Smooth Representation Clustering

Han Hu, Zhouchen Lin, Jianjiang Feng and Jie Zhou
1State Key Laboratory on Intelligent Technology and Systems, TNList Department of Automation, Tsinghua University
2Key Lab. of Machine Perception, School of EECS, Peking University
huh04@mails.thu.edu.cn, zlin@pku.edu.cn, {jfeng,jzhou}@tsinghua.edu.cn

Abstract
Subspace clustering is a powerful technology for clustering data according to the underlying subspaces. Representation based methods are the most popular subspace clustering approach in recent years. In this paper, we analyze the grouping effect of representation based methods in depth. In particular, we introduce the enforced grouping effect conditions, which greatly facilitate the analysis of grouping effect. We further find that grouping effect is important for subspace clustering, which should be explicitly enforced in the data self-representation model, rather than implicitly implied by the model as in some prior work. Based on our analysis, we propose the SMooth Representation (SMR) model. We also propose a new affinity measure based on the grouping effect, which proves to be much more effective than the commonly used one. As a result, our SMR significantly outperforms the state-of-the-art ones on benchmark datasets.

1. Introduction
In many computer vision and machine learning problems, the data can be viewed as points drawn from multiple low-dimensional subspaces, with each subspace corresponding to one category or class, e.g., point trajectories of moving objects captured by an affine camera [24], images of several subjects under varying illumination or under different poses [6], and local patches or texture features of pixels/superpixels on an image [20]. A basic task for processing such kind of data is to cluster the points according to the underlying subspace. Such a task is called subspace clustering [26].

1.1. Related Work
Existing methods for subspace clustering can be roughly grouped into three categories: algebra based, statistics based, and spectral clustering based [26].

Most of the early studies on subspace clustering are algebra or statistics based. The two most well known algebraic methods are perhaps the shape interaction matrix (SIM) [2] and generalized principal component analysis (GPCA) [27]. Although with elegant formulations, in general the performance of these methods drops quickly in the presence of noise, degeneracy, or partially coupled subspaces. The statistics based methods treat subspace clustering as a mixed data inference problem and thus some popular methods from the more general statistical learning field can be used, such as random sample consensus (RANSAC) [5] and expectation maximization (EM) [12]. Although there have been several new techniques developed to improve the criterion (e.g., agglomerative lossy compression (ALC) [22]), model selection (e.g., Branch and Bound (BB) [10]), the performance of these methods is limited by their dependency on estimating the exact subspace models.

Many of the recent studies focus on the spectral clustering based methods [21, 30, 3, 16, 29, 19, 18, 17]. The major differences among these methods lie in the way they build the affinity matrix. A direct way is to compute affinity matrix from existing algebraic methods [21] or by defining a point-to-subspace or subspace-to-subspace distance metric [30]. More recently, many works apply the self-representation idea to compute affinities [3, 16, 29, 19, 18, 17], i.e., represent every sample by a linear combination of pixels/superpixels on an image [20]. A basic task for processing such kind of data is to cluster the points according to the underlying subspace. Such a task is called subspace clustering [26].

\[
\min_Z \alpha \|X - A(X)Z\|_l + \Omega(X, Z),
\]

\[
s.t. \quad Z \in C,
\]

where \(X \in \mathbb{R}^{d \times n}\) is the data matrix with each column being a sample vector, \(A(X)\) is a dictionary matrix which could be learnt or be simply set as \(A(X) = X\), \(\| \cdot \|_l\) is a proper norm, \(\Omega(X, Z)\) and \(C\) are the regularizer and constraint set on \(Z\), respectively, and \(\alpha > 0\) is a trade-off parameter. Then \(Z^*\) is used to compute an affinity matrix, e.g.,
∀ group effect if ∥CASS have the grouping effect. In contract, CASS have the grouping effect easily. Then we investigate why grouping effect is actually important in the representation model. In particular, we introduce the enforced grouping effect (EGE) conditions, which can help us identify the grouping effect easily. Then we investigate why grouping effect helps subspace clustering.

1.2. Contributions

Lu et al. [18, 17] discovered that Least Squares Regression (LSR) [18] and Correlation Adaptive Subspace Segmentation (CASS) [17] models both have the grouping effect defined as follows.

