NON-RIGID 3D SHAPE RECOGNITION VIA DICTIONARY LEARNING

Yin Zhou†, *Kai Liu*‡, *Kenneth E. Barner*†

†University of Delaware, Newark, DE, USA, 19716

‡School of Electrical Engineering and Information, Sichuan University, China

ABSTRACT

Non-rigid 3D shape recognition is an important and challenging research topic in computer vision and pattern recognition. This paper presents a novel algorithm, called dictionary learning based on supervised locally linear representation (DL-SLLR), for efficient 3D shape recognition using shape descriptors. Specifically, we introduce a novel locality-preservation error term along with a label approximation error term into the objective function. The proposed algorithm optimizes a dictionary for its capability in representation as well as its locality-preservation capability, which thus allows more consistent encoding of similar descriptors compared with sparse coding. In addition, the proposed SLLR coding yields a closed-form solution, compared to many sparse coding algorithms. Experimental results demonstrate that using majority voting, DL-SLLR outperforms D-KSVD and SVM over a newly generated SLI 3D Face Dataset and the SHREC'11 Contest Dataset.

***Index Terms*—** Shape recognition, Point cloud classification, Dictionary learning, Classification, Sparse Coding

1. INTRODUCTION

Accurately recognizing non-rigid 3D objects in real world has been a challenging topic in machine/computer-vision-based applications such as robotic control, surveillance, automatic navigation, assistive technology, etc [1]. To achieve this objective, effective feature extraction strategies and discriminative classification algorithms are much needed.

In order to extract robust features from 3D surface of non-rigid objects, many algorithms have been proposed. Typically, they can be categorized into global feature extraction, e.g., shape histograms [2], shape moments [3], spherical harmonics [4], etc, and local feature extraction, e.g., heat kernel signatures [5], meshSIFT [6], 3D SURF [1], etc. Experiments have demonstrated that the local feature based methods have obvious advantages for dealing with issues of noise and partial occlusion [1, 7]. In this paper, we employ meshSIFT [6] algorithm to build 3D shape descriptors.

Once the features of an object are extracted, an effective classification algorithm is desired to identify the class of the object. Among those proposed classification methods so far, we mainly investigate dictionary learning based approaches. Sparse coding solves for a compact and representational dictionary such that the large amount of training data can be expressed linearly by a few atoms in the dictionary. It has been proven that this model is effective in image restoration [8,9], image denoising [10,11], image classification [12–14], etc. In particular, Zhang *et al.* [14] proposed Discriminative-KSVD (D-KSVD) for face recognition by introducing label information into an objective function such that a representational dictionary and a linear classifier can be jointly optimized by using KSVD [10]. However, sparse coding based algorithms do not exploit the dependency information among local features and may therefore yield inconsistent representations of similar features [15, 16].

Contact: {zhouyin, barner}@udel.edu, kailiu@scu.edu.cn
This work is supported by NSF under Grant No. 0812458.

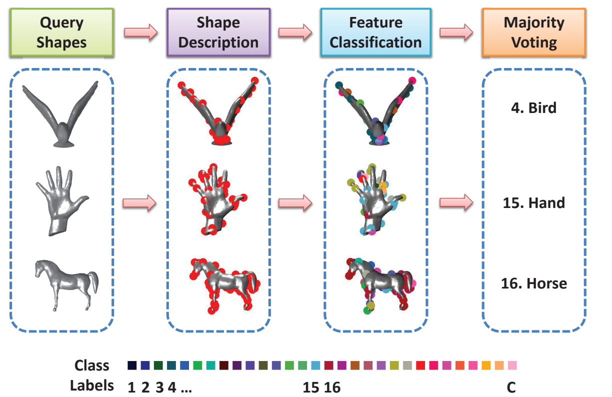


Fig. 1: The proposed classification strategy. Given a query shape S, extract shape descriptors on it and then perform classification per descriptor. Finally the label of S is determined by majority voting over descriptor decisions.

Locality-based coding was recently developed [17, 18] to address this issue. Particularly, Yu *et al.* [17] introduced Locally Coordinate Coding (LCC) to approximately express nonlinear functions as a linear combination of anchor points. Unfortunately, their coding strategy is based on a modification of LASSO (Least Absolute Shrinkage and Selection Operator) and hence suffers from high computation cost. Wang *et al.* [18] further proposed Locality-constrained Linear Coding (LLC) as a fast approximation to LCC achieving impressive performance in image classification by using LLC codes as features and Support Vector Machine (SVM) as classifier. Nevertheless, little efforts has been made to apply the aforementioned sparse or local coding techniques to non-rigid 3D shape recognition.

