levels. This occurs because the LpOSC algorithm uses smaller feature dimensions compared with the original pixels. In general, our algorithm performs better under severe noise conditions, though it may take longer. However, to apply it in real-time applications, distributed optimization techniques should be applied to improve the computational speed. Such speed improvements are an important future research direction.

6 CONCLUSION

In this paper, we present a novel algorithm that considers both the local and global features between adjacent samples in the time domain and the continuous characteristics amongst neighbouring pixels in the spatial domain. The affinity matrix is constructed using a wavelet-HOG dictionary to self-represent image samples, and the p-norm is integrated as the regular order term to achieve effective clustering for sequential data. Simultaneously, to emphasize both the differences between the various clusters and the consistency within individual clusters, we implement the denoising principal component selection method prior to the block-diagonal structure method of the coefficient matrix to increase the robustness of image sequence clustering. Experimental results obtained using a handwriting digit dataset, the extended Yale B face dataset, video segmentation data, and COIL clearly demonstrate the effectiveness of the proposed method. In the future, we will investigate how to use distributed optimization techniques to improve our method's efficiency.

ACKNOWLEDGEMENTS

The paper is the extended version of paper EDMA1 submitted to the workshop EDMA-2017. This research work is supported by National Natural Science Foundation of China under grant numbers 61672157, 61772134, and 41601477 and by Natural Science Foundation of Fujian Province under grant number 2018J01776 and by the Leading Program in the Science and Technology Department of Fujian Province under grant number 2015Y0054.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

ORCID

Liping Chen https://orcid.org/0000-0002-5135-4007
Gongde Guo https://orcid.org/0000-0002-5118-7261

REFERENCES

Bo, L., Ren, X., & Fox, D. (2010). Kernel Descriptors for Visual Recognition. Advances in Neural Information Processing Systems 23:, Conference on Neural Information Processing Systems 2010, Proceedings of A Meeting Held 6-9 December 2010, Vancouver, British Columbia, Canada (pp.244-252). DBLP. http://papers.nips.cc/paper/4147-kernel-descriptors-for-visual-recognition.pdf
Boult, T. E., & Brown, L. G. (1991). Factorization-based segmentation of motions. Proc IEEE Workshop on Visual Motion, 179–186. https://doi.org/10.1109/WVM.1991.212809
Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations & Trends in Machine Learning, 3(1), 1–122. https://doi.org/10.1561/2200000016
Bradley, P. S., & Mangasarian, O. L. (2000). K-plane clustering. Journal of Global Optimization, 16(1), 23–32. https://doi.org/10.1023/A:1008324625522
Chen, J., & Yang, J. (2014). Robust subspace segmentation via low-rank representation. IEEE Transactions on Cybernetics, 44(8), 1432–1445. https://doi.org/10.1109/TCYB.2013.2286106
Chen, L., Guo, G., & Chen, L. (2017). Smooth representation clustering based on Kernelized random walks. Web and Big Data. https://doi.org/10.1007/978-3-319-69781-9_1
Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vol.1, pp.886–893). IEEE. https://doi.org/10.1109/CVPR.2005.177.
Deng, C. X., Li, S., & Fu, Z. X. . (2010). The reproducing kernel Hilbert space based on wavelet transform. International Conference on Wavelet Analysis & Pattern Recognition. IEEE. https://doi.org/10.1109/ICWAPR.2010.5576389
Elhamifar, E., & Vidal, R. (2009). Sparse subspace clustering. IEEE Conference on Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2009.5206547
Elhamifar, E., & Vidal, R. (2013). Sparse subspace clustering: algorithm, theory, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(11), 2765–2781. https://doi.org/10.1109/TPAMI.2013.57

Feng, J., Lin, Z., Xu, H., & Yan, S. (2014). Robust subspace segmentation with block-diagonal prior. IEEE Conference on Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2014.482

Georgiades, A. S., Belhumeur, P. N., & Kriegman, D. J. (2002). From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis & Machine Intelligence, 23(6), 643–660. https://doi.org/10.1109/34.927464

Ho, J., Yang, M. H., Lim, J., Lee, K. C., & Kriegman, D. (2006). Clustering appearances of objects under varying illumination conditions. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2003.1211332

Hull, J. J. (2002). A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis & Machine Intelligence, 16(5), 550–554. https://doi.org/10.1109/34.291440

Lécun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Ma, Y., Yang, A. Y., Derksen, H., & Fossum, R. (2008). Estimation of subspace arrangements with applications in modeling and segmenting mixed data. SIAM Review, 50(3), 413–458. https://doi.org/10.1137/060655523

Mallat, S. G., & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. IEEE Trans on Signal Processing, 41(12), 3397–3415. https://doi.org/10.1109/78.258082

Nene, S. A., Nayar, S. K., & Murase, H. (1996). Columbia object image library (COIL-20). Technical Report CUCS-005-96.