Definition 1 (Grouping Effect): Given a set of d-dimensional data points X = [x₁, ..., xₙ], ∈ Rᵈˣₙ, a self-representation matrix Z = [z₁, ..., zₙ], ∈ Rⁿˣⁿ has grouping effect if ∥xᵢ − xⱼ∥ᵢ → 0 ⇒ ∥zᵢ − zⱼ∥ᵢ → 0, ∀i ≠ j.

Inspired by [18, 17], we analyze the grouping effect of representation based method in depth. In particular, we introduce the enforced grouping effect (EGE) conditions, which can greatly facilitate the analysis of grouping effect of a representation based method. By the EGE conditions, we easily find new models that also have the grouping effect. In contrast, Lu et al. [18, 17] proved the grouping effect of LSR and CASS in a case-by-case way. Their proofs are specific and hence cannot be applied to other models.

We further find that grouping effect is actually important for subspace clustering. So we propose to explicitly enforce the grouping effect in the representation model. In contract, prior work [18, 17] only passively discovered that LSR and CASS have the grouping effect.

Finally, based on our analysis we propose the SMooth Representation (SMR) model. We also propose a novel affinity measure based on the grouping effect, which proves to be much more effective than the commonly used measure (∥Z⁺∥₁ + ∥Z⁺T∥₁)/2. Our experiments on benchmark datasets show that our SMR significantly outperforms the state-of-the-art approaches.

2. Grouping Effect

The grouping effect was first explicitly stated by Lu et al. [18, 17], who showed that in LSR and CASS when the samples are close to each other their representation coefficients are also close to each other. Their proofs are specific for LSR and CASS and cannot be applied to other models. In this section, we analyze the grouping effect of reconstruction based models in depth. We first introduce the Enforced Grouping Effect (EGE) conditions, which can help us identify the grouping effect easily. Then we investigate why grouping effect helps subspace clustering.

2.1. Enforced Grouping Effect Conditions

We introduce general sufficient conditions for the grouping effect as follows.

Definition 2 (Enforced Grouping Effect Conditions): The EGE conditions on problem (1) are:

1. A(X) is continuous with respect to X and Ω(X, Z) is continuous with respect to X and Z ∈ C;
2. Problem (1) has a unique solution Z⁺ and Z⁺ is not an isolated point of C;
3. Z ∈ C if and only if ZP ∈ C, and Ω(X, Z) = Ω(XP, ZP), for all permutation matrix P.
4. A(XP) = A(X)P, Z ∈ C if and only if PT ZP ∈ C, and Ω(X, Z) = Ω(XP, PT ZP), for all permutation matrix P.

Then we have the following lemma.

Lemma 1: If Problem (1) satisfies the EGE conditions (1) and (2), then its optimal solution Z⁺ is a continuous function of X.

Proof: It is obvious that Z⁺ can be regarded as a function of X according to EGE condition (2). In the following, we prove the continuity of Z⁺ w.r.t X.

Suppose Z⁺ is discontinuous with respect to X and X = X₁ is a discontinuity point. We have: ∃ε₁ > 0, ∀δ₁ > 0, there exist ∥X₂ − X₁∥₁ < δ₁ and ∥Z⁺₂ − Z⁺₁∥₁ > ε₁.

Denote f(X, Z) = ∥X − A(X)Z∥₁ + αΩ(X, Z). Since Problem (1) has a unique solution, we have ∃ε₂ > 0, f(X₂, Z⁺₂) < f(X₂, Z⁺) − ε₂. According to EGE condition (1), f(X, Z) is continuous with respect to X: for any ε₃ > 0, there exists δ₂ > 0 such that for all ∥X − X₂∥₁ < δ₂ ⇒ |f(X, Z⁺₂) − f(X₂, Z⁺₂)| < ε₃ and there exists some number δ₃ > 0 such that for all ∥X − X₂∥₁ < δ₃ ⇒ |f(X, Z⁺) − f(X₂, Z⁺)| < ε₃.
Suppose $2\varepsilon_3 < \varepsilon_2$, $\delta_1 \leq \delta_2$, and $\delta_1 \leq \delta_3$. We have

\[
\begin{align*}
  f(X_1, Z_2^2) &< f(X_2, Z_2^2) + \varepsilon_3 \\
  &< f(X_2, Z_1^2) + \varepsilon_3 - \varepsilon_2 \\
  &< f(X_1, Z_1^2) + 2\varepsilon_3 - \varepsilon_2 \\
  &< f(X_1, Z_1^2).
\end{align*}
\] (2)

Eq. (2) indicates that $Z_2^2$ is a better solution of Problem (1) when $X = X_1$, which is a contradiction. Hence the continuity of $Z^*$ w.r.t. $X$ is proved. □

Then we have the following proposition.