In this paper, we propose a novel algorithm, called dictionary learning based on supervised locally linear representation (DL-SLLR) for efficient 3D shape recognition. The main contribution is explicitly incorporating a locality-preservation error term and the label approximation error term into the objective function. Unlike sparse coding based algorithms [13, 14, 19–22], the proposed SLLR coding yields a closed-form solution. Moreover, the dictionary is optimized for both reconstruction and locality preservation, which therefore allows not only faithful reconstruction but also more consistent encoding of similar descriptors [15]. The proposed DL-SLLR is also different from recently proposed locality-based coding algorithms [17, 18, 23] in that 1) The SLLR coding is supervised such that training descriptors can only be coded by its same-class neighboring atoms, which thus yields a more discriminative dictionary; 2) A simple yet effective linear mapping is explicitly formulated into the unified objective function for classification.

To classify a query shape, we aggregate the predicted results of all descriptors using majority voting. Such a scheme requires negligible computational complexity and is invariant to rigid (rotation, scaling, and shift) and non-rigid (e.g., stretch, shrink and twist) transformations. Experiments over a newly generated SLI 3D Face Dataset and the SHREC'11 Contest Dataset validate the effectiveness of the proposed framework, *i.e.,* DL-SLLR in conjunction with majority voting.

2. DICTIONARY LEARNING

Consider a C-label 3D shape classification problem. Let $\mathbf{Y}_i \in \mathbb{R}^{m\times n_i}$ be a set of m-dimensional n_i shape descriptors extracted from 3D objects with label i. Assign label i to all descriptors in $\mathbf{Y}_i$. Set $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_C] \in \mathbb{R}^{m\times N}$ as the training set for all classes, where $N = \sum_{i=1}^{C} n_i$.

Let $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C] \in \mathbb{R}^{m\times L}$ be a structured dictionary, where each $\mathbf{D}_i \in \mathbb{R}^{m\times K}$ is a class-specific sub-dictionary trained for $\mathbf{Y}_i$ and $L = KC$. Denote $\mathbf{x}_j \in \mathbb{R}^L$ as the sparse code of $\mathbf{y}_j$ over $\mathbf{D}$, where $\mathbf{y}_j \in \mathbf{Y}$ is the j-th descriptor in $\mathbf{Y}$, for $j = 1, \ldots, N$. We define $\boldsymbol{\Omega}_k(\mathbf{y}_j)$ as the same-class neighborhood with respect to $\mathbf{y}_j$ containing k-nearest-neighbor atoms from one particular sub-dictionary, which is pertaining to the same label as $\mathbf{y}_j$. Correspondingly, define $\boldsymbol{\Lambda}_{\mathbf{d}_i} \triangleq \{\mathbf{y}_j |\ \forall j,\ x_{ij} \neq 0,\ \mathbf{y}_j \in \mathbf{Y}\}$ as a neighborhood of $\mathbf{d}_i$, containing all $\mathbf{y}_j$ that are concurrently selecting $\mathbf{d}_i$ as one of their neighboring atoms, where x_{ij} is the i-th element in vector $\mathbf{x}_j$.

The goal at hand is achieving two objectives. The first is establishing a discriminative dictionary structured as $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C]$ such that each $\mathbf{D}_i$ is independently trained for $\mathbf{Y}_i$ and every atom $\mathbf{d}_i$ preserves the locality of its neighborhood $\boldsymbol{\Lambda}_{\mathbf{d}_i}$. The second objective is realizing a linear mapping $\mathbf{W} \in \mathbb{R}^{C\times L}$ that transforms the sparse code $\mathbf{x}_j$ of every descriptor $\mathbf{y}_j$ to its label vector $\mathbf{h}_j = [0, \ldots, 1, \ldots, 0]^{\mathrm{T}}$, where the index of element 1 indicates the label of $\mathbf{y}_j$. Thus, the dictionary learning problem (DL-SLLR) is formalized as:

$$\min_{\mathbf{D},\mathbf{W},\mathbf{X}} \quad \begin{pmatrix} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_F^2 + \alpha \sum_{i=1}^{K} \sum_{\mathbf{y}_j \in \boldsymbol{\Lambda}_{\mathbf{d}_i}} \|\mathbf{y}_j - \mathbf{d}_i\|_2^2 \\ +\beta \|\mathbf{H}-\mathbf{W}\mathbf{X}\|_F^2 + \gamma \|\mathbf{X}\|_F^2 + \mu \|\mathbf{W}\|_F^2 \end{pmatrix} \tag{1}$$

