BoSCC: Bag of Spatial Context Correlations for Spatially Enhanced 3D Shape Representation

Zhizhong Han, Zhenbao Liu, *Member, IEEE*, Chi-Man Vong, *Senior Member, IEEE*, Yu-Shen Liu, Shuhui Bu, *Member, IEEE*, Junwei Han, *Senior Member, IEEE*, and C. L. Philip Chen, *Fellow, IEEE*

***Abstract*— Highly discriminative 3D shape representations can be formed by encoding the spatial relationship among virtual words into the *Bag of Words* (BoW) method. To achieve this challenging task, several unresolved issues in the encoding procedure must be overcome for 3D shapes, including: 1) *arbitrary mesh resolution*; 2) *irregular vertex topology*; 3) *orientation ambiguity on the 3D surface*; and 4) *invariance to rigid and non-rigid shape transformations*. In this paper, a novel spatially enhanced 3D shape representation called *bag of spatial context correlations* (BoSCCs) is proposed to address all these issues. Adopting a novel local perspective, BoSCC is able to describe a 3D shape by an occurrence frequency histogram of spatial context correlation patterns, which makes BoSCC become more compact and discriminative than previous global perspective-based methods. Specifically, the spatial context correlation is proposed to simultaneously encode the geometric and spatial information of a 3D local region by the correlation among spatial contexts of vertices in that region, which effectively resolves the aforementioned issues. The spatial context of each vertex is modeled by Markov chains in a multi-scale manner, which thoroughly captures the spatial relationship by the transition probabilities of intra-virtual words and the ones of inter-virtual words. The high discriminability and compactness of BoSCC are effective for classification and retrieval, especially in the scenarios of limited samples and partial shape retrieval. Experimental results show that BoSCC outperforms the state-of-the-art spatially enhanced BoW methods in three common applications: global shape retrieval, shape classification, and partial shape retrieval.**

***Index Terms*— Bag of spatial context correlations, spatial context correlation, spatial context, 3D shape representations.**

Manuscript received March 12, 2016; revised November 21, 2016 and February 17, 2017; accepted May 9, 2017. Date of publication May 16, 2017; date of current version June 7, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61672430, Grant 61573284, Grant 61522207, Grant 61472202, and Grant 61272229, in part by the NWPU Basic Research Fund under Grant 3102016JKBJJGZ08, in part by the University of Macau under Grant MYRG2016-00134-FST and Grant MYRG2014-00083-FST, and in part by FDCT Macau under Grant 050/2015/A. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Chunming Li. *(Corresponding author: Yu-Shen Liu.)*

Z. Han is with Tsinghua University, Beijing, China, and also with Northwestern Polytechnical University, Xi'an 710072, China (e-mail: h312h@mail.nwpu.edu.cn).

Z. Liu, S. Bu, and J. Han are with Northwestern Polytechnical University, Xi'an 710072, China (e-mail: liuzhenbao@nwpu.edu.cn; bushuhui@nwpu.edu.cn; jhan@nwpu.edu.cn).

C.-M. Vong is with the Department of Computer and Information Science, University of Macau, Macau 99999, China (e-mail: cmvong@umac.mo).

Y.-S. Liu is with Tsinghua University, Beijing 100084, China (e-mail: liuyushen@tsinghua.edu.cn).

C. L. P. Chen is with the Faculty of Science and Technology, University of Macau, Macau 99999, China (e-mail: philip.chen@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2017.2704426

I. Introduction

BAG OF WORDS (BoW) method is originally developed for natural language processing [1] and document classification [2]. With BoW, a text, e.g. a sentence or a document, is represented as an occurrence frequency histogram of words in a dictionary, as shown in Fig. 1 (a). Regarding 2D images or 3D shapes as documents in computer vision, BoW representations have also been used to facilitate understanding 2D [3]–[9] or 3D [10]–[14] shapes based on "virtual words", as illustrated in Fig. 1 (b). Through clustering over a set of local features (e.g., features around 2D pixels and features around 3D vertices or faces), cluster centers are obtained and considered as *the types of virtual words*. Subsequently, each local feature is assigned to its nearest type of virtual word in order to compute the occurrence frequency of each type of virtual words in a shape. However, the BoW method disregards the spatial relationship among words or virtual words, which is in fact an important information source to enhance the discriminability of BoW representations. Thus, effective spatially enhanced BoW methods are highly demanded in 3D shape analysis, where the spatial information is regarded as a key point. However, it is a very challenging task to encode the spatial relationship among virtual words for 3D shapes. The main unresolved issues in the encoding procedure for 3D shapes include i) arbitrary mesh resolution, ii) irregular vertex topology, iii) orientation ambiguity on the 3D surface and iv) invariance to various rigid and non-rigid shape transformations, such as translation, rotation, and articulation.

An intuitive idea of encoding the spatial information into BoW representations for 2D images resorts to the absolute positional relationship, such as "above" [15], "top-left" [16] or the relationships described by coordinates in 2D grids [17]–[20]. However, the absolute positional relationship is not rotation-invariant, which makes the encoded spatial information unsuitable for discriminating 3D shapes. Therefore, one class of approaches for 3D shapes encodes the relative positional relationship between local BoW representations. Similar to the spatial pyramid matching for 2D images [17], [20], local BoW representations are first calculated from segmented local 3D regions, such as patches [21], [22], regions between concentric spheres centered at the barycenter [23], and regions between intrinsic isocontours on articulated shapes [24]. Then, the relative positional relationship is encoded by concatenating local BoW representations in a consistent order [23], [24], or the pairwise Euclidean distance [21], [23] and the neighboring indica-

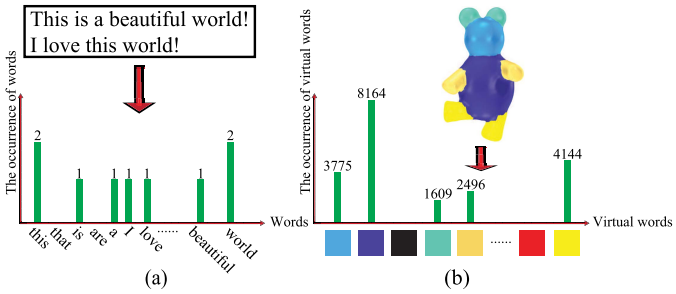

Fig. 1. Illustration of BoW for texts and 3D shapes. (a) BoW for texts. (b) BoW for 3D shapes.

tor [22] between local BoW representations. Nevertheless, these approaches suffer from a common problem, that is, only coarse spatial information can be encoded. This is because the spatial information in each segmented local region is neglected under local BoW representations. To encode more detailed spatial information, another class of approaches employs the pairwise distance between local features or virtual words, such as the geodesic or diffusion distance between local features on 3D articulated shapes [25], [26] or the Euclidean distance between local features and the barycenter of artificial 3D shapes [27]. However, the pairwise distance may cause spatial ambiguity, which fails to encode the spatial relationship among multiple pairs of virtual words. Furthermore, these approaches in both classes adopt the global perspective to directly encode the global spatial information. This global perspective results in high dimensional 3D shape representations but with coarse spatial information. Under this kind of representation, the performance of subsequent classification or partial shape retrieval is significantly affected. Therefore, for various applications such as shape retrieval and classification, it remains a research challenge to produce a compact (i.e. lower dimensional yet sufficiently informative) 3D shape representation, that is encoded with the spatial information invariant to rigid and non-rigid shape transformations.

To achieve this challenging requirement, we propose a novel spatially enhanced 3D shape representation, named as *Bag of Spatial Context Correlations* (BoSCC). Unlike the global perspective adopted in [23], [24], and [26], BoSCC employs a novel local perspective which indirectly describes the global spatial information by the local spatial information. Specifically, *spatial context correlation* is proposed to simultaneously encode the geometric and spatial information of a local region via correlation among spatial contexts of vertices, where the spatial context of each vertex is modelled as Markov chains [28] in a multi-scale manner. With the help of the novel spatial context correlation, BoSCC is formed by an occurrence frequency histogram of spatial context correlation patterns, which captures more detailed spatial information in a more compact way than other state-of-the-art methods. The resulting BoSCC representation is compact and with high discriminability for classification and retrieval, especially significant for training of classifiers with limited samples and partial shape retrieval. Finally, the significant contributions of our work are summarized as follows:

i) The Markov chain is introduced to model the spatial context around a vertex, which effectively overcomes the obstacles of arbitrary mesh resolution, irregular vertex topology and orientation ambiguity on 3D surface.
ii) The spatial context correlation is proposed to simultaneously encode the geometric and spatial information of 3D local regions.
iii) BoSCC is proposed as a spatially enhanced 3D shape representation, which is compact and with high discriminability for shape classification and retrieval. The compactness of BoSCC is especially helpful when training classifiers with limited available samples.
iv) A novel local perspective is introduced to indirectly encode the global spatial information by spatial context correlation patterns of local regions. In this way, BoSCC becomes more suitable for partial shape retrieval than other global perspective based methods.

The paper is organized as follows. Section II presents the related work of BoW and spatially enhanced BoW in 3D domain. The core techniques of BoSCC including spatial context modelling and the computation of spatial context correlation are detailed in Section III. Experimental setup and results with analysis are described in Section IV. Finally, the conclusion is drawn in Section V.

II. Related Work

In this section, three categories of related work are briefly reviewed, including i) BoW for 3D shapes, ii) BoW with encoded spatial information for 3D shapes and iii) spatially enhanced techniques for 2D images.

A. BoW for 3D Shapes

The BoW method offers a basic framework to represent a 3D shape using local features of vertices, faces or regions. Based on virtual words from a learned virtual dictionary, each local feature is assigned to its nearest type of virtual word whose index labels the corresponding vertex, face or region. Consequently, the BoW representation of a 3D shape is formed by an occurrence frequency histogram of all types of virtual words.

Recently, the BoW method has been employed for 3D shape recognition and retrieval [11], [23], [29]–[31]. The main difference among these studies lies in the different local features used for learning the virtual dictionary. For example, Toldo *et al.* [32] employ different geometry descriptors to characterize regions, Ohbuchi *et al.* [30] and Lian *et al.* [31] adopt SIFT features to describe projections of 3D local regions, Liu *et al.* [29] and Li and Godil [23] use the spin image [33] as the features of sampled vertices. With the powerful spectral descriptors such as the Heat Kernel Signature [34] and the Local Spectral Descriptor [10], the BoW method obtains better results in [10] and [35]. To enable efficient fusion of different types of local features and modalities for learning the virtual dictionary, Tabia *et al.* [14] introduce the covariance matrix to fuse local features in a local region. Since covariance matrices lie on Riemannian manifolds, the geodesic distance

on Riemannian manifolds is further derived to measure the similarity between covariance matrices.

B. Spatially Enhanced BoW for 3D Shapes

The disadvantage of BoW method is the disregard of spatial relationships among virtual words. To remedy this disadvantage, one class of approaches employs the spatial relationship among local BoW representations. Similar to spatial pyramid matching [17], local BoW representations are computed from local 3D regions, such as patches [21], regions between concentric spheres centered at the barycenter [23], and regions between intrinsic isocontours for articulated shapes [24].