**Proposition 1:** The optimal solution $Z^*$ to problem (1) has grouping effect if EGE conditions (1), (2), and (3) are satisfied.

**Proof:** We first instantiate $X$ by $X_1$. Consider two sufficiently close points $x_i$ and $x_j$ in $X_1$. For simplicity we informally write $|a - b|_F \to 0$ to denote that $a$ and $b$ are close to each other. Exchanging the two columns $x_i$ and $x_j$, we get a new data matrix $X_2 = X_1 P_{ij}$, where $P_{ij}$ is the permutation matrix by exchanging the $i^{th}$ and $j^{th}$ columns of the identity matrix. It is obvious that $\|X_2 - X_1\|_F \to 0$ and $A(X_2) - A(X_1)|_F \to 0$.

Given EGE condition (3), it is easy to check that $Z_2^2 = Z_1^2 P$ is the unique optimal solution of problem (1) when $X = X_2$. By Lemma 1, we have that $\|X_2 - X_1\|_F \to 0 \Rightarrow \|Z_2 - Z_1\|_F \to 0$. Therefore, $\|z_i - z_j\|_2 \to 0$ as $Z_2^2$ and $Z_1^2$ only differ in the $i^{th}$ and $j^{th}$ columns.

We now check the grouping effect of existing representation based methods listed in Table 1 by the above EGE conditions. SSC [3, 4] does not satisfy the conditions. Indeed, it does not have the grouping effect. For example, considering $X = [x_1, x_2, x_3]$ with $x_1 = x_2 = x_3$, any permutation matrix satisfying $\text{diag}(Z) = 0$ would be the optimal solution to SSC. For LSR, all EGE conditions are satisfied. So it has the grouping effect. Figures 1(a)-(c) also exemplify our above observations. For LRR [16, 15], it is obvious that EGE conditions (1) and (3) are satisfied. The uniqueness of the optimal solution to LRR can also be proven, as stated in Proposition 2. Hence, LRR has the grouping effect. The optimal solution of CASS [17] is unique according to [8], and it also has the grouping effect.

**Proposition 2:** LRR has a unique optimal solution.

**Proof:** Please find it in the supplementary material. □

Proposition 1 not only provides a way to determine the grouping effect of existing methods, it may also help us to design new methods with grouping effect. For example, the following families of methods have grouping effects.

**Proposition 3:** Problems (1) with the following $\Omega(Z)$ and $C$ have grouping effect:

\[
\begin{align*}
  (1) \, \Omega(Z) &\equiv \sum_{j=1}^{n} \left( \sum_{i=1}^{n} |Z_{ij}|^p \right)^{q}, \, p > 1, \, q \geq \frac{1}{p}, \, C = \emptyset. \\
  (2) \, \Omega(Z) &\equiv \text{tr}(Z^THZ^T)^p, \, H > 0, \, p \geq 1/2, \, C = \emptyset.
\end{align*}
\]

**Proof:** We put it in the supplementary material. □

It should be noted that the constraint set $C = \{Z_{ii} = 0, \forall i\}$, as used by some existing methods, such as SSC [4], SSQP [29], MSR [19] and LSR-Z [18], does not satisfy the EGE condition (3). Accordingly, these methods do not have the grouping effect in a strict sense. However, these methods also perform well. So we generalize the concept of grouping effect as follows in order to comply with such an observation.

**Definition 3 (Permutated Grouping Effect):** Given a set of data points $X = [x_1, \ldots, x_n] \in \mathbb{R}^{d \times n}$, a self-representation matrix $Z = [z_1, \ldots, z_n] \in \mathbb{R}^{n \times n}$ has permuted grouping effect if $\|x_i - x_j\|_2 \to 0 \Rightarrow \|P^T_{ij}z_i - z_j\|_2 \to 0$, where $P^T_{ij}$ is the permutation matrix by exchanging the $i^{th}$ and the $j^{th}$ rows of the identity matrix.