$$\text{s.t.} \quad x_{ij} = 0 \quad \text{if} \ \ \mathbf{d}_i \notin \boldsymbol{\Omega}_k(\mathbf{y}_j) \tag{$*$}$$

$$\mathbf{1}^{\mathrm{T}}\mathbf{x}_j = 1 \quad \forall i, j \tag{$**$}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{L\times N}$ contains the SLLR codes (discussed in the next section) for descriptors in $\mathbf{Y} \in \mathbb{R}^{m\times N}$, and x_{ij} is the i-th element in column vector $\mathbf{x}_j \in \mathbf{X}$. Obeying the standard meaning, the first and the third terms represent the reconstruction and the label approximation errors respectively. The second term is the novel locality-preservation error term, which ensures that every atom is close to those training samples that concurrently choose it as one of their neighboring atoms. It therefore encourages atom consistency in local representations of similar descriptors [15]. Note that $\gamma \|\mathbf{X}\|_F^2$ and $\mu \|\mathbf{W}\|_F^2$ are regularization penalty terms, included for numerical stability, with γ and μ small positive constants. In addition, the constraint $(*)$ requires each descriptor to be reconstructed only by its same-class neighboring atoms, ensuring that every sub-dictionary $\mathbf{D}_i$ is trained from the $\mathbf{Y}_i$ independently. The constraint $(**)$ allows the coding to be shift-invariant, in which $\mathbf{1}$ is a column vector of all ones.

3. OPTIMIZATION

The dictionary learning problem can be solved by iteratively repeating the following two steps to reduce the objective function, *i.e.,* first solving for the code matrix $\mathbf{X}$ with the other two variables fixed, and then updating $\mathbf{D}$ and $\mathbf{W}$, respectively. The iterations are terminated if either the objective function value is below some preset threshold or a maximum number of iterations has been reached.

3.1. Supervised Locally Linear Representation (SLLR)

Consider first solving for the SLLR code $\mathbf{x}_j \in \mathbf{X}$, for all $j = 1, \ldots, N$, with $\mathbf{D}$, $\mathbf{W}$ fixed. Define $\hat{\mathbf{y}}_j = \left[\mathbf{y}_j^{\mathrm{T}}, \sqrt{\beta}\mathbf{h}_j^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{m+C}$ as the j-th augmented training sample in the augmented training set $\hat{\mathbf{Y}} = \left[\mathbf{Y}^{\mathrm{T}}, \sqrt{\beta}\mathbf{H}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{(m+C)\times N}$. Likewise denote $\hat{\mathbf{D}} = \left[\mathbf{D}^{\mathrm{T}}, \sqrt{\beta}\mathbf{W}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{R}^{(m+C)\times L}$ as the augmented dictionary. Furthermore, set $\hat{\boldsymbol{\Omega}}_k(\mathbf{y}_j) = \{\hat{\mathbf{d}}_i |\ \forall i,\ \mathbf{d}_i \in \boldsymbol{\Omega}_k(\mathbf{y}_j),\ \hat{\mathbf{d}}_i \in \hat{\mathbf{D}}\}$ as the augmented neighborhood with respect to $\mathbf{y}_j$, with $\hat{\mathbf{d}}_i$ being the i-th column in $\hat{\mathbf{D}}$. Thus, minimizing Eq. (1) with respect to $\mathbf{x}_j$, is equivalent to solving the following locally linear representation (LLR) problem [24, 25] under the same-class neighborhood constraint.

$$\begin{aligned} \min_{\mathbf{x}_j} \quad & \|\hat{\mathbf{y}}_j - \hat{\mathbf{D}}\mathbf{x}_j\|_2^2 + \gamma \|\mathbf{x}_j\|_2^2 \\ \text{s.t.} \quad & x_{ij} = 0 \quad \text{if} \ \ \hat{\mathbf{d}}_i \notin \hat{\boldsymbol{\Omega}}_k(\mathbf{y}_j) \\ & \mathbf{1}^{\mathrm{T}}\mathbf{x}_j = 1 \quad \forall i \end{aligned} \tag{2}$$