In the concentric BoW method [23], concentric spheres are adopted as the spherical harmonic descriptor [36] to segment a shape into different regions, followed by spatial relationship encoding between segmented regions. First, a 3D shape is partitioned by several concentric spheres centered at its barycenter. Then, the 3D shape is represented by a concatenation of local BoW representations, and a local BoW representation is computed from the region between each pair of neighboring spheres. Although the encoded spatial information is invariant to rigid shape transformation (e.g., translation and rotation), it cannot resist non-rigid shape transformation (e.g., bending of arm). This is because non-rigid shape transformation changes the distribution of virtual words between the neighboring spheres. The same issue also exists in the spatially enhanced BoW method [21]. Although the patches used for computing the local BoW representations are directly segmented on 3D shape surfaces, the spatial relationship among these patches is still not invariant to non-rigid shape transformation due to its encoding under pairwise Euclidean distances. To resolve this issue, Intrinsic Spatial Pyramid Matching (ISPM) [24] uses isocontours of the second eigenfunction of the Laplace-Beltrami operator [37] to cut the articulated shapes into regions. The spatial information encoded by ISPM is invariant to non-rigid shape transformation, but it is only suitable for articulated shapes. This is because the rigid shape cannot be consistently segmented under the isocontours employed in [24].

In a nutshell, the main problem of aforementioned approaches is that the spatial information in each segmented local region is disregarded by the local BoW representation, leading to unsatisfactory discriminability.

To encode more detailed spatial information, spatial sensitive Bag-of-Features (SSBoF) employs the pairwise geodesic or heat diffused distance between local features, which is invariant to rigid and non-rigid shape transformations [25], [26]. SSBoF provides a simple 3D shape representation which sums over all pairwise distance-weighted features. However, the spatial information is encoded merely by the pairwise geodesic or diffusion distances, which is ambiguous and restricted. The spatial ambiguity is mainly caused by uncertain relative location between two virtual words. This is because the location of one virtual word determined by a specific distance to another virtual word can be anywhere on a circle rather than an unambiguous location. This leads SSBoF to the failure of encoding the spatial relationship among multiple pairs of virtual words. As a result, the significantly discriminative spatial information in a local region cannot be effectively captured. The same issue also exists in the case of the Euclidean distance between the local feature and the barycenter in [27], where the Euclidean distance cannot resist non-rigid shape transformation and is only meaningful for rigid 3D shapes.

The aforementioned methods encode the spatial information from the global perspective, which directly encodes the global spatial information into shape representations. This makes the resulting representation with high dimensions but coarse spatial information, which is incompact for the subsequent processing. To capture more detailed spatial information in a more compact way, the proposed BoSCC encodes the global spatial information from the local perspective. BoSCC represents 3D shapes by the patterns of encodings of geometric and spatial information in 3D local regions.

C. Encoding the Spatial Information for 2D Images

Features are always important for 2D images understanding [38]–[40]. Many methods have been proposed to encode the spatial relationship among virtual words in 2D images [15]–[18], [25], [41]–[45]. Among these methods, correlation-based methods present several attractive properties, such as low computational complexity and robustness with respect to basic geometric transformations. In the pioneer work of Huang et al [41], the correlation between colors in images is employed for indexing and classifying images. Following this idea, Savarese *et al.* [42] propose to capture the spatial arrangement of virtual words based on the co-occurrence matrix of virtual words. However, local BoW representations are extracted from local regions in [42], leading to the spatial information loss in local regions. By considering the correlation between local features, the work [46] significantly speeds up the virtual word learning process while maintaining accuracy. To benefit from the advantages of correlation, we propose to capture the complex spatial information in a local region by the correlation among spatial contexts of vertices.

In addition, the Markov chain [28] is also employed to encode the spatial relationship between virtual words. In [44] and [45], a Markov chain is used to incorporate the spatial information into classical histogram features. In [44], the spatial co-occurrence matrix based Markov chain is introduced to encode relationships among intra-virtual words and inter-virtual words, respectively. Accordingly, Markov stationary features are proposed by combining the initial and stationary distributions of Markov chain. To eliminate the inherent ambiguities of Markov stationary features, homogeneity-aware Markov stationary features are derived in [45], which only considers mutually distinct pairs of virtual words in the spatial co-occurrence matrix. In our method, the high performance of homogeneity-aware Markov stationary features facilities the modelling of spatial context around each vertex in a multi-scale manner.

III. Bag of Spatial Context Correlations

BoSCC is introduced in detail in this section. First, the overview of BoSCC is presented. Next, several subsections

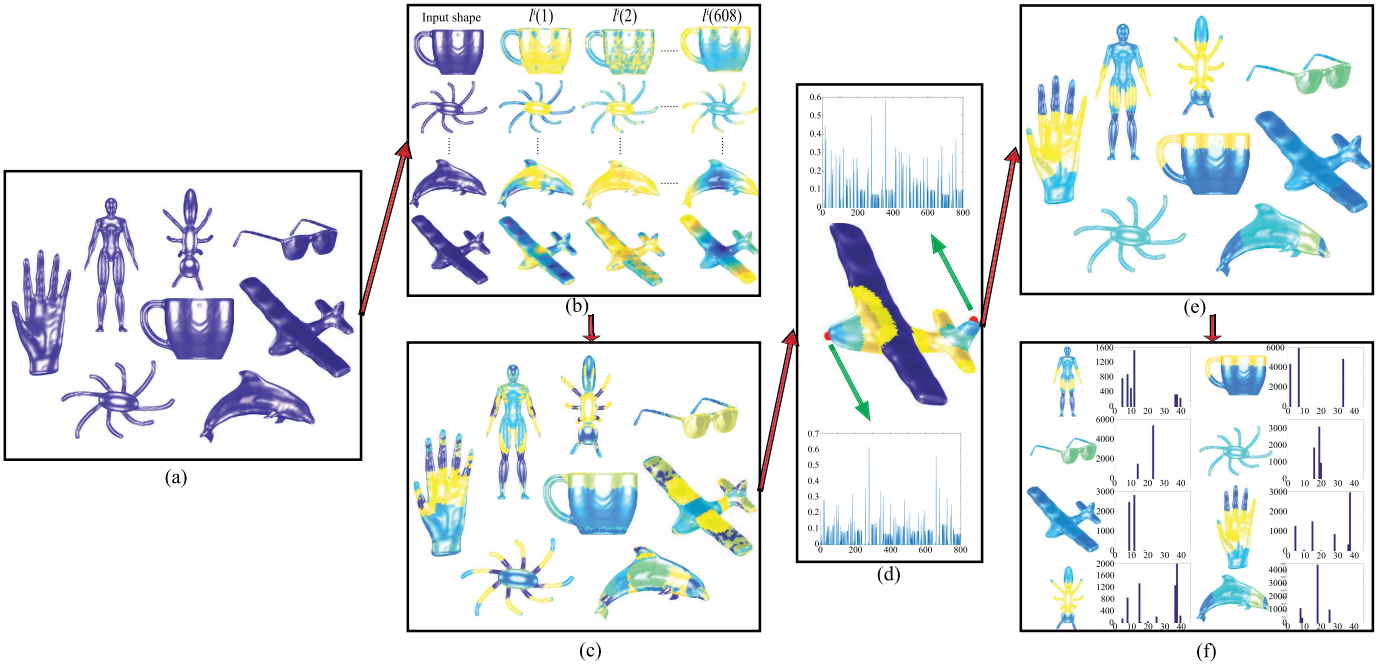

Fig. 2. The framework of computing the bag of spatial context correlations. For a set of 3D shapes briefly shown in (a), the local features of each shape are first extracted as shown in (b), which are used for learning a low-level virtual dictionary shown in (c). Based on the low-level virtual dictionary, the spatial context correlation in a local region centered at each vertex is calculated in (d), which is used for high-level virtual dictionary learning in (e). The proposed novel 3D shape representation, bag of spatial context correlations, are shown in (f).

are included to elaborate on the main steps or techniques mentioned in the overview, including low-level virtual dictionary learning, spatial context modelling, the computation of spatial context correlation, high-level virtual dictionary learning and the computation of BoSCC.

A. Overview of BoSCC

An overview of BoSCC is introduced in the following four steps and also illustrated in Fig 2.

1) **Low-level feature extraction.** Given a set of 3D shapes represented as meshes, as briefly shown in Fig. 2 (a), the local feature around each vertex on all shapes is extracted by several local descriptors. To distinguish from the following involved features, the local feature is called as the *low-level feature* of vertex. In Fig. 2 (b), the colors shown on each shape represent the values at the same dimension in all low-level features of vertices on that shape, where the 1st, 2nd and 608th dimensions are typically visualized. For the i-th shape, the values shown by colors are denoted as $l^i(1)$, $l^i(2)$ and $l^i(608)$, respectively. This step will be described in detail in Section III-B.
2) **Low-level virtual dictionary learning.** A *low-level virtual dictionary* is learned via clustering over the extracted low-level features from all shapes, where the cluster centroids are regarded as the types of *low-level virtual words*. In the learning process, each vertex on all shapes is labelled by the index of its nearest type of low-level virtual word. These indices on each shape are shown in Fig. 2 (c), separately. This step will be also described in detail in Section III-B.
3) **The computation of spatial context correlation.** With the low-level virtual dictionary, the geometric and spatial information of a local region can be described by the proposed spatial context correlation. First, the spatial context around each vertex is modelled to capture the distribution of neighboring low-level virtual words. Then, the correlation between spatial contexts is able to simultaneously encode i) low-level virtual words and ii) their spatial relationship, containing geometric and spatial information in the local region. Comparing to the low-level feature, the spatial context correlation also encodes the spatial information of a local region, thus, it is called as the *high-level feature* of vertex in this paper. The spatial context correlations of two vertices on an airplane are demonstrated in Fig. 2 (d). More details of the spatial context modelling and the computation of spatial context correlation are presented in Section III-C and Section III-D, respectively.
4) **High-level virtual dictionary learning.** Using spatial context correlations as high-level features, a *high-level virtual dictionary* is learned, that is constituted by spatial context correlation patterns. Such patterns are obtained by clustering over all spatial context correlations (computed in Step 3) from all shapes. The clustered patterns (or centroids) are regarded as the types of *high-level virtual words*. In the learning process, each vertex on all shapes is labelled by the index of its nearest type of high-level virtual word, as shown in Fig. 2 (e). The detail of this step is presented in Section III-E.
5) **The computation of BoSCC.** Finally, BoSCC represents each 3D shape as a frequency histogram, that is

constructed by counting the occurrence of each type of high-level virtual word, as shown in Fig. 2 (f). More details are introduced in Section III-F.

B. Low-Level Virtual Dictionary Learning

Let $\mathbb{S} = \{\mathbf{S}^i | i \in [1, N]\}$ be a set of 3D shapes, where the vertices on each shape $\mathbf{S}^i$ is denoted as a set $\{\boldsymbol{v}_j^i | j \in [1, N_V^i]\}$. Since we do not know in advance whether the shape $\mathbf{S}^i$ is rigid or non-rigid, we construct the low-level feature $\boldsymbol{l}_j^i$ of each vertex $\boldsymbol{v}_j^i$ using many informative local descriptors. Similar to [47], the following descriptors are included, i.e., multi-scale surface curvature, average geodesic distances [48], singular values extracted from principal component analysis of local regions, distances from medial surface points [49], the shape diameter function [50], shape contexts [51], and spin images [33]. For each descriptor, different parameters are also employed to generate more comprehensive features. Overall, these descriptors generate 64, 18, 60, 270, 72, 100, and 24 dimensional feature vectors, respectively. The low-level feature $\boldsymbol{l}_j^i$ is formed via concatenating all the aforementioned feature vectors into a 608-dimensional vector as defined by $\boldsymbol{l}_j^i = [l_j^i(1), l_j^i(2), \ldots, l_j^i(608)]$. Then, the low-level features of vertices on all shapes are collected into a set which is denoted as $\mathbb{L} = \{\boldsymbol{l}_j^i | i \in [1, N], j \in [1, N_V^i]\}$. Readers can refer to [47] for more details of the involved descriptors.