Then we have the following proposition.

**Proposition 4:** The optimal solution $Z^*$ to problem (1) has permuted grouping effect if EGE conditions (1), (2), and (4) are satisfied.

**Proof:** Similar to the proof of Proposition 1, we form a new data matrix $X_2 = X_1 P_{ij}$. By EGE conditions (2) and (4), $Z_2^2 = P^T_{ij}Z_1^2P_{ij}$ is the unique optimal solution to problem (1) when $X = X_2$. If $\|x_i - x_j\|_F \to 0$, then by Lemma 1 we have $\|Z_2^2 - Z_1^2\|_F \to 0$, implying $\|P^T_{ij}z_i - z_j\|_2 \to 0$. □

One may check that the constraint set $C = \{Z_{ii} = 0, \forall i\}$ satisfy the EGE condition (4). So by Proposition 4, SSQP [29], MSR [19] and LSR-Z [18] have the permuted grouping effect.

**2.2. Why Grouping Effect?**

It was claimed by Lu et al. [18] that the effectiveness of LSR comes from the grouping effect. However, in [18] there is no convincing evidence to support this claim. In this section, we provide two viewpoints to advocate this property for representation based methods.

We first analyze the grouping effect from the viewpoint of optimization. The first term in Problem (1) penalizes the reconstruction error, which can be regarded as a first-order energy encoding the whole subspace structure of the data. The grouping effect of a representation matrix indicates that $\Omega(Z)$ and $C$ in Problem (1) must include a second-order energy to penalize the discontinuities in the representation coefficients, either implicitly or explicitly. With this second-order energy, the representation will be stable. On the other hand, spatially close data points may help each other to prevent over-fitting in reconstruction the samples. For example, in Figure 1(a), we consider the faces marked by the green and purple squares. They are very close in appearance but with a large part shadowed, hence violating the subspace
3. Smooth Representation Clustering

In this section, based on the detailed analysis on grouping effect, we propose a novel subspace clustering method, called Smooth Representation (SMR) clustering. We first introduce how to explicitly enforce grouping effect in the representation model, then present the SMR model.

3.1. Enforcing Grouping Effect

As stated in Section 2.1, LRR and LSR utilize the grouping effect implicitly. The grouping effect can be understood as the smooth dependence of feature on the sample. We may write the regularization term of LSR as follows:

\[ \Omega(Z) = \text{tr}(ZZ^T) \]

\[ = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|z_i - z_j\|_2^2 + \frac{1}{n} \|Z^T e\|_2^2, \tag{3} \]

where \( e \) is the all ones vector. It can be viewed as assigning equal weights to all pairs of representations, regardless of whether the representations are close to each other or not.

By the analysis in Section 2, we should enforce the grouping effect explicitly by the affinity of samples. One possibility is adopting the following regularization term:

\[ \Omega(Z) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \|z_i - z_j\|_2^2 \]

\[ = \text{tr}(ZLZ^T), \tag{4} \]

where \( W = (w_{ij}) \) is the weight matrix measuring the spatial closeness of points and \( L = D - W \) is the Laplacian matrix, in which \( D \) is the diagonal degree matrix with \( D_{ii} = \sum_{j=1}^{n} w_{ij} \). A common way to construct \( W \) is to use the \( k \) nearest neighbor (k-nn) graph with a heat kernel or 0-1 weights [9]. We find that the simple 0-1 weighted \( k \)-nn graph performs well enough in our experiments, as exemplified in Figure 1(d). So we use the 0-1 weighted \( k \)-nn graph (a default value of \( k \) is 4) in all our experiments. There are also more complex graphs, e.g., affinity graphs produced by other subspace clustering algorithms. However, usually we did not observe evident improvements over the \( k \)-nn graph.