Taking both of the constraints into consideration simultaneously, and using Lagrange multiplier, we get

$$\mathcal{J}(\tilde{\mathbf{x}}_j, \eta) = \|\hat{\mathbf{y}}_j - \hat{\boldsymbol{\Omega}}_k \tilde{\mathbf{x}}_j\|_2^2 + \gamma \|\tilde{\mathbf{x}}_j\|_2^2 + \eta \left(\mathbf{1}^{\mathrm{T}}\tilde{\mathbf{x}}_j - 1\right) \tag{3}$$

where for simplicity we express $\hat{\boldsymbol{\Omega}}_k(\mathbf{y}_j)$ as $\hat{\boldsymbol{\Omega}}_k \in \mathbb{R}^{(m+C)\times k}$, and $\tilde{\mathbf{x}}_j$ is a succinct vector containing only the nonzero coefficients for those $\hat{\mathbf{d}}_i \in \hat{\boldsymbol{\Omega}}_k(\mathbf{y}_j)$. Denote as $\mathbf{G} = (\hat{\boldsymbol{\Omega}}_k - \hat{\mathbf{y}}_j\mathbf{1}^{\mathrm{T}})^{\mathrm{T}}(\hat{\boldsymbol{\Omega}}_k - \hat{\mathbf{y}}_j\mathbf{1}^{\mathrm{T}})$ the local covariance matrix. Then Eq. (3) can be written as:

$$\mathcal{J}(\tilde{\mathbf{x}}_j, \eta) = \tilde{\mathbf{x}}_j^{\mathrm{T}}(\mathbf{G} + \gamma\mathbf{I})\tilde{\mathbf{x}}_j + \eta \left(\mathbf{1}^{\mathrm{T}}\tilde{\mathbf{x}}_j - 1\right) \tag{4}$$

where $\mathbf{I}$ is the identity matrix. Setting $\nabla_{\tilde{\mathbf{x}}_j}\mathcal{J}(\tilde{\mathbf{x}}_j, \eta)$ and $\nabla_{\eta}\mathcal{J}(\tilde{\mathbf{x}}_j, \eta)$ to zero yields the desired closed-form solution, as

$$\tilde{\mathbf{x}}_j = \frac{(\mathbf{G} + \gamma\mathbf{I})^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}(\mathbf{G} + \gamma\mathbf{I})^{-1}\mathbf{1}} \tag{5}$$

As suggested by [24], a more efficient way to compute $\tilde{\mathbf{x}}_j$ is by first solving the linear system of equations $(\mathbf{G} + \gamma\mathbf{I})\tilde{\mathbf{x}}_j = \mathbf{1}$ and then normalizing $\tilde{\mathbf{x}}_j$ to satisfy the sum-to-one constraint. We adopt this for practical implementation.

Note that the proposed SLLR is different from LLR [24, 25] in that 1) SLLR coding is supervised, which yields discriminative local reconstruction coefficients; 2) SLLR is performed over a compact dictionary and is combined with dictionary optimization, which in turn helps further reduce the reconstruction error.

3.2. Updating the Dictionary and the Mapping

Next, consider the update of $\mathbf{D}$ and $\mathbf{W}$, with $\mathbf{X}$ fixed. We individually optimize each atom of $\mathbf{D}$. Let $\mathbf{d}_i \in \mathbb{R}^m$ be the i-th atom in $\mathbf{D}$ and define $\mathbf{x}_{i*} \in \mathbb{R}^{1\times N}$ as the i-th row of $\mathbf{X}$. With $\mathbf{X}$ and the other atoms fixed, we rewrite Eq. (1) and cast the optimization problem with respect to $\mathbf{d}_i$ as

$$\min_{\mathbf{d}_i} \ \mathcal{H}(\mathbf{d}_i) = \left\| \mathbf{Y} - \sum_{l \neq i} \mathbf{d}_l \mathbf{x}_{l*} - \mathbf{d}_i \mathbf{x}_{i*} \right\|_F^2 + \alpha \left\{ \sum_{\mathbf{y}_j \in \mathbf{\Lambda}_{\mathbf{d}_i}} \|\mathbf{y}_j - \mathbf{d}_i\|_2^2 + \sum_{l \neq i} \sum_{\mathbf{y}_j \in \mathbf{\Lambda}_{\mathbf{d}_l}} \|\mathbf{y}_j - \mathbf{d}_l\|_2^2 \right\} \tag{6}$$

Letting $\mathbf{E} = \mathbf{Y} - \sum_{l \neq i} \mathbf{d}_l \mathbf{x}_{l*}$ and rearranging Eq. (6), we have

$$\min_{\mathbf{d}_i} \ \mathcal{H}(\mathbf{d}_i) = \mathrm{Tr}\left\{ (\mathbf{E} - \mathbf{d}_i \mathbf{x}_{i*})(\mathbf{E} - \mathbf{d}_i \mathbf{x}_{i*})^{\mathrm{T}} \right\} + \alpha \sum_{\mathbf{y}_j \in \mathbf{\Lambda}_{\mathbf{d}_i}} \left[(\mathbf{y}_j - \mathbf{d}_i)^{\mathrm{T}} (\mathbf{y}_j - \mathbf{d}_i) \right] \tag{7}$$