Subsequently, a low-level virtual dictionary $\boldsymbol{\Phi}$ is learned by clustering over all the low-level features in $\mathbb{L}$. This is implemented based on K-means clustering method via optimizing iteratively the objective function defined as follows,

$$\underset{\boldsymbol{\phi}_k \in \boldsymbol{\Phi}}{\operatorname{argmin}} \sum_{k=1}^{K_L} \sum_{\boldsymbol{l}_j^i \in \mathbf{L}_k} \|\boldsymbol{l}_j^i - \boldsymbol{\phi}_k\|^2. \tag{1}$$

In our implementation, the initial cluster centers are randomly selected for K-means. To alleviate the effect caused by the random initialization of cluster centers, the final clustering result is optimally selected from 10 candidate clustering results which are differently initialized.

Using Eq. (1), the low-level features in $\mathbb{L}$ are clustered into K_L clusters, and each cluster is denoted by $\mathbf{L}_k$, where $k \in [1, K_L]$. Here, $\mathbf{L}_k$ is a subset of $\mathbb{L}$ such that $\mathbb{L}$ can also be described as $\{\mathbf{L}_k | k \in [1, K_L]\}$. The center of each cluster $\mathbf{L}_k$ serves as one type of low-level virtual word ϕ_k in $\boldsymbol{\Phi}$, that is, $\boldsymbol{\Phi} = \{\phi_k | k \in [1, K_L]\}$. The selection of K_L will be discussed in Section IV. Finally, each low-level feature $\boldsymbol{l}_j^i$ in $\mathbb{L}$ is labelled by the index of its nearest type of low-level virtual word in $\boldsymbol{\Phi}$. With the low-level virtual dictionary $\boldsymbol{\Phi}$, the distribution of low-level virtual words can be presented on each shape as shown in Fig. 2 (c) and Fig. 3 (a).

Based on the low-level virtual dictionary $\boldsymbol{\Phi}$, the spatial context around each vertex can be further modelled, as described in the following subsection.

C. Spatial Context Modelling

1) Overview: For a local region $\mathbf{R}_{\boldsymbol{v}}$ centered at a vertex $\boldsymbol{v}$, the spatial context around $\boldsymbol{v}$ is modelled in a multi-scale manner to capture the distribution of low-level virtual words in $\mathbf{R}_{\boldsymbol{v}}$. In the modelling process, a Markov chain is adopted in each scale of $\boldsymbol{v}$ to extract the distribution information from co-occurrence counts of pairwise types of low-level virtual words. To detail, the spatial relationship between the *intra-virtual words* and the one between the *inter-virtual words* are encoded by the initial and the stationary distributions of a Markov chain, respectively.

Note that the co-occurrence counts between pairwise low-level virtual words can be directly obtained from 3D surfaces, and the counts are invariant to rigid shape transformation. In addition, the counts are independent of mesh resolution, vertex topology and orientation on 3D surfaces, which makes the Markov chain effectively overcome the issues of arbitrary mesh resolutions, irregular vertex topology and orientation ambiguity on the 3D surface.

The procedures of spatial context modelling are illustrated in Fig. 3. After each vertex is labelled by the index of its nearest type of low-level virtual word, as shown in Fig. 3 (a), the spatial context around each vertex is modelled in multi-scales as demonstrated in Fig. 3 (b). In each scale shown in Fig. 3 (c), a Markov chain is constructed via forming the spatial co-occurrence matrix in Fig. 3 (d). Then, the initial and stationary distributions of the Markov chain are calculated according to the spatial co-occurrence matrix by the method proposed in [45], which encodes the spatial relationship between intra-virtual words and the one between inter-virtual words as shown in Fig. 3 (e) and Fig. 3 (f), respectively.

2) Spatial Co-Occurrence Matrix: The spatial context of $\boldsymbol{v}$ is modelled in the local region $\mathbf{R}_{\boldsymbol{v}}$. Firstly, $\mathbf{R}_{\boldsymbol{v}}$ is separated into several ring-like scales. The r-th scale is denoted by $\mathbf{R}_{(\boldsymbol{v},r)}$, where $r \in [1, N_R]$ and N_R is the number of scales. $\mathbf{R}_{(\boldsymbol{v},r)}$ can be established via partitioning the vertices in $\mathbf{R}_{\boldsymbol{v}}$ using a set of geodesic distance thresholds $\{R_0, R_1, \ldots, R_M\}$, i.e. $\mathbf{R}_{(\boldsymbol{v},r)} = \{\boldsymbol{u} | R_{r-1} \leq Geo(\boldsymbol{v}, \boldsymbol{u}) \leq R_r, \boldsymbol{u} \text{ and } \boldsymbol{v} \in \mathbf{S}\}$, where $Geo(\boldsymbol{v}, \boldsymbol{u})$ is the geodesic distance between the vertex $\boldsymbol{u}$ and the centeral vertex $\boldsymbol{v}$.

In each scale $\mathbf{R}_{(\boldsymbol{v},r)}$, a Markov chain is constructed by forming a spatial co-occurrence matrix, $\mathbf{C}_{(\boldsymbol{v},r)} \in \mathbb{Z}^{K_L \times K_L}$. Every entry c_{ab} of $\mathbf{C}_{(\boldsymbol{v},r)}$ counts co-occurrence frequency of the a-th and b-th types of low-level virtual words, where $a \in \{1, K_L\}$ and $b \in \{1, K_L\}$. A co-occurrence of two low-level virtual words is determined if they are connected by an edge. The c_{ab} is defined as below

$$c_{ab} = Num\{p_{\boldsymbol{u}}^i = a, p_{\boldsymbol{v}}^i = b | Edge(\boldsymbol{u}, \boldsymbol{v})\}, \tag{2}$$

where Num indicates the number of vertex pairs which satisfy all the conditions listed in the set, $p_{\boldsymbol{u}}^i$ and $p_{\boldsymbol{v}}^i$ are the indices of the nearest types of low-level virtual words by which the vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ are labelled, $Edge(\boldsymbol{u}, \boldsymbol{v})$ denotes that the vertices $\boldsymbol{u}$ and $\boldsymbol{v}$ are connected by an edge. Note that the spatial co-occurrence matrix $\mathbf{C}_{(\boldsymbol{v},r)}$ is symmetric and nonnegative [45].

The construction of $\mathbf{C}_{(\boldsymbol{v},r)}$ is illustrated in Fig. 4 (a) and Fig. 4 (b). In Fig. 4 (a), the central vertex $\boldsymbol{v}$ and its neighboring vertices form a local patch. In this example, the neighboring vertices are labeled by three types of low-level virtual words, such as a-th, b-th and c-th types of low-level virtual words.

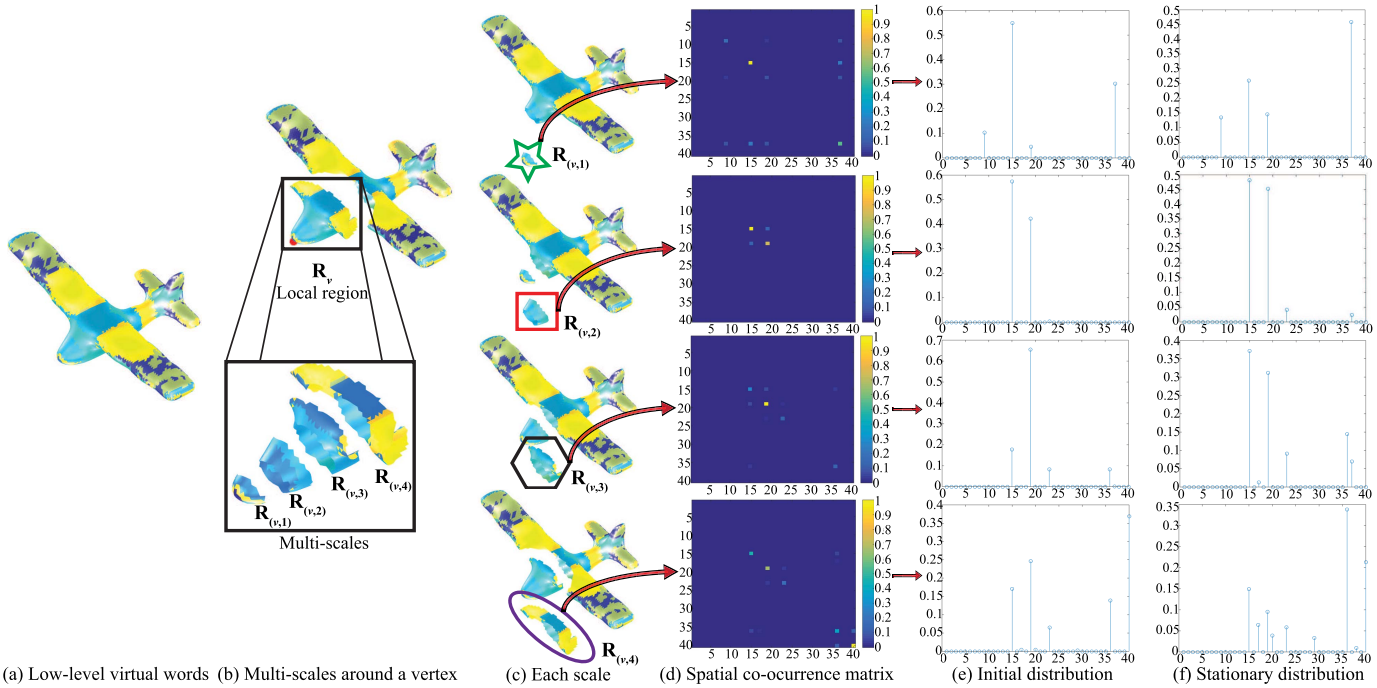

Fig. 3. The procedures of modelling the spatial context around a vertex marked by a red point. For a 3D shape with vertices labelled by low-level virtual words in (a), the spatial context around the vertex is modelled as shown from (b) to (e). The local region centered at the vertex is separated into multi-scales as shown in (b). Each scale detailed in (c) is used to construct the spatial co-occurrence matrix of a Markov chain, as shown in (d). According to (d), the initial and stationary distributions of the corresponding Markov chain are calculated, as shown in (e) and (f), respectively.

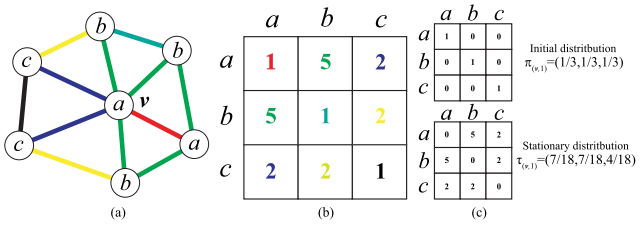

Fig. 4. The formation of spatial co-occurrence matrix. (a) The vertices connected by edges are labeled by the a-th, b-th and c-th types of low-level virtual words. (b) The spatial co-occurrence matrix is formed from (a), where the color of entries corresponds to the color of edges in (a). (c) The initial and stationary distributions of the Markov chain represented in (b).

The spatial co-occurrence matrix $\mathbf{C}_{(\boldsymbol{v},r)}$ derived from this local patch is given in Fig. 4 (b), where each entry of $\mathbf{C}_{(\boldsymbol{v},r)}$ indicates the number of edges with the same color as it in Fig. 4 (a). Taking $c_{ab} = 5$ for example, the green entry 5 means that the a-th and b-th types of low-level virtual words are co-occurred 5 times on the local patch, where the corresponding edges connecting a and b are also marked in the same green color.