3.2. Smooth Representation

To avoid numerical instability issue, we enforce \( L \) to be strictly positive definite by adding a \( \epsilon I \) instead, where \( I \) is the identity matrix and \( 0 < \epsilon < 1 \). A default value of \( \epsilon \) is 0.01. Then we get our smooth representation model:

\[ \min_Z f(Z) = \alpha \|X - XZ\|_F^2 + \text{tr}(ZL^T), \tag{5} \]
Problem (5) is a smooth convex program. Differentiating the objective function \( f(Z) \) with respect to \( Z \) and setting it to zero, we get the optimal solution \( Z^* \) satisfies
\[
\alpha X^T XZ^* + Z^* \bar{L} = \alpha X^T X.
\] (6)

The above equation is a standard Sylvester equation \([1]\). It has a unique solution.

**Proposition 6:** The Sylvester equation (6) has a unique solution.

**Proof:** \( X^T X \) is positive semi-definite. So all of its eigenvalues are nonnegative: \( \lambda_i \geq 0, \forall i \). \( \bar{L} \) is positive definite. So all of its eigenvalues are positive \( \mu_j > 0, \forall j \). Hence, for any eigenvalues of \( X^T X \) and \( \bar{L} \), \( \lambda_i + \mu_j > 0 \). According to \([14]\), the Sylvester equation (6) has a unique solution. □

A classical algorithm for the Sylvester equation is the Bartels-Stewart algorithm \([1]\), which consists of transforming the coefficient matrices into Schur forms by QR decomposition, and then solving the resulting triangular system via back-substitution. The algorithm has a computational complexity of \( O(n^3) \).

The solution to Problem (5) also has several nice properties according to Proposition 7.

**Proposition 7:** The solution to Problem (5) has the following properties: (1) it has grouping effect; (2) it is block diagonal when the subspaces are independent and the data is noise free.

**Proof:** (1) We can easily check that Problem (5) satisfies EGE conditions (1), (2), and (3). According to Proposition 1, its solution has the grouping effect.

(2) When the columns of the data matrix \( X \) is permuted by any permutation matrix \( P \), we have \( \bar{L}(XP) = P^T \bar{L}(X)P \). Hence, \( \Omega(ZP) = \text{tr}(ZP \bar{L}(XP)P^T Z^T) = \text{tr}(Z \bar{L}(X)Z^T) = \Omega(Z) \). Denote \( Z^D = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \), where \( A \) and \( D \) are from \( Z = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \). Substituting \( Z \) and \( Z^D \) into Equation (4), we have \( \Omega(Z) \geq \Omega(Z^D) \), where the equality holds if and only if \( B = C = 0 \), and \( \Omega(Z^D) = \Omega(A) + \Omega(D) \). So \( \Omega(Z) \) satisfies the Enforced Block Diagonal Condition \([18]\). According to Theorem 2 in \([18]\), its optimal solution is block diagonal when the subspaces are independent and the data is noise free. □

Furthermore, benefiting from the strengthening of grouping effect, the representations derived by SMR usually have much larger gap between the within-class and the between-class distances than those by LRR and LSR, as illustrated in qualitatively Figure 1(d) and quantitatively Figure 2. This property implies the SMR can derive more salient within-class affinities with regard to the between-class ones than the methods when using the measures described in Section 3.3.

![Figure 2. Comparing gaps between the within-class (wc.) and between-classes (bc.) representation distances of LRR, LSR and SMR on the USPS handwritten digit dataset. For each digit image, the 5% closest images in the wc. and bc. sets respectively are selected for illustration. We can see that SMR has much larger gaps than LRR and LSR.](image)

### 3.3. Subspace Clustering by SMR

After obtaining the self-representation matrix \( Z^* \), a common way for subspace clustering is to define an affinity matrix as
\[
J_1 = (\|Z^*\| + \|Z^{*T}\|)/2
\] (7)
and use the spectral clustering algorithm \([23]\) to produce the final clustering results, as has been used by SSC, LRR and LSR.

The effectiveness of \( J_1 \) mainly comes from the block diagonal property of \( Z^* \), leaving the nice property of grouping effect unexploited. To exploit the merit of grouping effect, we define a new affinity matrix as
\[
J_2 = \left( \frac{z_i^T z_j^*}{\|x_i\| \|x_j\|} \right)^\gamma,
\] (8)
where \( \gamma > 0 \) is used to control the affinity variances. The new affinity measure can be regarded as the inner product of the new representation vectors normalized by the norms of their original features. The normalization prevents the affinity measure from biased by the original feature amplitudes, which is very common in the motion segmentation problem whose trajectories usually vary a lot in amplitude. Figure 3 shows the affinity matrices of the two measures based on \( Z^* \) derived by SMR. It can be seen that \( J_2 \) strengthens the affinities within each cluster and weakens them across clusters.

