Pattern Recognition ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Fisher encoding of differential fast point feature histograms for partial 3D object retrieval

Michalis A. Savelonas [a,b,*], Ioannis Pratikakis [a,b], Konstantinos Sfikas [a,c]

[a] ATHENA Research and Innovation Center, Branch of Xanthi, GR-67100 Kimmeria, Xanthi, Greece
[b] Department of Electrical and Computer Engineering, Democritus University of Thrace, Building B′, Panepistimioupolis, GR-67100 Kimmeria, Xanthi, Greece
[c] Department of Computer and Information Science, NTNU, Trondheim, Norway

ARTICLE INFO

Article history:
Received 20 April 2015
Received in revised form
23 October 2015
Accepted 4 February 2016

Keywords:
3D object retrieval
Partial matching
Local descriptors
Fisher encoding

ABSTRACT

Partial 3D object retrieval has attracted intense research efforts due to its potential for a wide range of applications, such as 3D object repair and predictive digitization. This work introduces a partial 3D object retrieval method, applicable on both point clouds and structured 3D models, which is based on a shape matching scheme combining local shape descriptors with their Fisher encodings. Experiments on the SHREC 2013 large-scale benchmark dataset for partial object retrieval, as well as on the publicly available Hampson pottery dataset, demonstrate that the proposed method outperforms seven recently evaluated partial retrieval methods.

1. Introduction

Several methods have been proposed for partial 3D object retrieval (P3DOR) in the last five years [1]. Still, as noted in the recent comparative study of Sipiran et al. [2], the problem is very challenging and far from being solved. Apart from limitations in retrieval performance, most existing P3DOR methods require structured 3D data as input, despite the fact that a variety of sensors (e.g. Velodyne spinning LIDAR) provide their output in point cloud form.

P3DOR is usually based on shape matching in a local fashion. Such an approach is motivated by the observation that a partial query and its corresponding complete model are by definition expected to be locally similar. However, apart from localized shape information, the global shape of each partial query is intuitively expected to provide an additional cue for P3DOR. The bag-of-visual-words (BoVW) framework provides a tool for deriving global shape representations from local shape descriptors and has already been successfully employed for global or partial 3D object retrieval [3,4].

Despite their advantages, the existing BoVW-based P3DOR methods [1] have not sufficiently addressed the partiality of the query object since they tend to uniformly consider the frequency of occurrence of all available codewords. However, codewords characterizing a complete 3D model may not characterize its associated partial query, as schematically described in Fig. 1. Note that apart from the case illustrated in Fig. 1, where some codewords associated with the target model are completely absent from the partial query, another potential case considered concerns some codewords with a frequency of occurrence which differs between the partial query and the target model. A more selective weighting of codeword distribution, adapted to the partiality of each query, is expected to enhance the performance in the partial retrieval task.

Lavoué [4] introduced a BoVW-based P3DOR alternative, which combines the standard and the spatially sensitive BoVW approach. However, this method relies on the basic k-means BoVW variant, whereas it employs a simple $L1$-based distance for shape matching, which uniformly considers the frequency of occurrence of all available codewords. Bronstein et al. [3] proposed another spatially sensitive BoVW-based P3DOR method which employs the scale invariant heat kernel signature (SI-HKS) [5] descriptor. Instead of using uniform distance metrics, similarity-sensitive hashing (SSH) is used, in order to adjust weighting by considering the training set as a whole. Still, such an approach is not adaptive to the codeword frequency of occurrence in each specific partial query. Moreover, both methods of Lavoué and Bronstein et al. require structured query input and cannot be applied in the case of point clouds, without employing a point cloud-to-mesh conversion stage.

The BoVW-based work of Li et al. [6] recently employed a scheme for adaptive weighting of shape distance terms related to multiple descriptors, which is based on particle swarm optimization (PSO) [7]. In this scheme, PSO is performed offline and accordingly the resulting weights are not adapted to the codeword

* Corresponding author at: ATHENA Research and Innovation Center, Branch of Xanthi, GR-67100 Kimmeria, Xanthi, Greece. Tel.: +30 25410 79577.
E-mail address: msavelonas@ieee.org (M.A. Savelonas).