3) Spatial Context: Based on $\mathbf{C}_{(\boldsymbol{v},r)}$, the transition matrix $\mathbf{T}_{(\boldsymbol{v},r)}$ of a Markov chain is derived to describe the feature of the Markov chain. The entry t_{ab} of $\mathbf{T}_{(\boldsymbol{v},r)}$ is defined in the following,

$$t_{ab} = c_{ab} / \sum_{b=1}^{K_L} c_{ab}. \tag{3}$$

The definition of $\mathbf{T}_{(\boldsymbol{v},r)}$ satisfies some basic properties of a Markov chain, including $t_{ab} > 0, \forall\ (a,b) \in Edge(a,b)$, and $\sum_{b=1}^{K_L} t_{ab} = 1, a \in [1, K_L]$.

However, the representation of $\mathbf{T}_{(\boldsymbol{v},r)}$ is of K_L^2 dimensions and may not be robust [45]. To model the spatial information in each scale more effectively, the homogeneity-aware Markov stationary features [45] are employed in our method. The homogeneity-aware Markov stationary features are formed by the initial and stationary distributions of the corresponding Markov chain, whose representation is a $2K_L$ dimensional vector defined as follows.

The initial distribution of various types of low-level virtual words over $\mathbf{R}_{(\boldsymbol{v},r)}$, denoted by $\boldsymbol{\pi}_{(\boldsymbol{v},r)}$, is a K_L dimensional vector. $\boldsymbol{\pi}_{(\boldsymbol{v},r)}$ encodes the intra-virtual words transition information in the r-th scale around the vertex $\boldsymbol{v}$. $\boldsymbol{\pi}_{(\boldsymbol{v},r)}$ consists of the normalized diagonal entries of $\mathbf{C}_{(\boldsymbol{v},r)}$ as defined in the following,

$$\boldsymbol{\pi}_{(\boldsymbol{v},r)}(k) = c_{kk} / \sum_{k=1}^{K_L} c_{kk}, \tag{4}$$

where $\boldsymbol{\pi}_{(\boldsymbol{v},r)}$ derived from $\mathbf{C}_{(\boldsymbol{v},r)}$ is illustrated in Fig. 4 (c).

The stationary distribution of various types of low-level virtual words over $\mathbf{R}_{(\boldsymbol{v},r)}$, denoted by $\boldsymbol{\tau}_{(\boldsymbol{v},r)}$, is also a K_L dimensional vector. $\boldsymbol{\tau}_{(\boldsymbol{v},r)}$ encodes the inter-virtual words transition information in the r-th scale around the vertex $\boldsymbol{v}$. It is the limiting probability of each type of low-level virtual word that the Markov chain will be in infinitely. Therefore, $\boldsymbol{\tau}_{(\boldsymbol{v},r)}$ satisfies $\boldsymbol{\tau}_{(\boldsymbol{v},r)} = \boldsymbol{\tau}_{(\boldsymbol{v},r)} \times \mathbf{T}_{(\boldsymbol{v},r)}$, which makes it capable of becoming a unique and invariant measure of the Markov chain. We employ the informative solution from the homogeneity-aware Markov stationary features as defined below,

$$\boldsymbol{\tau}_{(\boldsymbol{v},r)}(k) = \sum_{q \neq k} c_{kq} / \sum_{k=1}^{K_L} \sum_{q \neq k} c_{kq}, \tag{5}$$

where $\boldsymbol{\tau}_{(\boldsymbol{v},r)}$ derived from $\mathbf{C}_{(\boldsymbol{v},r)}$ is also illustrated in Fig. 4 (c).

Finally, the spatial context around each vertex $\boldsymbol{v}$ is represented by the set of pairs of $\boldsymbol{\pi}_{(\boldsymbol{v},r)}$ and $\boldsymbol{\tau}_{(\boldsymbol{v},r)}$ in terms of different scales, i.e. $\{(\boldsymbol{\pi}_{(\boldsymbol{v},r)}, \boldsymbol{\tau}_{(\boldsymbol{v},r)})|r \in [1, N_R]\}$. The spatial context encodes the spatial relationship between the low-level virtual words in the same type and the one between the low-level virtual words in different types. Based on these spatial contexts, the spatial context correlation is computed as detailed in the following subsection.

D. The Computation of Spatial Context Correlation

1) Overview: The co-occurrence based Markov chain overcomes the obstacles of 3D shapes in spatial context modelling, and it keeps the spatial context invariant to rigid shape transformations. However, the co-occurrence makes the spatial context still suffer from two drawbacks. On one hand, the co-occurrence indicated by two neighboring low-level virtual words is a kind of short-range connection, which makes the spatial context unable to model the long-range spatial relationship among low-level virtual words in a local region. This aspect determines that the spatial information encoded by the spatial context is coarse and restricted. On the other hand, the co-occurrence information is vague and easily affected by noisy virtual words on 3D shapes due to non-rigid shape transformation. To make up these disadvantages, the novel spatial context correlation is proposed to encode the long-range spatial information in a local region in a robust way, which alleviates the influence caused by non-rigid shape transformation.

The spatial context correlation is implemented via capturing the correlation between two spatial contexts in a local region. One is the spatial context of the vertex centered at the region, the other is the mean spatial context among vertices that are labeled by the same type of low-level virtual word in the same region. The spatial context correlation implicitly models how the two spatial contexts impact each other via the correlation on 3D surfaces. Taking the local region centered at a vertex $\boldsymbol{v}_j^i$ for example, the spatial context correlation of this region is denoted as $\boldsymbol{h}_j^i$, which is also implemented in a multi-scale way. In each scale of $\boldsymbol{v}_j^i$, $\boldsymbol{h}_j^i$ is obtained as the correlation between the spatial context of $\boldsymbol{v}_j^i$ and the mean spatial context of vertices assigned to each type of low-level virtual word.

The reason why the correlation can capture such complex spatial information of a 3D shape mainly lies in three aspects. First, the low-level virtual words are spatially distributed in different scales of $\boldsymbol{v}_j^i$, and hence, the correlation is also computed in a multi-scale manner. Second, there are spatial overlaps between the scales of $\boldsymbol{v}_j^i$ and the scales of vertices assigned to each type of low-level virtual word, which implicitly bridges the gap between the spatial context change of $\boldsymbol{v}_j^i$ and the one of vertices assigned to each type of low-level virtual word. Thus, the changes in common help to alleviate the effect of noisy virtual words caused by non-rigid shape transformation. Third, the long-range spatial information can be captured by averaging spatial contexts of vertices assigned to the same type of low-level virtual word in different scales of $\boldsymbol{v}_j^i$.

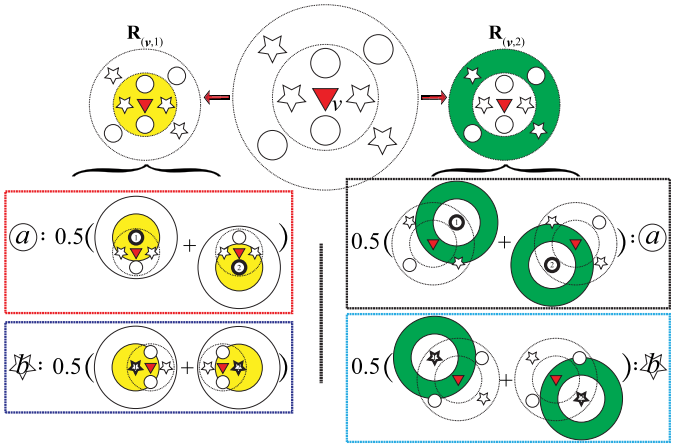

Fig. 5. An illustration of spatial context correlation. The circle and star nodes denote vertices assigned to two types of low-level virtual words in a local region centered at $\boldsymbol{v}$, where $\boldsymbol{v}$ is marked by a red triangle. In each of the two scales of $\boldsymbol{v}$, the mean spatial context of vertices assigned to each type of low-level virtual word is computed.

An illustration of spatial context correlation is shown in Fig 5, where the spatial context correlation of $\boldsymbol{v}$ (denoted as a red triangle) is computed over the neighboring region of $\boldsymbol{v}$, $\mathbf{R}_{\boldsymbol{v}}$. $\boldsymbol{v}$ is surrounded by the vertices assigned to the a-th and b-th types of low-level virtual words (denoted as the circle nodes and the star nodes, respectively) in two scales of $\boldsymbol{v}$, i.e. $\mathbf{R}_{(\boldsymbol{v},1)}$ and $\mathbf{R}_{(\boldsymbol{v},2)}$. The yellow regions and green regions denote the spatial context of each vertex in two different scales, respectively. In each scale of $\boldsymbol{v}$, the mean spatial context among vertices assigned to each type of low-level virtual word is computed. Taking the two vertices assigned to the a-th type of low-level virtual word in $\mathbf{R}_{(\boldsymbol{v},1)}$ for example, the computation of mean spatial context is detailed in the red rectangle, where the yellow regions representing the spatial contexts of two vertices are averaged. In addition, the mean spatial context of vertices assigned to the a-th type of low-level virtual word in $\mathbf{R}_{(\boldsymbol{v},2)}$ is calculated, as illustrated via averaging the two green regions in the black rectangle. Similarly, the mean spatial contexts of vertices assigned to the b-th type of low-level virtual word in $\mathbf{R}_{(\boldsymbol{v},1)}$ and $\mathbf{R}_{(\boldsymbol{v},2)}$ are calculated via repeating the above process, as shown in the blue and cyan rectangles, respectively.

2) Spatial Context Correlation: Specifically, $\boldsymbol{h}_j^i$ includes two parts in each scale of $\boldsymbol{v}_j^i$, which is defined as $\boldsymbol{h}_j^i = \{(\boldsymbol{\Pi}_{\boldsymbol{v}}(r), \boldsymbol{\Gamma}_{\boldsymbol{v}}(r))|r \in [1, N_R]\}$. In the r-th scale of $\boldsymbol{v}_j^i$, $\boldsymbol{\Pi}_{\boldsymbol{v}}(r)$ is the correlation between $\boldsymbol{\pi}_{(\boldsymbol{v},r)}$, while $\boldsymbol{\Gamma}_{\boldsymbol{v}}(r)$ is the correlation between $\boldsymbol{\tau}_{(\boldsymbol{v},r)}$. $\boldsymbol{\Pi}_{\boldsymbol{v}}(r)$ and $\boldsymbol{\Gamma}_{\boldsymbol{v}}(r)$ are two matrices, where each row of $\boldsymbol{\Pi}_{\boldsymbol{v}}(r)$ and $\boldsymbol{\Gamma}_{\boldsymbol{v}}(r)$ is the mean spatial context of vertices assigned to the same type of low-level virtual word in the r-th scale. Each row of $\boldsymbol{\Pi}_{\boldsymbol{v}}(r)$ and $\boldsymbol{\Gamma}_{\boldsymbol{v}}(r)$ is defined as follows, respectively,

$$\boldsymbol{\Pi}_{\boldsymbol{v}}(k,r) = \sum_{\boldsymbol{u} \in \mathbf{M}_k(r)} \boldsymbol{\pi}_{(\boldsymbol{u},r)} / |\mathbf{M}_k(r)|, \tag{6}$$

$$\boldsymbol{\Gamma}_{\boldsymbol{v}}(k,r) = \sum_{\boldsymbol{u} \in \mathbf{M}_k(r)} \boldsymbol{\tau}_{(\boldsymbol{u},r)} / |\mathbf{M}_k(r)|, \tag{7}$$

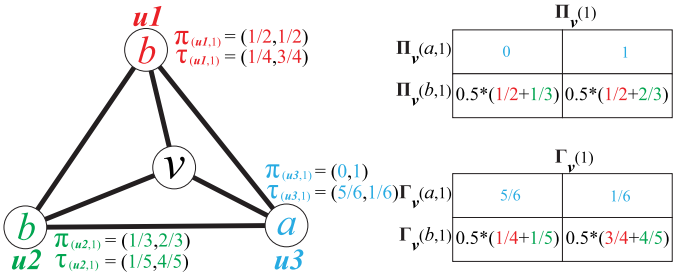

Fig. 6. A demonstration of computing the spatial context correlation of vertex $\boldsymbol{v}$ in one scale. The procedure of computation is shown in the right table, where the colors of entries correspond to the colors of $\boldsymbol{\pi}$ and $\boldsymbol{\tau}$.

where $\mathbf{M}_k(r) = \{\boldsymbol{u} | \boldsymbol{u} \in \mathbf{R}_{(\boldsymbol{v},r)}, p^i_{\boldsymbol{u}} = k\}$ denotes all the vertices $\boldsymbol{u}$ assigned to the k-th type of low-level virtual word in the r-th scale of $\boldsymbol{v}$, and $|\mathbf{M}_k(r)|$ is the number of vertices in $\mathbf{M}_k(r)$, $r \in [1, N_R]$, $k \in [1, K_L]$.