The whole procedure of subspace clustering by SMR is summarized in Algorithm 1.
Algorithm 1 Subspace Clustering by SMooth Representation (SMR)

Require: Data points $X = [x_1, \cdots, x_n] \in \mathbb{R}^{d \times n}$, the number of subspaces $m$

1: Build a $k$-nn graph $W$ and compute the corresponding Laplacian matrix $L$.
2: Solve the Sylvester equation (6) by the Bartels-Stewart algorithm to get a representation matrix $Z^*$.
3: Compute affinity matrix by either (7) or (8).
4: Use spectral clustering algorithm to obtain $m$ clusters.

4. Experiments

In this section, we apply our SMR\footnote{Codes available at https://sites.google.com/site/hanlushomepage/} to three applications of subspace clustering: motion segmentation, face clustering, and handwritten digit clustering. We also compare SMR with representative reconstruction based methods, such as SSC, LRR and LSR, whose performances are state-of-the-art.

4.1. Datasets and Evaluation Metric

We use three datasets for our experiments: Hopkins155 [25], Extended Yale Face B [7] and USPS [11], which are the most popular benchmark datasets used in the literature for evaluating subspace clustering algorithms. For all the algorithms, the best results are reported.

Hopkins155 [25] is a motion segmentation dataset, consisting of 155 video sequences with extracted feature points and their tracks across frames. See Figure 4 for some sample sequences. We use PCA to project the data into a 12-dimensional subspace. The same as in most literatures, for each algorithm, we use the same parameters for all sequences [26].

Extended Yale Face B [7] is a face clustering dataset, which consists of $192 \times 168$ pixel cropped face images, under varying poses and illuminations, from 38 human subjects. We use all the 64 frontal face images of the first 10 subjects, and resize the images to $32 \times 32$. We also use PCA to project the data into a $10 \times 6$-dimensional subspace.

USPS [11] is a handwritten digit dataset of 9298 images, with each image having $16 \times 16$ pixels. We use the first 100 images of each digit for experiments.

The same as in most literatures, we use clustering error ($CE$) to measure the accuracy [25]. $CE$ is the minimum error by matching the clustering result and the ground truth under the optimal permutation, formally defined as:

$$CE = 1 - \frac{1}{N} \sum_{i=1}^{N} \delta(p_i, \text{map}(q_i)),$$

where $q_i$, $p_i$ represent the output label and the ground truth one of the $i$th point; $\delta(x, y) = 1$ if $x = y$, and $\delta(x, y) = 0$ otherwise; map($q_i$) is the best mapping function that permutes clustering labels to match the ground truth labels and it can be efficiently computed by the Kuhn-Munkres algorithm [13].

4.2. Experimental Results

For fair comparison, we adopt the Frobenius norm for the reconstruction error term for all the algorithms. Table 2 lists the motion segmentation errors of the four methods on the Hopkins155 datasets using the typical affinity measure (7). SMR achieves a clustering error of 2.27\%, while the best result of other algorithms is 3.01\% by LSR. It should be noted that the numbers in Table 2 are different from those listed in [18] because they used an approximate computation of $CE$, which is also observed by [4]. Noting that most sequences are easy to be segmented and hence all the algorithms get zero errors on them, the performance improvement by SMR over others is significant. The computational costs of all the algorithms are also listed in Table 2. SMR is a bit slower than LSR but much faster than SSC and LRR.

Table 2. Clustering errors ($CE$) using affinity measure (7) and the computation times on Hopkins155 datasets. The computation time includes only the computation of $Z^*$ in Problem (1).