Pham, D. S., Budhaditya, S., Phung, D., & Venkatesh, S. (2012). Improved subspace clustering via exploitation of spatial constraints. Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2012.6247720

Rao, S. R., Yang, A. Y., Sastry, S. S., & Ma, Y. (2010). Robust algebraic segmentation of mixed rigid-body and planar motions from two views. International Journal of Computer Vision, 88(3), 425–446. https://doi.org/10.1007/s11263-009-0314-1

Tierney, S., Gao, J., & Guo, Y. (2014). Subspace clustering for sequential data. Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2014.134

Tipping, M. E., & Bishop, C. M. (1999). Mixtures of principal component analyzers. Neural Computation, 11(2), 443. https://doi.org/10.1007/s11263-009-0043-8

Unser, M. (2015). Wavelets: On the virtues and applications of the mathematical microscope. Journal of Microscopy, 255(3), 123–127. https://doi.org/10.1111/jmi.12151

Vageeswaran, P., Mitra, K., & Chellappa, R. (2013). Blur and illumination robust face recognition via set theoretic characterization. IEEE Transactions on Image Processing, 22(4), 1362–1372. https://doi.org/10.1109/TIP.2012.2228498

Vidal, R., & Favaro, P. (2014). Low rank subspace clustering (lrs). Pattern Recognition Letters, 43(1), 47–61. https://doi.org/10.1101/essy1.12365

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. Science, 287(5456), 1273–1276. https://doi.org/10.1126/science.287.5456.1273

Wang, W. W., Xiao-Ping, L. I., Feng, X. C., & Wang, S. Q. (2015). A survey on sparse subspace clustering. Acta Automatica Sinica, 41(8), 1373–1384. https://doi.org/10.16383/j.ias.2015.140891

Wu, Y., Zhang, Z., Huang, T. S., & Lin, J. Y. (2001). Multibody grouping via orthogonal subspace decomposition. IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE. https://doi.org/10.1109/CVPR.2001.990968

Xin, B., Wang, Y., Gao, W., Wipf, D., & Xin, B. (2017). Building invariances into sparse subspace clustering. IEEE Transactions on Signal Processing, 66, 1–1. https://doi.org/10.1109/tsp.2017.2762279, 462.

Xu, J., Xu, K., Chen, K., & Ruan, J. (2015). Reweighted sparse subspace clustering. Computer Vision & Image Understanding, 138, 25–37. https://doi.org/10.1016/j.cviu.2015.04.003

**AUTHOR BIOGRAPHIES**

**Liping Chen** is an associate professor at Fujian Normal University. She received B.S. degree in Computer Science and M.S. degree in Applied Mathematics from Fujian Normal University. She is currently pursuing her PhD at Fujian Normal University, China. Her research interests include video analysis, subspace clustering, and data mining.

**Gongde Guo** is a professor at Fujian Normal University. He received the Bachelor of Engineering degree in computer software from Zhejiang University, China in 1985 and PhD degree in computer science from University of Ulster in 2004. He is currently working at School of Mathematics and Computer Science, Fujian Normal University as a professor. His research interests include intelligent computing and algorithms development, machine learning and data mining, and the application in data classification/categorisation and health food discrimination, with 130+ articles published.

**Hui Wang** is Professor of Computer Science at Ulster University and is visiting professor at Fujian Normal University. His research interests are machine learning, logics and reasoning, combinatorial data analytics, and their applications in image, video, spectra, and text analysis. He has over 200 publications in these areas.

---

How to cite this article: Chen L, Guo G, Wang H. An ordered sparse subspace clustering algorithm based on p-Norm. Expert Systems. 2021;38:e12368. https://doi.org/10.1111/exsy.12368