Note that $\mathcal{H}(\mathbf{d}_i)$ is convex. Hence, setting the gradient of $\mathcal{H}(\mathbf{d}_j)$ with respect to $\mathbf{d}_i$ to zero yields the updated atom $\mathbf{d}_i^{new}$ as

$$\mathbf{d}_i^{new} = \frac{1}{(\mathbf{x}_{i*}\mathbf{x}_{i*}^{\mathrm{T}} + \alpha |\mathbf{\Lambda}_{\mathbf{d}_i}|)} \left(\mathbf{E}\mathbf{x}_{i*}^{\mathrm{T}} + \alpha \sum_{\mathbf{y}_j \in \mathbf{\Lambda}_{\mathbf{d}_i}} \mathbf{y}_j \right) \tag{8}$$

where $|\mathbf{\Lambda}_{\mathbf{d}_i}|$ denotes the cardinality of set $\mathbf{\Lambda}_{\mathbf{d}_i}$. Applying Eq. (8) to all $\mathbf{d}_i$, for $i = 1, \ldots, L$, completes the dictionary update in the current iteration.

In order to update $\mathbf{W}$, we solve the multivariate ridge regression [26] problem as

$$\mathbf{W}^{new} = \arg\min_{\mathbf{W}} \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_F^2 + \mu \|\mathbf{W}\|_F^2 \tag{9}$$

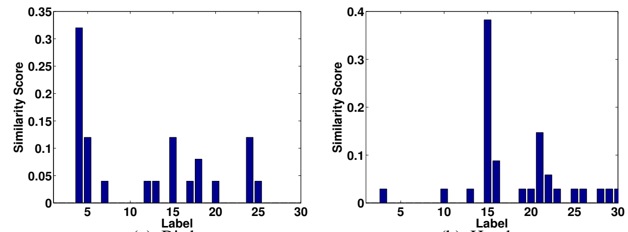
where μ is a small positive constant for numerical stability. The solution is easily obtained as

$$\mathbf{W}^{new} = \mathbf{H}\mathbf{X}^{\mathrm{T}}(\mathbf{X}\mathbf{X}^{\mathrm{T}} + \mu \mathbf{I})^{-1} \tag{10}$$

Minimizing the objective function, we will obtain the optimal dictionary $\mathbf{D}$, which is representational for reconstructing training descriptors and capable of preserving locality of the data manifold, and get the optimal mapping $\mathbf{W}$ in approximating the label matrix.

4. CLASSIFICATION

Human can distinguish different classes of objects (see Fig. 1) with mutual similarity in shape, even without using the clue of size, color and texture because we can make judgements based on seeking and comparing the most distinctive shape characteristics among those objects, despite the presence of a large portion of mutual similarity [27]. In other words, it is the most distinctive shape features of an object that plays the critical role in a successful recognition. We propose to emulate this process by employing majority voting and apply it to classifying 3D objects based on the newly proposed dictionary learning algorithm. Simplistically, we may assume that the votes from nondistinctive shape descriptors are approximately evenly spread across similar classes. Thus the outcome is the class that wins the most votes from the distinctive descriptors. Majority voting is an aggregation process (in which we need no explicit knowledge about which descriptors are distinctive or not) and its result is determined by the highest accumulated votes on a particular class. Visualizing the vote distribution of two objects from the



(a) Bird (b) Hand

Fig. 2: Majority voting results after normalization on SHREC'11 Contest Dataset. The two objects are bird (a) and hand (b). The bird is associated to label 4 while the hand is associated to label 15. Since the number of extracted descriptors varies across different objects, we normalize the voting results for better visualization.

SHREC'11 Contest Dataset [28], we can see in Fig. 2 that although a large portion of votes go to incorrect classes, the true class clearly receives the highest number of votes compared with any incorrect class.