http://dx.doi.org/10.1016/j.patcog.2016.02.003
0031-3203/

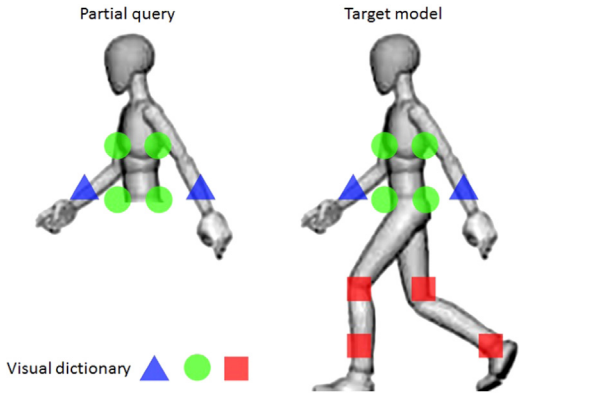

Fig. 1. Some visual words of the target model (e.g. red square) can be underrepresented in the partial queries. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

distribution of each particular partial query. Moreover, the weights determined by PSO are only locally optimal [8].

Fisher encoding [9,10] improves the retrieval performance over standard BoVW, by means of difference encoding and subtracting the mean of a Gaussian fit to all observations. The resulting measures comprise the Fisher vector, which has been shown to provide a generalized, enhanced version of a variant of *k*-means-based BoVW: the vector of locally aggregated descriptors (VLAD) [11]. As demonstrated by Jegou et al. [11], information which is not distinctive for each sample (i.e. it is also shared by other samples) is approximately discarded from the Fisher vector. Moreover, Fisher encoding requires much smaller vocabularies and has been associated with enhanced retrieval performance [11,12]. It has also been supported in a recent comparative study [13], when compared to the basic *k*-means-based BoVW and the support vector encoding [14].

The proposed P3DOR method, which can be applied on both point clouds and structured 3D models, is based on a hybrid shape matching scheme, defined so as to account for both local and global shape similarities, as well as to address the partiality of the query object. We introduce the differential fast point feature histogram (dFPFH), which extends the FPFH point cloud descriptor [15] in order to more accurately capture local geometry transitions. Global shape similarity is estimated by means of a weighted distance of Fisher vectors. A non-uniform weighting of both local and global distances is employed in order to reduce the influence of the most dissimilar pairs, following the intuition that certain parts or codewords of the target object can be missing or underrepresented in the partial query. Overall, local and global distances, which are derived for multiple scales, are combined to obtain a ranked list of the most similar complete 3D objects.

Experimental evaluation on the SHREC 2013 large-scale benchmark dataset for partial object retrieval [16] supports the proposed method against five recently evaluated P3DOR methods [2], with respect to standard retrieval performance measures. Additional experimentation on the publicly available Hampson pottery dataset provides a real-world application scenario in the cultural heritage (CH) domain, along with extra favorable comparisons with two recent P3DOR applications [17,18], that have been evaluated on this dataset. Two preliminary conference versions of this work have been accepted for publication. A variant which only uses global Fisher vectors and uniform weighting for P3DOR [17] and a variant which uses both local descriptors and Fisher vectors [19]. Compared to the latter version, this paper provides a more complete presentation of the related literature and methodology, whereas it incorporates more detailed experimentation and analysis, including evaluation of computational cost and additional experiments with real partial queries, obtained by the Breuckmann Optoscan scanner.

The remainder of this paper is organized as follows: Section 2 presents related work on local shape descriptors and P3DOR. Sections 3 and 4 describe the shape representation and matching components of the proposed method, respectively. The experimental evaluation including comparisons with state-of-the-art P3DOR methods is presented in Section 5. Conclusions and future research directions are discussed in Section 6.

2. Related work

This section provides an overview of local shape descriptors, as well as of state-of-the-art in P3DOR.

2.1. Local shape descriptors

There has already been a considerable amount of research in local 3D shape descriptors. One of the most popular is the spin-image [20], which has been widely applied on both structured and unstructured data. A spin-image of an oriented point is a 2D representation of its surrounding surface, which is constructed on a pose-invariant 2D coordinate system by accumulating the coordinates of neighboring points. However, Steder et al. [21] have shown that range value patches are more reliable than spin-images. Spin-images also do not explicitly take empty space into account. For example, in the case of a square plane, the spin images for points in the center and for corners would be identical.

Ohbuchi et al. [22] introduced the multiple orientation depth Fourier transform descriptor (MODFD) in the context of an appearance-based 3D object retrieval method. MODFD encompasses model projections from 42 viewpoints so as to cover all possible view aspects. The method of Ohbuchi et al. is designed for ill-defined model representations, most notable of which is the polygon soup.