Using Eq. (6) and Eq. (7), the correlation between the spatial context of vertex $\boldsymbol{v}^i_j$ and the mean spatial context of vertices in $\mathbf{M}_k(r)$ is computed. It implicitly encodes how the mean spatial context impacts the spatial context of $\boldsymbol{v}^i_j$ in the r-th scale. For vertices in $\mathbf{M}_k(r)$, the mean spatial context is obtained by averaging their spatial contexts in the same r-th scale.

Eq. (6) and Eq. (7) are further explained via demonstrating the computation of spatial context correlation in one scale in Fig. 6. The vertex $\boldsymbol{v}$ is surrounded by three vertices ($\boldsymbol{u}1$, $\boldsymbol{u}2$ and $\boldsymbol{u}3$) located in the first scale of $\boldsymbol{v}$. For a better presentation, the three vertices are only assigned to the a-th and b-th types of low-level virtual words. The spatial contexts of three vertices are shown beside. The computed spatial context correlation $\{(\boldsymbol{\Pi}_{\boldsymbol{v}}(1), \boldsymbol{\Gamma}_{\boldsymbol{v}}(1))\}$ is shown as two matrices on the right. The colors of vertices and spatial contexts are used to show the computation process more clearly.

3) A More Compact Representation: For each scale r, $\boldsymbol{\Pi}_{\boldsymbol{v}}(r)$ and $\boldsymbol{\Gamma}_{\boldsymbol{v}}(r)$ are two $K_L \times K_L$ matrices, resulting in a computation-inefficient representation of $\boldsymbol{h}^i_j$, which is $2 \times K^2_L \times N_R$ dimensional. However, such a high dimension further burdens the learning of high-level virtual words.

To remedy this disadvantage, we regard the $\boldsymbol{\Pi}_{\boldsymbol{v}}(r)$ and $\boldsymbol{\Gamma}_{\boldsymbol{v}}(r)$ as the spatial co-occurrence matrices of two new Markov chains. Then, the initial and stationary distributions are respectively computed using Eq. (4) and Eq. (5) to represent the two new spatial co-occurrence matrices. This procedure provides a more compact representation of $\boldsymbol{h}^i_j$ without any information loss, in which the dimension of $\boldsymbol{h}^i_j$ is significantly reduced from $2 \times K^2_L \times N_R$ to $4 \times K_L \times N_R$.

E. High-Level Virtual Dictionary Learning

The spatial context correlation $\boldsymbol{h}^i_j$ is computed from the local region centered at each vertex $\boldsymbol{v}^i_j$. Since $\boldsymbol{h}^i_j$ contains the spatial information of a local region, it is regarded as a high-level feature of $\boldsymbol{v}^i_j$ in this paper. The $\boldsymbol{h}^i_j$ extracted from each vertex on all shapes is collected into a high-level feature set $\mathbb{H}$, such that $\mathbb{H} = \{\boldsymbol{h}^i_j | i \in [1, N], j \in [1, N^i_V]\}$. In order to learn the high-level virtual dictionary $\boldsymbol{\Psi}$, all the high-level features in $\mathbb{H}$ are clustered into K_H clusters, where each cluster is denoted by $\mathbf{H}_d$, then, $\mathbb{H}$ can also be described as $\{\mathbf{H}_d | d \in [1, K_H]\}$. The learning of $\boldsymbol{\Psi}$ is implemented based on K-means clustering method via optimizing iteratively the object function defined as follows,

$$\underset{\boldsymbol{\psi}_d \in \boldsymbol{\Psi}}{\operatorname{argmin}} \sum_{d=1}^{K_H} \sum_{\boldsymbol{h}^i_j \in \mathbf{H}_d} \|\boldsymbol{h}^i_j - \boldsymbol{\psi}_d\|^2, \tag{8}$$

where $\boldsymbol{\Psi} = \{\boldsymbol{\psi}_d | d \in [1, K_H]\}$ and the center of each cluster $\mathbf{H}_d$ serves a type of high-level virtual word $\boldsymbol{\psi}_d$. Similar to the learning of low-level virtual dictionary, the final clustering result is also optimally selected from 10 candidate clustering results which are differently initialized.

The selection of K_H will be discussed in Section IV. The learning result labels each $\boldsymbol{h}^i_j$ by the index of its nearest type of high-level virtual word $\boldsymbol{\psi}_d$. Then, the distribution of high-level virtual words can be presented on each shape, as shown in Fig. 3 (e).

Based on the high-level virtual dictionary $\boldsymbol{\Psi}$, BoSCC can be computed for each shape via counting the occurrence frequency of all types of high-level virtual words.

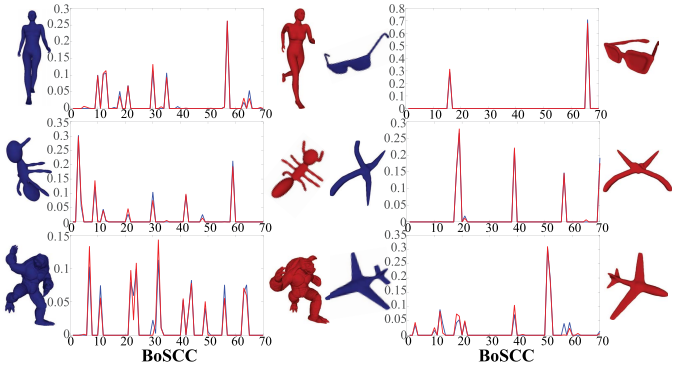

Fig. 7. The BoSCC representations of different rigid and non-rigid shapes in different classes. The BoSCC representation is drawn in the same color as the corresponding shape in each class. For example, the blue woman shape is with the blue histogram and the red woman shape is with the red histogram.

F. The Computation of BoSCC

The proposed BoSCC, denoted as $\boldsymbol{f}^i$, is computed via quantizing all $\boldsymbol{h}^i_j$ in shape $\mathbf{S}^i$ into a K_H dimensional vector. The element of $\boldsymbol{f}^i$ represents the occurrence frequency of each type of high-level virtual word $\boldsymbol{\psi}_d$ in $\mathbf{S}^i$. Described by the spatial context correlation patterns, $\boldsymbol{f}^i$ is more compact and encodes more detailed spatial information than other methods.

In Fig. 7, the BoSCC representations of different shapes in different classes are shown. In each class, although shapes are transformed by different rigid and non-rigid shape transformations, such as rotations or different poses of human and ants, the BoSCC representations of shapes in the same class are almost identical, in addition, the ones of shapes in different classes are significantly different from each other, thereby promoting the shape discriminability. The result in this example shows that BoSCC is capable of discriminating shapes from different classes and resisting rigid and non-rigid shape transformations by a compact representation.

The computation of BoSCC is carried out using MATLAB on a computer with an i7 CPU and 16GB RAM. Based on

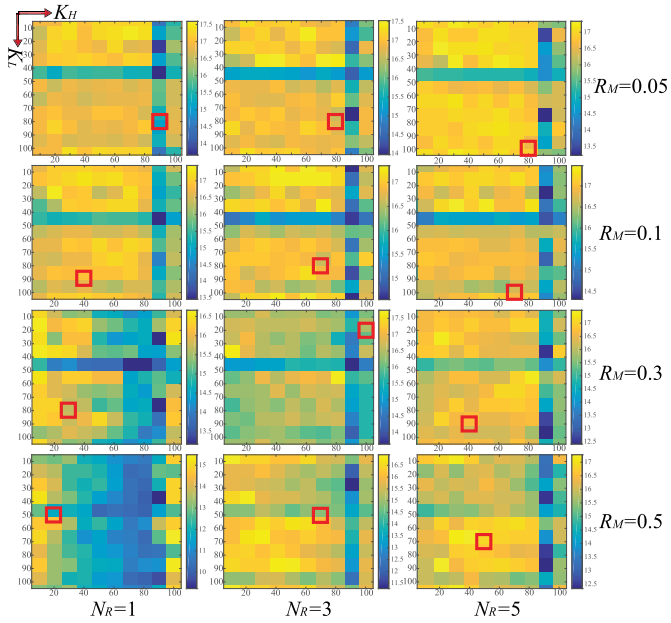

Fig. 8. The comparison of BoSCC parameters. The involved parameters include the local region size R_M, the number of scales N_R, the number of low-level virtual words K_L and the number of high-level virtual words K_H. The red square marks the best performance in each pair of specific N_R and R_M.

the low-level virtual words, the average time for computing the spatial context correlation of a local region is about 0.005 second, where the region with three scales is determined by a threshold of 0.1 times maximum geodesic distance. The time complexities of K-means for learning low-level and high-level virtual words are both $O(t)$, where t is proportional to three factors: the number of clusters, the number of samples, and the number of iterations.

IV. Results and Analysis

In this section, the results and analysis of the performance of BoSCC are presented. The setup of BoSCC parameters is first discussed. By demonstrating how these parameters affect the discriminability of BoSCC for global shape retrieval, the parameters tuning procedure is explained. Then, BoSCC is compared with the state-of-the-art methods under several different shape benchmarks for three common applications of 3D shape analysis, including global shape retrieval, shape classification and partial shape retrieval. The shapes employed in the experiments are with various mesh resolutions, irregular vertex topology, and different rigid or non-rigid shape transformations.

A. The Setup of BoSCC Parameters

The setup of BoSCC parameters is performed through 3D shape retrieval under a shape set which is formed by the first 280 shapes in the labeledPSB dataset [47]. Note that other shape sets can also be used for this setup. For clear comparison, the area under the Precision and Recall (PR) curve is used to compare the retrieval results obtained by different sets of parameters.

The BoSCC parameters include the size of low-level virtual dictionary K_L, the size of high-level virtual dictionary K_H, the number of scales N_R, and the size of local region around each vertex which is determined by the maximum geodesic distance threshold R_M.