Given a query object S, denote $\mathbf{Q}_S = [\mathbf{q}_1, \ldots, \mathbf{q}_n] \in \mathbb{R}^{m \times n}$ as the set of n extracted shape descriptors. The local reconstruction code $\mathbf{x}_j$ for each $\mathbf{q}_j$ is computed by solving

$$\begin{aligned} \min_{\mathbf{x}_j} \quad & \|\mathbf{q}_j - \mathbf{D}\mathbf{x}_j\|_2^2 + \gamma \|\mathbf{x}_j\|_2^2 \\ \text{s.t.} \quad & x_{ij} = 0 \qquad \text{if } \ \mathbf{d}_i \notin \mathbf{\Gamma}_t(\mathbf{q}_j) \ \ \forall i \\ & \mathbf{1}^{\mathrm{T}}\mathbf{x}_j = 1 \end{aligned} \tag{11}$$

where x_{ij} is the i-th element in vector $\mathbf{x}_j \in \mathbb{R}^L$ and $\mathbf{\Gamma}_t(\mathbf{q}_j)$ is a neighborhood set consisting of t nearest-neighbor atoms of $\mathbf{q}_j$. The solution is given previously as Eq. (5).

Next, compute the projection $\mathbf{r} = \mathbf{W}\mathbf{x}_j \in \mathbb{R}^C$ and assign the label l_j to descriptor $\mathbf{q}_j$ according to

$$l_j = \arg\max_i (\mathbf{r} = [r_1, \ldots, r_i, \ldots, r_C]^{\mathrm{T}}) \tag{12}$$

Applying the same procedure to all $\mathbf{q}_j \in \mathbf{Q}$, a label vector is formed as $\mathbf{l} = [l_1, l_2, \ldots, l_n]$. Finally, we count the votes for each class label based on $\mathbf{l}$ and classify the query shape S according to the label receiving the most votes.

5. EXPERIMENT

The proposed DL-SLLR algorithm is evaluated using majority voting as classification scheme over two large datasets, the SLI 3D Face Dataset and the SHREC'11 Contest Dataset [28][1]. The proposed method is compared with D-KSVD [14] using majority voting and with a baseline SVM [29] method with Gaussian kernel using bag-of-words histogram (BoWH + SVM). The shape descriptors are extracted using meshSIFT [6]. Training parameters for DL-SLLR are $k \in \{2, 3, 4, 5\}$, $\alpha = \beta = 0.01$, $\gamma = \mu = 0.001$ over both datasets. The neighborhood size t for classification is set to 10 and 6 for the face and the SHREC'11 datasets respectively.

5.1. SLI 3D Face Dataset

First presented are classification results over a newly generated Structured Light Illumination 3D Face Dataset (SLI 3D Face Dataset) [30]. This dataset is collected using the algorithm and hardware implementation developed in [31, 32]. It contains 576

[1] Accessible at:http://www.itl.nist.gov/iad/vug/sharp/contest/2011/NonRigid/data.html

(a) Neutral (b) Sad (c) Happy (d) Anger

(e) Right 45° (f) Up-front 0° (g) Left 45°

Fig. 3: Examples for SLI 3D Face Dataset. Top row are four expressions at angle $0°$; Bottom row are manually cropped face area with normal expression under three view angles.

Table 1: Recognition results on SLI 3D Face Dataset.

Method	Proposed	D-KSVD [14]	BoWH + SVM [29]	Smeet's [6]
Accuracy	**96.00%**	95.78%	90.63%	91.67%

Table 2: Computation time for classifying one query object on SLI 3D Face Dataset.

Method	Proposed	D-KSVD [14]
Computation time	0.0776s	0.5068s

high-quality dense 3D point clouds (approximately 5000 points per face) for 24 subjects with 4 static facial expressions under 3 different view angles. The population of 24 volunteers consists of 7 females and 17 males. Data for each individual is collected over two recording sessions in a dark room. During each session, an individual is required to face the camera at 3 different angles, *i.e.,* $\pm 45°$ (frontal right/left) and $0°$ (up-front), while at each angle performing 4 kinds of static facial expressions, *i.e.,* neutral, sad, happy, and anger.

Preprocessing the point clouds for classification, we use the depth information to segment subjects from the background (top row of Fig. 3) and then manually crop the face area for each subject with a 3D bounding box (bottom row of Fig. 3). We employ the same subset of the database as [30] for evaluation. The total number of meshSIFT descriptors extracted from training faces is approximately 70,000. For DL-SLLR and D-KSVD, a dictionary of $L = 4800$ atoms is trained for classification, *i.e.,* $K = 200$ atoms per class. The results are reported based on 4-fold cross-validation over a repetitions. As shown in Table 1, the proposed approach outperforms other competitive methods yielding the highest recognition rate of 96.00%. The proposed DL-SLLR is also compared with D-KSVD [14] in terms of the average computation time for classifying one 3D face. As shown in Table 2, DL-SLLR is approximately 6 time faster than D-KSVD.