Normal aligned radial features (NARF) have been introduced by Steder et al. [23] as an interest point extraction method, along with a feature descriptor in 3D range data. The interest point extraction method has been designed with two specific goals: (i) the selected points are supposed to be in positions where the surface is stable, so as to ensure a robust estimation of the normal, and where there are sufficient changes in the immediate vicinity, (ii) the outer shapes of objects as seen from a certain perspective are used, considering that the focus is on partial views.

Kernel descriptors provide a principled approach for converting pixel attributes to patch-level features. They were initially proposed for RGB images, and have not been used for depth maps and point clouds until the work of Bo et al. [24]. In this work, the match kernel framework involves the following main steps: (i) pixel attributes are defined, (ii) match kernels are designed to measure the similarities of image patches, based on the defined attributes, and (iii) approximate, low-dimensional match kernels are determined. Besides using gradient and local binary patterns in their framework, the authors have also developed size, PCA and spin-based kernel descriptors, which capture diverse yet complementary cues.

Point feature histograms (PFH) [25] have been introduced as point cloud descriptors. They are calculated over a neighborhood centered at each point of interest, encoding patterns of point distances. This type of information is the closest one to point coordinates, as provided in raw point cloud input data. PFH [25] and its more efficient sibling, fast PFH (FPFH) [15], have been proposed by Rusu et al. as pose invariant local shape descriptors, which represent underlying surface model properties. They rely upon geometrical relations between k nearest neighbors, defined by means of 3D point coordinates (x, y, z) along with estimated surface normals (nx, ny, nz).

Most local shape descriptors, with MODFD being an exception, have not been initially proposed for 3D object retrieval. The most

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

usual application addressed is object recognition in scenes, whereas PFH has been proposed for registration. In principle, there is nothing prohibitive in applying the same descriptors for 3D object retrieval.

2.2. Partial 3D object retrieval

Most P3DOR methods can be roughly classified as: (i) view-based, with prominent examples in [26–28], (ii) part-based [29,30], (iii) BoVW-based [3,4,6], and finally (iv) hybrid methods combining these three main paradigms [18,31]. Five recent P3DOR methods, encompassing elements of these categories, were recently compared on SHREC 2013 large-scale partial retrieval benchmark:

- Two methods incorporating 2D-3D alignment [32], as well as entropy-based adaptive view clustering [33]. We refer to these methods as 'SBR2D-3D' and 'SBR-VC', respectively. SBR2D-3D employs a sketch-based algorithm based on 3D model features and 2D relative shape matching. For each target model, it precomputes the view context and relative shape context features of a set of densely sampled views. For the query model, it generates its silhouette feature view and then similarly computes its view context and relative shape context features. Based on the view context of the silhouette feature view and the sample views of the query model, it performs a query-target alignment by short listing several model views to correspond with the silhouette feature view. Finally, query-target matching is based on the shape context matching between the silhouette feature view of the query model and the candidate sample views of the target 3D model. SBR-VC employs a visual complexity metric, which is based on the viewpoint entropy distribution of multiple sample views of the 3D model. The metric is used to adaptively decide the number of the representative views to perform fuzzy k-means view clustering. This is followed by a more accurate and parallel relative shape context matching.
- Two methods using data-aware partitioning [34] and BoVW [2]. We refer to these methods as 'Data-aware' and 'S-BoVW', respectively. The data-aware method employs feature detection by means of Harris 3D keypoints in the Euclidean space using adaptive clustering [35]. The minimum enclosing spheres of the detected keypoints is used to define model partitions. Both local and global shape representations are derived by means of the DESIRE descriptor [36], with the former derived from shape partitions and the latter derived from overall shape. S-BoVW starts from local point cloud descriptors, which include rectangular and polar spin images [37], shape contexts [38] and FPFH [15], in order to calculate the codewords using k-means clustering.
- A method proposed in [16], which uses spin images and signature quadratic form distance. We refer to this method as 'SQFD'. SQFD starts from Harris 3D keypoints to build a feature set composed of normalized local descriptors. Next, a local clustering algorithm [39] is applied to obtain a set of representative descriptors. Shape matching is performed with the signature quadratic form distance.