<table>
<thead>
<tr>
<th>method</th>
<th>SSC</th>
<th>LRR</th>
<th>LSR</th>
<th>SMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$CE$ (%)</td>
<td>3.90</td>
<td>4.11</td>
<td>3.01</td>
<td>2.27</td>
</tr>
<tr>
<td>time (s)</td>
<td>2.50</td>
<td>2.03</td>
<td><strong>0.12</strong></td>
<td>0.40</td>
</tr>
</tbody>
</table>
Table 3. Clustering errors (CE) using affinity measure (8) with $\gamma = 1$ on Hopkins155 datasets.

<table>
<thead>
<tr>
<th>method</th>
<th>SSC</th>
<th>LRR</th>
<th>LSR</th>
<th>SMR</th>
<th>SMR(J_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE (%)</td>
<td>6.27</td>
<td>2.83</td>
<td>1.90</td>
<td>1.13</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. The performance of SMR with varying $\alpha$ on Hopkins155 datasets (4-nn graph is used).

As stated in Section 3.3, the affinity measure (8) is better in exploiting the grouping effect. So we also use affinity measure (8) for experiments and report the clustering errors in Table 3. It can be seen that the performances of LRR, LSR and SMR are significantly improved, and SMR get the minimum segmentation error with 1.13%. The SSC with the new affinity measure (8) performs worse than using the traditional affinity (7). These results support the use of affinity measure (8) rather than (7) for subspace clustering when the self-representation model has grouping effect.

We also test the performance of SMR with varying parameters $\alpha$ (in the objective function of (1)) and $k$ (for constructing the $k$-nn graph). The results are shown in Figure 5 and Figure 6. SMR performs very stably with varying $\alpha$ and the number $k$ of neighborhood. Since other algorithms rely on only one parameter $\alpha$, to be fair we use a 4-nn graph for all our experiments without tuning $k$ on different datasets.

Tables 4 and 5 show the clustering errors on Extended Yale Face B and USPS datasets, respectively. To make fair comparison, we use the traditional affinity measure (7) in all the algorithms. The number of our algorithm using the new affinity (8) is also shown in the last column for reference. It can be seen that SMR outperforms the others significantly, especially on the USPS datasets. We also illustrate the affinity matrices using (7) in Figure 7, where we can clearly observe better grouping effect from SMR than from the others. When the affinity (8) $J_2$ is used, the performances are further improved. For example, we achieve 3.75% on Extended Yale Face B.

Table 4. Clustering errors (CE) on Extended Yale Face B datasets.

<table>
<thead>
<tr>
<th>method</th>
<th>SSC</th>
<th>LRR</th>
<th>LSR</th>
<th>SMR</th>
<th>SMR(J_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE (%)</td>
<td>48.81</td>
<td>35.00</td>
<td>27.50</td>
<td>26.56</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Clustering errors (CE) on USPS datasets.

<table>
<thead>
<tr>
<th>method</th>
<th>SSC</th>
<th>LRR</th>
<th>LSR</th>
<th>SMR</th>
<th>SMR(J_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE (%)</td>
<td>43.10</td>
<td>22.60</td>
<td>26.10</td>
<td>12.70</td>
<td>11.20</td>
</tr>
</tbody>
</table>

5. Conclusions and Future Work

In this paper, we analyze the grouping effect of representation based methods in depth. We introduce Enforced Grouping Effect conditions to verify whether a representation based model has the grouping effect in a systematic manner. We also provide insights to the importance of grouping effect for subspace clustering. Based on our detailed analysis, we propose a novel subspace clustering model, Smooth Representation, to explicitly enforce the grouping effect in the model. We further propose a novel affinity measure that better utilizes the grouping effect among representation coefficients. Extensive experiments on benchmark datasets testify to the great advantage of SMR over the state-of-the-art subspace clustering methods. In the future, we plan to utilize the grouping effect in a wider scope, e.g., semi-supervised learning.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grants 61225008, 61020106004 and 61373074, the National Basic Research Program of China under Grant 2014CB349304, the Ministry of Education of China under Grant 2012002110033, and the Tsinghua University Initiative Scientific Research Program. Zhouchen Lin is supported by NSF China (grant nos. 61272341, 61231002, 61121002) and MSRA.
Figure 7. Affinity matrices derived by SSC, LRR, LSR, and SMR on USPS datasets using (7). The affinities are normalized by $0.6 \cdot \max(Z^*)$ to have better view. The grouping effect of SMR is much more salient than those of others.

References