5.2. SHREC'11 Contest Dataset

The SHREC'11 Contest Dataset [28] consists of 600 non-rigid 3D objects from 30 classes represented as watertight triangle meshes, including alien, horse, lamp, etc., as shown in Fig. 4 and Fig. 5. Each class equally has 20 objects. The total number of shape descriptors extracted from training objects is approximately 380,000. For DL-SLLR and D-KSVD, a classification dictionary of $L = 6000$ atoms is trained, *i.e.,* $K = 200$ atoms per class. We conduct 10-fold cross-validation over the entire dataset and report averaged recognition results over 20 repetitions. As shown in Table 5, the proposed DL-SLLR with majority voting achieves the highest recognition rate of 99.67%. In addition, DL-SLLR is compared with D-KSVD [14] in terms of the average computation time for classifying one query object. As shown in Table 4, the proposed algorithm is approximately 5 times faster than D-KSVD. Finally, we study the robustness of aforementioned methods against to partial occlusions. Fig. 6 shows the performance of the methods under the conditions of varying percentage of occlusion. Clearly, the proposed approach (DL-SLLR in conjunction with majority) outperforms other methods.

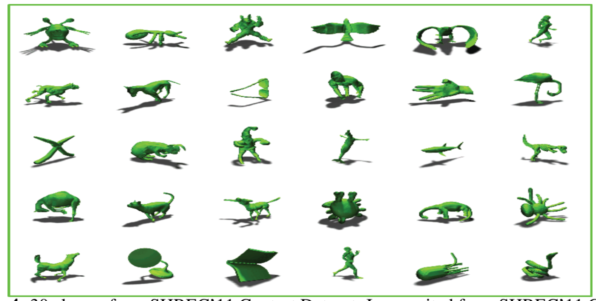


Fig. 4: 30 classes from SHREC'11 Contest Dataset. Image cited from SHREC'11 Contest website.

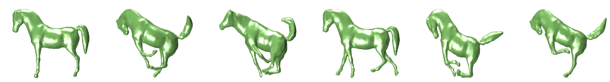


Fig. 5: 3D Nonrigid shapes from object class horse.

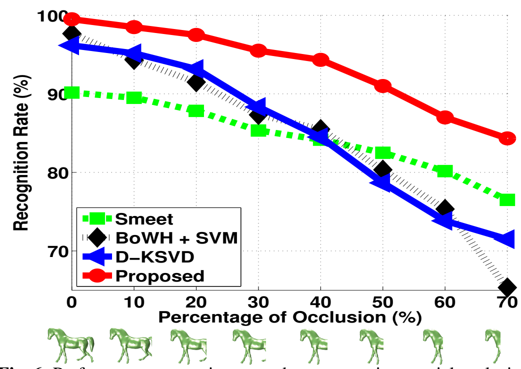


Fig. 6: Performance comparison on robustness against partial occlusion.

Table 3: Recognition results on SHREC'11 Contest Dataset.

Method	Proposed	D-KSVD [14]	BoWH + SVM [29]	Smeet's [28]
Accuracy	**99.67%**	96.67%	98.00%	90.00%

Table 4: Computation time for classifying one query object on SHREC'11 Contest Dataset.

Method	Proposed	D-KSVD [14]
Computation time	0.0122s	0.0697s

6. CONCLUSION AND FUTURE WORK

We present a novel dictionary learning algorithm (DL-SLLR) for 3D shape recognition. The main contribution is integrating the locality-preservation error term and the label approximation error term into the objective function for discriminative dictionary learning. The dictionary is learned in such a way as to be simultaneously representational and locality-preserving. Moreover, the proposed SLLR coding yields a closed-form solution. Experimental results show that the proposed DL-SLLR in conjunction with majority voting, achieves impressive classification performance over two large-scale 3D datasets and outperforms state-of-the-art methods, *i.e.,* D-KSVD and SVM. Future work includes 1) the refinement of coding strategy by combining nearest neighbor search with sparse coding; 2) more systematic evaluation of the proposed method using other shape descriptors and over more standard datasets; 3) benchmarking our newly generated SLI 3D Face Dataset and contribute it to the research community.

7. REFERENCES

[1] Jan Knopp, Mukta Prasad, Geert Willems, Radu Timofte, and Luc Van Gool, "Hough transform and 3d surf for robust three dimensional classification," in *ECCV'10*.