In addition, two recent P3DOR methods have been applied on the publicly available Hampson pottery dataset:

- A panoramic, view-based method, proposed in [18]. We refer to this method as 'Panoramic'. Panoramic uses an enhanced variant of the dense scale invariant feature transform (SIFT) descriptor [40], calculated on panoramic views of each 3D model. The resulting feature vectors are encoded by means of k-means-based BoVW.
- A method proposed in [17], which is a preliminary conference version of this work and addresses partial retrieval by means of Fisher encoding in a purely global fashion. We refer to this method as 'Global Fisher'. Global Fisher applies an adaptive variant of the FPFH, which depends on point cloud density, followed by Fisher encoding.

3. Shape representation

PFH represents a point cloud in a local fashion by analyzing the relationships between the normals of pairs of neighboring points. It has been originally defined as follows: (i) for each point p, all of its neighbors enclosed in the sphere of a given radius r are selected (r-neighborhood), (ii) for every pair of points p_i and p_j ($i \neq j$) in the r-neighborhood of p and their PCA-estimated normals n_i and n_j (p_i being the point with a smaller angle between its associated normal and the line connecting the points) [41], a Darboux uvn frame ($u = n_i$, $v = (p_i - p_j) \times u$, $n = u \times v$) is defined and the angular variations of n_i and n_j are computed as follows: $\alpha = u \cdot n_j$, $\phi = u \cdot (p_j - p_i)/\|p_j - p_i\|$, $\theta = \arctan(w \cdot n_j, u \cdot n_j)$. The histograms which constitute the PFH descriptor have b binning subdivisions for each one of α, ϕ and θ angle, where b is implementation-dependent. This leads to a histogram size equal to $3b$.

The fast point feature histogram (FPFH) [15] has been proposed in order to accelerate PFH computations by employing a subset of neighboring points for histogram calculation. For a given query point p_q, its single point feature histogram (SPFH) values are first estimated by creating pairs between itself and its r-neighbors. This is repeated for all points in the object, followed by re-weighting of the SPFH values using the SPFH values of r-neighbors, in order to create the FPFH for p_q.

In this work, FPFH is extended in order to capture local geometric transitions by measuring the differences in feature histograms, which are associated with concentric spheres. The proposed FPFH extension, namely differential FPFH (dFPFH) has $6b$ bins (the standard $3b$ bins associated with a concentric sphere of radius r plus the histogram of $3b$ bins, which quantifies the transitions of FPFH in a ribbon around r):

$$\mathrm{dFPFH}(q_p, r) = [\mathrm{FPFH}(q_p, r)\ \ \Delta\mathrm{FPFH}(q_p, r)] \tag{1}$$

where $\Delta\mathrm{FPFH}(q_p, r) = \mathrm{FPFH}(q_p, r_{outer}) - \mathrm{FPFH}(q_p, r_{inner})$. Fig. 2 provides an intuitive explanation of dFPFH. In the case of the smooth surface of the cup illustrated in Fig. 2(a), the FPFH histograms of the two concentric spheres are rather similar, resulting in histogram differences approximating zero. On the other hand, the irregularity of the ant surface in Fig. 2(b) is reflected in much larger differences of the FPFH histograms.

Each object may well be scanned using various types of scanning equipment, from various distances or with different settings, resulting in multiple point cloud densities. Aiming to alleviate the effects of this variability on retrieval performance, we introduce two extra preprocessing steps: (i) the input object is downsampled with voxelized grid filtering, in which all points within a voxel are approximated by their centroid. Multiscale information can be derived by considering multiple voxel sizes, (ii) the neighborhood radius r considered in dFPFH calculations is adaptively estimated for each point cloud as a linear function of the mean point distance over all r-neighborhoods. It should also been noted that dFPFH is scale invariant, as is the case with its originating descriptor, PFH. The latter has been shown by Rusu et al. [42].