Four candidates are considered for comparing the impact made by the increasing size of local region, such as $R_M \in \{0.05, 0.1, 0.3, 0.5\}$. For each R_M, the number of scales N_R is set as $\{1, 3, 5\}$, respectively. The incremental N_R is able to provide more detailed spatial information around each vertex for the spatial context correlation to capture. For each combination of R_M and N_R, both K_L and K_H iterate from 10 to 100 with an increment of 10. Under the above parameter settings, the performance of BoSCC for 3D shape retrieval is analyzed comprehensively. The obtained results with different parameters are shown in Fig. 8, where the area under PR curves is shown by colors in each matrix with K_L as column and K_H as row. The nine matrices are indexed by R_M in row and N_R in column.

1) The Size of Local Region R_M: R_M determines the size of local region centered at each vertex. The results obtained by different sizes of local regions are compared in Fig. 9 (a), where each colored point is the mean of all entries of the matrix determined by the specific R_M and N_R in Fig. 8. From the comparison, we could see that comparable results are obtained with $R_M = 0.05$ and 0.1, which are better than the ones obtained with other region sizes. Moreover, the performance of BoSCC decreases while increasing the size of local regions. The comparison results imply that oversized local regions decrease the performance of encoding the spatial information. This is because the spatial context in each scale becomes too vague to be captured by a Markov chain. As a result, common spatial patterns cannot be effectively extracted through the correlation among spatial contexts. In the following experiments, BoSCC is calculated with $R_M = 0.1$, since the performance with $N_R = 5$ of $R_M = 0.1$ is better than the ones of $R_M = 0.05$. In addition, $N_R = 5$ is chosen as explained as follows.

2) The Number of Scales N_R: The larger N_R makes BoSCC encode more detailed spatial information, which increases the shape retrieval performance. This is verified by the results showing how the performance of BoSCC is impacted by N_R in Fig. 9 (b). In Fig. 9 (b), each colored point is the mean of values on all scales in Fig. 9 (a), and the performance of BoSCC increases along with increasing N_R. Thus, N_R is set to 5 for the following experiments. However, overlarge N_R would burden the computation of high-level virtual dictionary learning. Hence, no larger N_R is explored for better performance.

3) The Size of Low-Level Virtual Dictionary K_L: K_L mainly depends on the extracted low-level features. In Fig. 8, the same set of low-level features are used. Therefore, the impacts of K_L to the performance of BoSCC over different pairs of R_M and N_R are almost the same. This impact is further elaborated as follows. On one hand, it is impossible to discriminate patterns of low-level features when K_L is very small, e.g. $K_L = 10$, leading to the unsatisfactory performance for BoSCC. On the other hand, the impact of K_L is almost unchanged when K_L is

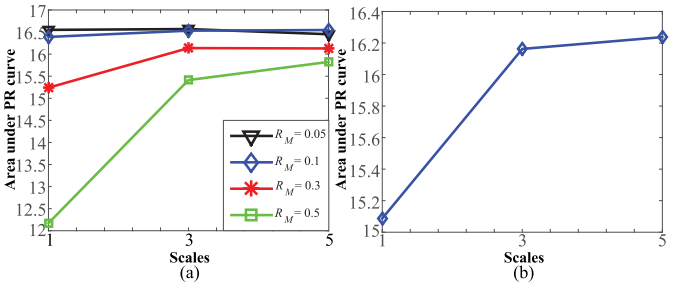

Fig. 9. The comparison between different results obtained with different R_M in (a) and different N_R in (b).

sufficiently large. Thus, the K_L is set to 100 in the following experiments, where the best performance is obtained with $K_L = 100$, $K_H = 70$, $R_M = 0.1$ and $N_R = 5$.

4) The Size of High-Level Virtual Dictionary K_H: K_H specifies the number of spatial context correlation patterns. Similar to K_L, the spatial context correlation patterns cannot be effectively discriminated with too small K_H, e.g. $K_H = 10$. In addition, K_H is determined based on the complexity of the spatial context correlations. It is necessary to choose smaller K_H for less complex spatial context correlations. This can be clearly observed from the results obtained with $N_R = 1$. With increasing R_M from 0.1 to 0.5, the results obtained with a large K_H are getting worse. This is because, with only one scale ($N_R = 1$), the increasing R_M makes the spatial context capture more limited spatial information of local regions, which further affects the performance of spatial context correlation. Thus, it is enough for small K_H to discriminate the spatial context correlation patterns. Adopting the same strategy as K_L, K_H is set to 70 in the following experiments.

B. Global Shape Retrieval

Global shape retrieval aims to search the globally similar shapes in a large dataset for a given query shape. In the following retrieval experiments, BoSCC is compared with several state-of-the-art methods of encoding the spatial relationship among virtual words for 3D shapes, such as SSBoF [25], [26], ISPM [24], and Bag of Spatial Context (BoSC). In addition, BoW is regarded as a baseline method. Note that BoSC is a representation based on the high-level virtual words learned from the multi-scale spatial context, which aims to highlight the effect of correlation employed by BoSCC. In addition, the Euclidean distance in the feature space is used to measure the similarity between two arbitrary shapes.

The comparison experiments are carried out under several well-known datasets, including the LabeledPSB [47], McGill 3D shape benchmark, SHREC2007 dataset [52] and SHREC2010 dataset [53]. In each shape set, all the compared methods employ the same low-level and high-level virtual dictionaries.

To comprehensively evaluate the performance of different methods, the PR curve, Nearest Neighbor (NN), First-Tier (FT), Second-Tier (ST), E-Measures (E) and Discounted Cumulated Gain vector (DCG) are employed. The FT is the percentage of top $K_C - 1$ matches (excluding the query)

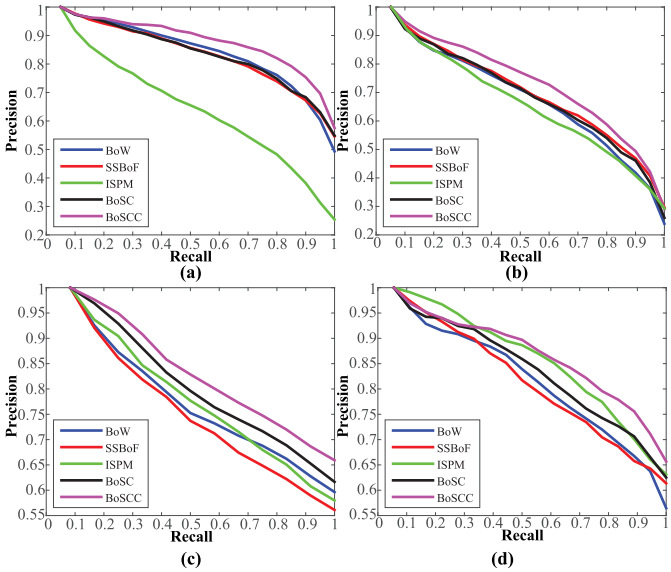

Fig. 10. The PR curves obtained by BoW, SSBoF, ISPM, BoSC and BoSCC under different shape benchmarks, including (a) LabeledPSB, (b) SHREC2007, (c) McGill and (d) SHREC2010.

from the class of the query, where K_C is the size of class. The ST is the same kind of metric, but for the top $2(K_C - 1)$ matches. The NN is the percentage of test in which the top match is from the class of query. The idea of E-Measures is to combine precision and recall into a single number to evaluate the whole performance. DCG measures the ranking quality.

The PR curves obtained by different methods under LabeledPSB, SHREC2007, McGill and SHREC2010 are shown in Fig. 10 (a), (b), (c) and (d), respectively. In all comparison results, our BoSCC obtains the best results among all the methods, which implies that BoSCC is able to effectively capture the geometric and spatial information of different shape regions by spatial context correlations. The outperforming result shows that BoSCC is superior to BoSC. This demonstrates that the correlation between the spatial context is able to capture the long-range spatial information in a more robust way than the spatial context, which is helpful for discriminating articulated shapes. Since rigid shapes cannot be split into local regions in a consistent manner as articulated shapes due to various geometric variations, the performance of ISPM is unsatisfactory when discriminating rigid shapes, such as tables and chairs. Some other retrieval measures are listed in Table I, where the bold numbers highlight the high performance of BoSCC in a more detailed way.

BoSCC is also compared with some other state-of-the-art methods under SHREC2007, including the Hybrid BoW of Lavoué [22], the curve based method of Tabia *et al.* [54], the BoW method of Toldo *et al.* [32] and Covariance Descriptors of Tabia *et al.* [14]. The PR curves under these methods are shown in Fig. 11 while the corresponding numerical evaluations are presented in Table II. In this experiment, BoSCC obtains the best result, implying that the spatial information encoded by BoSCC is with the highest discriminability. Note that the best result is obtained with $K_L = 280$ and $K_H = 100$, which is even better than the one with $K_L = 100$ and $K_H = 70$ as shown in Fig. 10 (b).

TABLE I

THE MEASURES OF SHAPE RETRIEVAL OBTAINED BY DIFFERENT METHODS UNDER SEVERAL 3D SHAPE BENCHMARKS

Benchmarks	Methods	Retrieval measures				
		NN	FT	ST	E	DCG
SHREC2007	BoW	0.8725	0.5829	0.7368	0.5152	0.8706
	SSBoF	0.8950	0.5953	0.7424	0.5238	0.8802
	ISPM	0.8872	0.5445	0.6736	0.4708	0.8531
	BoSC	0.8575	0.5836	0.7456	0.5188	0.8780
	BoSCC	**0.9125**	**0.6300**	**0.7664**	**0.5430**	**0.8962**
LabeledPSB	BoW	0.9607	0.7541	0.8610	0.6187	0.9419
	SSBoF	**0.9643**	0.7336	0.8642	0.6151	0.9394
	ISPM	0.8750	0.5434	0.6763	0.4702	0.8535
	BoSC	**0.9643**	0.7430	0.8612	0.6135	0.9397
	BoSCC	0.9536	**0.7705**	**0.8760**	**0.6276**	**0.9465**
McGill	BoW	0.8650	0.5122	0.6644	0.4831	0.8436
	SSBoF	0.8650	0.4952	0.6240	0.4632	0.8316
	ISPM	0.9024	0.4965	0.6562	0.4764	0.8467
	BoSC	0.9164	0.5534	0.7782	0.5252	0.9069
	BoSCC	**0.9319**	**0.5876**	**0.8202**	**0.5586**	**0.9260**
SHREC2010	BoW	0.9293	0.6768	0.8410	0.5884	0.9210
	SSBoF	0.9596	0.6644	0.8740	0.6018	0.9271
	ISPM	**0.9848**	0.7138	0.8952	0.6239	0.9359
	BoSC	0.9394	0.7039	0.8596	0.6052	0.9300
	BoSCC	0.9444	**0.7371**	**0.8982**	**0.6258**	**0.9428**

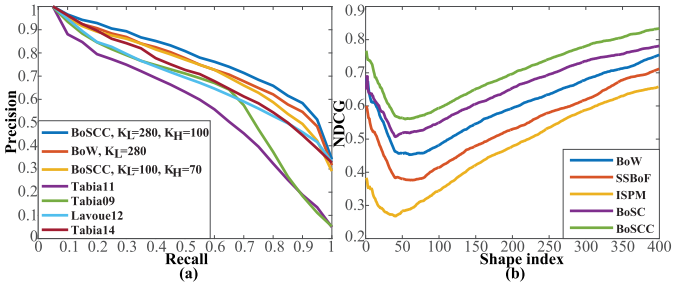

Fig. 11. (a) The comparison between PR curves which are obtained by BoSCC and some other state-of-the-art methods under SHREC2007. (b) The comparison between partial shape retrieval results obtained by BoW, SSBoF, ISPM, BoSC and BoSCC under SHREC2007 partial shape retrieval dataset.