[2] Robert Osada, Thomas Funkhouser, Bernard Chazelle, and David Dobkin, "Shape distributions," *ACM Trans. Graph.*, vol. 21, no. 4.

[3] Dietmar Saupe and Dejan V. Vranic, "3d model retrieval with spherical harmonics and moments," in *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*, 2001.

[4] Michael Kazhdan, Thomas Funkhouser, and Szymon Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3d shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*.

[5] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proceedings of the Symposium on Geometry Processing*, 2009.

[6] C. Maes, T. Fabry, J. Keustermans, D. Smeets, P. Suetens, and D. Vandermeulen, "Feature detection on 3d face surfaces for pose normalisation and recognition," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, sept. 2010, pp. 1 –6.

[7] Hien Van Nguyen and F. Porikli, "Concentric ring signature descriptor for 3d objects," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, sept. 2011, pp. 2893 –2896.

[8] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE TIP*, pp. 53 –69, 2008.

[9] Haichao Zhang, Jianchao Yang, Yanning Zhang, N.M. Nasrabadi, and T.S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in *ICCV 2011*, pp. 770 –777.

[10] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionarys for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311 – 4322, 11 2006.

[11] K. Engan, S.O. Aase, and J.H. Husoy, "Frame based signal compression using method of optimal directions (mod)," in *ISCAS '99*.

[12] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. PAMI*, vol. 31, no. 2, pp. 210 – 227, 2 2009.

[13] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *CVPR 2010.*

[14] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," *CVPR 2010.*

[15] Shenghua Gao, I.W. Tsang, Liang-Tien Chia, and Peilin Zhao, "Local features are not lonely: Laplacian sparse coding for image classification," in *CVPR 2010*.

[16] K. Kavukcuoglu, M.A. Ranzato, R. Fergus, and Yann Le-Cun, "Learning invariant features through topographic filter maps," in *CVPR 2009*.

[17] Kai Yu, Tong Zhang, and Yihong Gong, "Nonlinear learning using local coordinate coding," in *NIPS09*.

[18] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, T. Huang, and Yihong Gong, "Locality-constrained linear coding for image classification," in *CVPR 2010.*

[19] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman, "Supervised dictionary learning," in *In Adv. NIPS 2008*.

[20] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *CVPR 2008*, june 2008.

[21] Zhuolin Jiang, Zhe Lin, and L.S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," in *CVPR 2011.*

[22] Yi-Chen Chen, C.S. Sastry, V.M. Patel, P.J. Phillips, and R. Chellappa, "Rotation invariant simultaneous clustering and dictionary learning," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, march 2012, pp. 1053 –1056.

[23] Y. Zhou and K. E. Barner, "Locality constrained dictionary learning for nonlinear dimensionality reduction," *Signal Processing Letters, IEEE*, vol. 20, no. 4, pp. 335–338, April.

[24] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323 – 2326, 2000.

[25] Lijun Zhang, Chun Chen, Jiajun Bu, Deng Cai, Xiaofei He, and T.S. Huang, "Active learning based on locally linear reconstruction," *PAMI, IEEE Trans.*, vol. 33, no. 10, pp. 2026 –2038, 2011.

[26] Gene H. Golub, Per Christian Hansen, and Dianne P. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 1, pp. 185–194, Oct. 1999.

[27] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948 –1962, nov. 2006.

[28] "SHREC'11 track: shape retrieval on non-rigid 3D watertight meshes," in *Proc. Eurographics 2011 Workshop on 3D Object Retrieval (3DOR'11)*, 2011, pp. 79–88.

[29] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines, 2001. available at http://www.csie.ntu.edu.tw/cjlin/libsvm," .

[30] Yin Zhou, Kai Liu, Jinglun Gao, and Kenneth E. Barner, "High-speed structured light scanning system and 3d gestural point cloud recognition," in *Proceedings, IEEE Int. Conf. on Information Sciences and Systems*, Baltimore, USA, May 2013, to appear.

[31] Kai Liu, Yongchang Wang, Daniel L. Lau, Qi Hao, and Laurence G. Hassebrook, "Dual-frequency pattern scheme for high-speed 3-d shape measurement," *Opt. Express*, vol. 18, no. 5, pp. 5229–5244, Mar 2010.

[32] Kai Liu, Yongchang Wang, Daniel L. Lau, Qi Hao, and Laurence G. Hassebrook, "Gamma model and its analysis for phase measuring profilometry," *J. Opt. Soc. Am. A*, vol. 27, no. 3, pp. 553–562, Mar 2010.