TABLE II

THE MEASURES OF SHAPE RETRIEVAL OBTAINED BY DIFFERENT METHODS UNDER SHREC2007 DATASET

Methods	Retrieval measures			
	NN	FT	ST	DCG
BoSCC,$K_L = 280$ $K_H = 100$	**0.9410**	**0.6591**	**0.7962**	**0.9150**
BoW,$K_L = 280$	0.9325	0.6484	0.7802	0.9024
BoSCC,$K_L = 100$ $K_H = 70$	0.8875	0.5853	0.7580	0.8744
Tabia14	0.930	0.623	0.737	0.864
Lavoue12	0.918	0.590	0.734	0.841
Tabia11	0.853	0.527	0.639	0.719

C. Classification

Shape classification aims to classify shapes into different classes. The performance of BoSCC is further evaluated for shape classification task under the same four shape benchmarks including LabeledPSB, McGill, SHREC2007, and SHREC2010. Specifically, binary Support Vector Machine (SVM) [55] is employed to resolve the multi-class classification problem using the one-versus-one coding design, where $K_U(K_U-1)/2$ SVMs are trained and K_U is the number of shape classes. In addition, 10-fold cross validation on the training data are used to train the involved SVMs for better performance evaluation. For this task, the classification accuracy is measured using different numbers of training samples. 50%, 60%, 70%, 80% and 90% of shapes are randomly sampled from each class as training samples, and accordingly, the remaining shapes in each class are regarded as the testing samples.

TABLE III

THE CLASSIFICATION RESULTS OBTAINED BY DIFFERENT METHODS UNDER SEVERAL 3D SHAPE BENCHMARKS

Benchmarks	Methods	Training shape ratios				
		50%	60%	70%	80%	90%
SHREC2007	BoW	0.6850	0.2687	0.4250	0.5250	0.7000
	SSBoF	0.4800	0.0937	0.1083	0.2625	0.1250
	ISPM	0.0500	0.0500	0.0500	0.0500	0.0500
	BoSC	**0.6950**	0.4437	0.5983	0.7125	0.7500
	BoSCC	**0.6950**	**0.4687**	**0.6000**	**0.7375**	**0.8250**
LabeledPSB	BoW	0.8786	0.4196	0.7143	0.8929	0.9286
	SSBoF	0.6643	0.0714	0.1667	0.3571	0.1429
	ISPM	0.0000	0.1429	0.0714	0.0714	0.0714
	BoSC	**0.9000**	0.8304	0.9048	0.8929	0.9286
	BoSCC	**0.9000**	**0.9107**	**0.9405**	**0.9107**	**0.9643**
McGill	BoW	0.2381	0.3063	0.3093	0.4730	0.5184
	SSBoF	0.0682	0.1620	0.1415	0.0676	0.1357
	ISPM	0.0000	0.0482	0.0638	0.0625	0.0566
	BoSC	0.3957	0.4199	0.4531	0.4787	0.4749
	BoSCC	**0.4074**	**0.4514**	**0.5061**	**0.4883**	**0.6189**
SHREC2010	BoW	0.7172	0.7754	0.6594	0.6835	**0.8090**
	SSBoF	0.6061	0.1017	0.2367	0.3797	0.1011
	ISPM	0.0909	0.1000	0.1000	0.1000	0.1000
	BoSC	0.7172	0.7627	0.8430	0.7563	**0.8090**
	BoSCC	**0.7778**	**0.7881**	**0.8599**	**0.8038**	**0.8090**

The classification accuracy listed in Table III indicates that the performance of BoSCC is the best compared with BoW, SSBoF, ISPM, and BoSC. The high performance of BoSCC for classification benefits from not only the discriminability but also the compactness. Due to the compactness, BoSCC makes the training of classifier easily converged, especially when the training samples are limited. The compactness of BoSCC can be reflected from the dimension of representation vectors. For example, the dimensions of BoW, SSBoF, ISPM, BoSC and BoSCC are respectively 100 (equals to K_L), 10000 (equals to K_L^2), 25500 (equals to $K_L * K_{bin}$, where K_{bin} is the sum of region numbers of all resolutions.), 70 (equals to K_H) and 70 (equals to K_H). The result of ISPM is obtained under hard pooling with 8 resolutions, which concatenates the local BoW representations calculated from 1, 2, 4, 8, 16, 32, 64 and 128 regions split from a 3D shape, respectively. In the experiment, the results of SSBoF and ISPM are unsatisfactory since it is hard for a classifier to be trained well by only few training samples in such a high dimension.

In addition, the results of several methods with 60% training samples are visualized by the confusion matrix shown in Fig. 12. The lowest errors of BoSCC imply that the high performance of BoSCC is very suitable for 3D shape classification.

D. Partial Shape Retrieval

Partial shape retrieval aims to effectively search partially similar shapes as a given query in a database. SHREC2007 partial retrieval dataset is employed to evaluate the current experiment. Given the database of 400 watertight shapes and a set of 30 hybrid query shapes, each query is associated with a set of highly relevant shapes and a set of

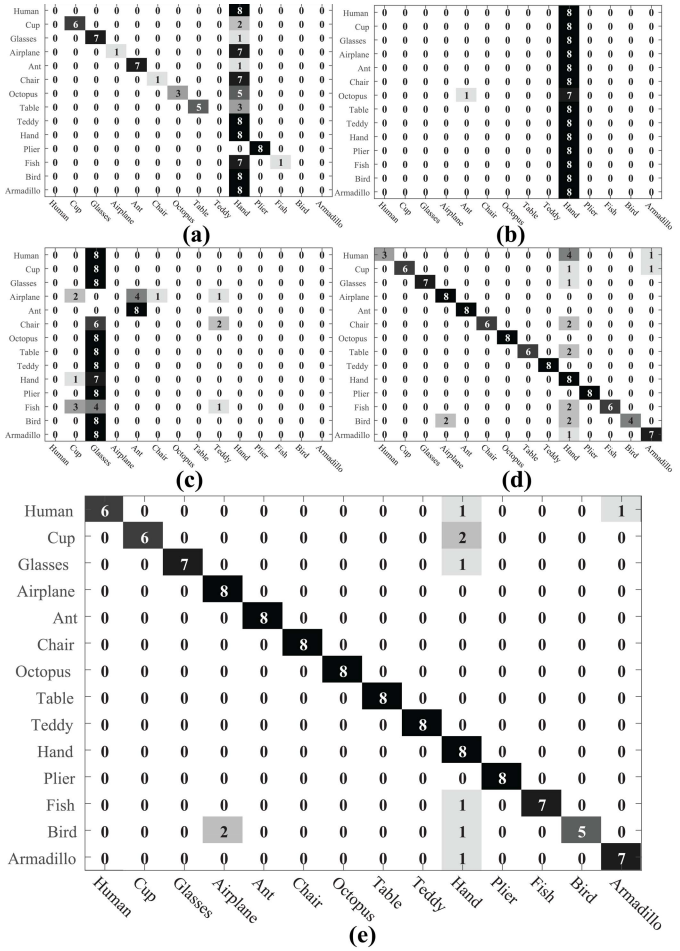

Fig. 12. The comparison shown by the confusion matrices with 60% training samples under LabledPSB dataset, which are obtained by (a) BoW, (b) SSBoF, (c) ISPM, (d) BoSC, (e) BoSCC.

marginally relevant shapes. In other words, these two sets of shapes respectively represent the classes which share a similar subpart with the query, and those which are reasonably similar to the query. The quantitative evaluation is computed using the Normalized Discounted Cumulated Gain (NDCG) vector [56]. NDCG is obtained via normalizing DCG by the ideal cumulated gain vector calculated from the ground truth. DCG is defined in the following,

$$DCG[i] = \begin{cases} G[i] & \text{if } i = 1; \\ DCG[i-1] + G[i]/\log_2 i & \text{otherwise,} \end{cases} \quad (9)$$

where $G[i]$ is a gain value representing the relevance of the i-th retrieved shape (2 for highly relevant, 1 for marginally relevant and 0 otherwise). For a given query, $DCG[i]$ is first computed, and then, $NDCG[i]$ is obtained via dividing the $DCG[i]$ by the ideal cumulated gain vector calculated from the ground truth.

The performance of BoSCC is compared with BoW, SSBoF, ISPM, and BoSC for partial shape retrieval. The NDCG values obtained by BoW, SSBoF, ISPM, BoSC and BoSCC are shown in Fig. 11 (b), and the best result is also obtained by BoSCC. Benefiting from the powerful ability of encoding the geometric and spatial information of local regions and the compactness,

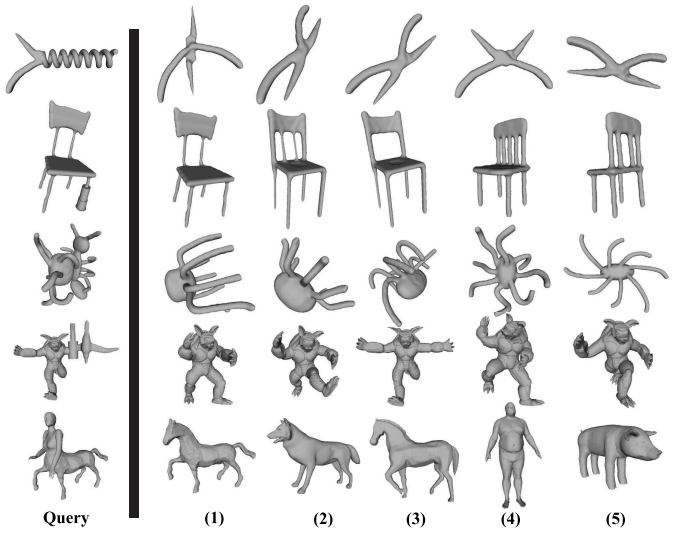

Fig. 13. Some examples of query shapes (the left column) and the top-5 retrieved shapes (the right 5 columns).

BoSCC is good at retrieving shapes queried with partially similar parts. The local perspective adopted by BoSCC is more effective to encode the local geometric and spatial information into global shape representations than the global perspective adopted by other state-of-the-art methods. This advantage can be further illustrated in Fig. 13, where five queries and their corresponding top-5 retrieved partially similar shapes are shown in each row, more importantly, the retrieved shapes are all highly relevant to the query shapes.

V. Limitation, Future Work and Conclusion

A. Limitation and Future Work

Although the proposed BoSCC achieves a high performance on shape retrieval and classification applications, there are still two limitations. One is that the performance of BoSCC might be affected by the discriminability of low-level virtual words, since the spatial context correlation patterns employed by BoSCC are based on the low-level virtual words. The other is that, although BoSCC can provide a compact 3D shape representation, the computation burden for learning high-level virtual words is still high. This is because the dimension of spatial context around each vertex increases while increasing the size of low-level virtual dictionary and the number of scales.

In the future, we plan to explore how to directly learn the high-level virtual words from low-level features and their spatial information, which aims to overcome the limitations of BoSCC. However, this is still a big challenge.

B. Conclusion

To remedy the significant issue of lacking spatial information in BoW for 3D shapes, BoSCC is proposed as a spatially enhanced 3D shape representation based on the novel spatial context correlation. It effectively encodes the spatial relationship among virtual words on the 3D surface into an

occurrence frequency histogram of spatial context correlation patterns. Benefiting from the novel local perspective, BoSCC is able to encode more detailed spatial information and become more compact than other global perspective based methods. The novel spatial context correlation overcomes the obstacles including arbitrary mesh resolution, irregular vertex topology, and orientation ambiguity on the surface in the spatial information encoding procedure, which facilitates the encoded spatial information to become invariant to rigid and non-rigid shape transformations. These advantages of BoSCC are justified by its high discriminability shown in the shape classification and retrieval experiments, especially when training classifiers with the limited number of samples or performing for partial shape retrieval, from which the existing state-of-the-art methods are still suffering.

REFERENCES

[1] J. M. Ponte and W. B. Croft, "A language modeling approach to information retrieval," in *Proc. 21st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 1998, pp. 275–281.

[2] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in *Proc. AAAI Workshop Learn. Text Categorization*, 1998, pp. 41–48.

[3] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 524–531.

[4] R. Ji, H. Yao, W. Liu, X. Sun, and Q. Tian, "Task-dependent visual-codebook compression," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2282–2293, Apr. 2014.

[5] Y. Gao, R. Ji, W. Liu, Q. Dai, and G. Hua, "Weakly supervised visual dictionary learning by harnessing image attributes," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5400–5411, Dec. 2014.

[6] L. Zheng, S. Wang, and Q. Tian, "Coupled binary embedding for large-scale image retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3368–3380, Aug. 2014.

[7] W. Zhou, H. Li, R. Hong, Y. Lu, and Q. Tian, "BSIFT: Toward data-independent codebook for large scale image search," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 967–979, Mar. 2015.

[8] S. Zhang, Q. Tian, Q. Huang, W. Gao, and Y. Rui, "USB: Ultrashort binary descriptor for fast visual matching and retrieval," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3671–3683, Aug. 2014.

[9] X. Yang, X. Gao, and Q. Tian, "Polar embedding for aurora image retrieval," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3332–3344, Nov. 2015.

[10] G. Lavoué, "Bag of words and local spectral descriptor for 3D partial shape retrieval," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2011, pp. 41–48.

[11] T. Darom and Y. Keller, "Scale-invariant features for 3-D mesh models," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2758–2769, May 2012.

[12] R. Ji, L.-Y. Duan, J. Chen, T. Huang, and W. Gao, "Mining compact bag-of-patterns for low bit rate mobile visual search," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3099–3113, May 2014.

[13] X. Bai, C. Rao, and X. Wang, "Shape vocabulary: A robust and efficient shape representation for shape matching," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3935–3949, Sep. 2014.

[14] H. Tabia, H. Laga, D. Picard, and P.-H. Gosselin, "Covariance descriptors for 3D shape matching and retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4185–4192.

[15] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorization using co-occurrence, location and appearance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[16] O. A. B. Penatti, F. B. Silva, E. Valle, V. Gouet-Brunet, and R. D. S. Torres, "Visual word spatial arrangement for image retrieval and classification," *Pattern Recognit.*, vol. 47, no. 2, pp. 705–720, 2014.

[17] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2006, pp. 2169–2178.

[18] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang, "Spatial-bag-of-features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3352–3359.

[19] Z. Liu, H. Li, W. Zhou, R. Zhao, and Q. Tian, "Contextual hashing for large-scale image search," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1606–1614, Apr. 2014.

[20] L. Xie, Q. Tian, M. Wang, and B. Zhang, "Spatial pooling of heterogeneous features for image classification," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1994–2008, May 2013.

[21] X. Li, A. Godil, and A. Wagan, "Spatially enhanced bags of words for 3D shape retrieval," in *Proc. Int. Symp. Vis. Comput.*, 2008, pp. 349–358.

[22] G. Lavoué, "Combination of bag-of-words descriptors for robust partial shape retrieval," *Vis. Comput.*, vol. 28, no. 9, pp. 931–942, Sep. 2012.

[23] X. Li and A. Godil, "Exploring the bag-of-words method for 3D shape retrieval," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 437–440.

[24] C. Li and A. B. Hamza, "Intrinsic spatial pyramid matching for deformable 3D shape retrieval," *Int. J. Multimedia Inf. Retr.*, vol. 2, no. 4, pp. 261–271, 2013.

[25] A. M. Bronstein and M. M. Bronstein, "Spatially-sensitive affine-invariant image descriptors," in *Proc. Eur. Conf. Comput. Vis.*, vol. 6312. 2010, pp. 197–208.

[26] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape Google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graph.*, vol. 30, no. 1, Jan. 2011, Art. no. 1.

[27] R. Wessel and R. Klein, "Learning the compositional structure of man-made objects for 3D shape retrieval," in *Proc. Eurograph. Workshop 3D Object Retr.*, May 2010, pp. 39–46.

[28] L. Breiman, *Probability*. Philadelphia, PA, USA: SIAM, 1992.

[29] Y. Liu, H. Zha, and H. Qin, "Shape topics: A compact representation and new algorithms for 3D partial shape retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2025–2032.

[30] R. Ohbuchi, K. Osada, T. Furuya, and T. Banno, "Salient local visual features for shape-based 3D model retrieval," in *Proc. Int. Shape Modeling Appl.*, Jun. 2008, pp. 93–102.

[31] Z. Lian, A. Godil, and X. Sun, "Visual similarity based 3D shape retrieval using bag-of-features," in *Proc. Shape Modeling Int. Conf.*, Jun. 2010, pp. 25–36.

[32] R. Toldo, U. Castellani, and A. Fusiello, "Visual vocabulary signature for 3D object retrieval and partial matching," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2009, pp. 21–28.

[33] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[34] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1383–1392, 2009.

[35] M. Ovsjanikov, A. M. Bronstein, M. M. Bronstein, and L. J. Guibas, "Shape Google: A computer vision approach to invariant shape retrieval," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Sep./Oct. 2009, pp. 320–327.

[36] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proc. Eurograph. Symp. Geometry Process.*, Jun. 2003, pp. 156–165.

[37] B. Levy, "Laplace–Beltrami eigenfunctions towards an algorithm that 'understands' geometry," in *Proc. IEEE Int. Conf. Shape Modeling Appl.*, Jun. 2006, pp. 13–20.

[38] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 215–232, 2016.

[39] D. Zhang, J. Han, L. Jiang, S. Ye, and X. Chang, "Revealing event saliency in unconstrained video collection," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1746–1758, Mar. 2017.

[40] X. Yao, J. Han, G. Cheng, X. Qian, and L. Guo, "Semantic annotation of high-resolution satellite images via weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3660–3671, Jun. 2016.

[41] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 762–768.

[42] S. Savarese, J. Winn, and A. Criminisi, "Discriminative object class models of appearance and shape by correlatons," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2006, pp. 2033–2040.

[43] S. Savarese, A. DelPozo, J. C. Niebles, and L. Fei-Fei, "Spatial-temporal correlatons for unsupervised action classification," in *Proc. IEEE Workshop Motion Video Comput.*, Jun. 2008, pp. 1–8.

[44] J. Li, W. Wu, T. Wang, and Y. Zhang, "One step beyond histograms: Image representation using Markov stationary features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[45] B. Ni, S. Yan, and A. Kassim, "Contextualizing histogram," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1682–1689.
[46] M. Shi, R. Xu, D. Tao, and C. Xu, "W-tree indexing for fast visual word generation," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1209–1222, Mar. 2013.
[47] E. Kalogerakis, A. Hertzmann, and K. Singh, "Learning 3D mesh segmentation and labeling," *ACM Trans. Graph.*, vol. 29, no. 3, Jul. 2010, Art. no. 102.
[48] M. Hilaga, Y. Shinagawa, T. Kohmura, and T. L. Kunii, "Topology matching for fully automatic similarity estimation of 3D shapes," in *Proc. ACM SIGGRAPH*, 2001, pp. 203–212.
[49] R. Liu, H. Zhang, A. Shamir, and D. Cohen-Or, "A part-aware surface metric for shape analysis," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 397–406, 2009.
[50] L. Shapira, A. Shamir, and D. Cohen-Or, "Consistent mesh partitioning and skeletonisation using the shape diameter function," *Vis. Comput.*, vol. 24, no. 4, pp. 249–259, 2008.
[51] G. Mori, S. Belongie, and J. Malik, "Shape contexts enable efficient retrieval of similar shapes," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, vol. 1. Dec. 2001, pp. 723–730.
[52] D. Giorgi, S. Biasotti, and L. Paraboschi, "Shape retrieval contest 2007: Watertight models track," *SHREC Competition*, vol. 8, no. 7, pp. 1–11, 2007.
[53] Z. Lian *et al.*, "SHREC'10 track: Non-rigid 3D shape retrieval," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2010, pp. 101–108.
[54] H. Tabia, M. Daoudi, J.-P. Vandeborre, and O. Colot, "A new 3D-matching method of non-rigid and partially similar models using curve analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 852–858, Apr. 2011.
[55] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[56] S. Marini, L. Paraboschi, and S. Biasotti, "Shape retrieval contest 2007: Partial matching track," in *Proc. SHREC Conjunct IEEE Shape Modeling Int.*, Jun. 2007, pp. 13–16.

Zhizhong Han is currently pursuing the Ph.D. degree with Northwestern Polytechnical University, China. He is majored in pattern recognition and machine intelligence. His research interests include machine learning, pattern recognition, and digital geometry processing.

Zhenbao Liu (M'11) received the bachelor's and master's degree from Northwestern Polytechnical University in 2001 and 2004, respectively, and the Ph.D. degree from the College of Systems and Information Engineering, University of Tsukuba, Tsukuba, Japan, in 2009. , He was a Visiting Scholar with Simon Fraser University, Canada, in 2012. He is currently a Professor with Northwestern Polytechnical University, China. He has published approximately 50 papers in major international journals and conferences. His research interests include pattern recognition, computer vision, and shape analysis.

Chi-Man Vong (M'09–SM'14) received the M.S. and Ph.D. degrees in software engineering from the University of Macau in 2000 and 2005, respectively. He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. His research interests include machine learning methods and intelligent systems.

Yu-Shen Liu received the B.S. degree in mathematics from Jilin University, China, in 2000, and the Ph.D. degree with the Department of Computer Science and Technology, Tsinghua University, China, in 2006. He was a Post-Doctoral Researcher with Purdue University from 2006 to 2009. He is an Associate Professor with the School of Software, Tsinghua University, Beijing, China. His research interests include shape analysis pattern recognition, machine learning, and semantic search.

Shuhui Bu (M'09) received the master's and Ph.D. degrees from the College of Systems and Information Engineering, University of Tsukuba, Japan, in 2006 and 2009. He was an Assistant Professor with Kyoto University, Japan, from 2009 to 2011. Currently, he is a Professor with Northwestern Polytechnical University, China. He has published approximately 40 papers in major international journals and conferences. His research interests are concentrated on computer vision and robotics.

Junwei Han (M'12–SM'15) is a currently a Professor with Northwestern Polytechnical University, Xi'an, China. He received the Ph.D. degree in pattern recognition and intelligent systems from the School of Automation, Northwestern Polytechnical University, in 2003. His research interests include multimedia processing and brain imaging analysis. He is an Associate Editor of the IEEE TRANSACTIONS ON HUMAN–MACHINE SYSTEMS, *Neurocomputing*, and *Multidimensional Systems and Signal Processing*.

C. L. Philip Chen (S'88–M'88–SM'94–F'07) received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1988. He was a Tenured Professor, a Department Head, and an Associate Dean in two different universities at U.S. for 23 years. He is currently the Dean of the Faculty of Science and Technology, University of Macau, Macau, China, and a Chair Professor with the Department of Computer and Information Science.