4. Shape matching

The dFPFH-based shape representation described in the previous section is used for shape matching. A hybrid scheme is proposed,

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

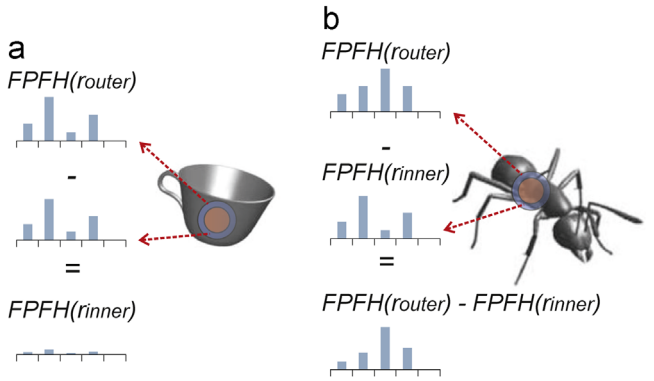

Fig. 2. A schematic representation of dFPFH: (a) smooth surfaces result in similar FPFH histograms for the concentric spheres (FPFH(r_{outer}) ≈ FPFH(r_{inner})) and histogram differences approximating zero, (b) irregular surfaces result in much larger differences of the FPFH histograms.

which incorporates the results of two distinct processes: (i) local shape similarity assessment by averaging the minimum weighted distances associated with pairs of dFPFH values calculated on the partial query and the target object respectively, (ii) global shape similarity assessment by means of a weighted distance of Fisher vectors.

4.1. Local shape similarity assessment

Aiming to assess the local similarity between the partial query object Q and each complete object T from the repository, we define the mean-minimum distance dm as follows:

$$dm(Q,T) = \text{mean}_{q_p \in Q}(\min_{t_p \in T}(L_{d1}(q_p, t_p))) \tag{2}$$

where q_p is a point of Q, t_p is a point of T, N and M denote the number of points of Q and T, respectively, whereas L_{d1} denotes the Manhattan distance L_1 between the dFPFH histograms of q_p and t_p. This strategy is justified by considering that the similarity of the partial query Q with T is not associated with the distance of histograms of all possible pairs of points (q_p, t_p), but only with the distance of pairs of histograms of similar points. The average of this distance forms dm. We selected L_1 over other distance alternatives (e.g. L_2) based on experimentation.

In addition, considering that in Eq. (2) the minimum of L_{d1} for each q_p depends on a single pair of points, we introduce the weighted mean-minimum distance dm_w, in which L_{d1} is replaced by a weighted average of the k smaller distances:

$$dm_w(Q,T) = \underset{q_p \in Q}{\text{mean}}\left[(1/k)\sum_{i=1,2,\ldots,k} w_i L_{d1}(q_p, t_p(i))\right] \tag{3}$$

where $t_p(i)$, $i = 1, 2, \ldots, k$ are the first k points of object T, when all points of T are sorted in increasing order with respect to their distance from q_p. The weights $w_i = (1-(i/k))$ are linearly decreasing, starting from the pair with the smaller distance ($i=1$). This weighting amplifies the influence of the more similar pairs of points, among the selected k pairs, whereas it ensures a smooth transition to zero, which is the weight associated to those points which are not among the k selected. We selected linearly decreasing weighting over other alternatives (e.g. quadratic decrease), since it is associated with less calculations and leads to comparable results, as found after experimentation.

4.2. Global shape similarity assessment

Aiming to assess the global similarity between the partial query object Q and each object T from the repository, we employ Fisher encoding, extending the purely global Fisher approach that has been proposed in [17]. The use of Fisher encoding instead of standard BoVW approaches has been experimentally supported in a recent comparative study [13]. A Gaussian mixture model (GMM) is estimated from local shape descriptors using an expectation maximization algorithm. The resulting GMM defines the visual codebook used [9,10].

Given a set of N dFPFH descriptors $\mathbf{x}_1, \ldots, \mathbf{x}_N \in R^D$, which are used for training, a GMM $p(\mathbf{x}|\theta)$ is the probability density on R^D given by

$$p(\mathbf{x}|\theta) = \sum_{k=1}^{K} p(\mathbf{x}|\mu_k, \Sigma_k)\pi_k \tag{4}$$

$$p(\mathbf{x}|\mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D \det\Sigma_k}} e^{-\frac{1}{2}(\mathbf{x}-\mu_k)^T \Sigma_k^{-1}(\mathbf{x}-\mu_k)} \tag{5}$$

where K is the number of Gaussian components used, θ is the vector of model parameters $(\pi_1, \mu_1, \Sigma_1, \ldots, \pi_K, \mu_K, \Sigma_K)$, including the prior probability values $\pi_k \in R_+$ (which sum to one), the means $\mu_k \in R^D$, and the positive definite covariance matrices $\Sigma_k \in R^{D\times D}$ of each Gaussian component. The covariance matrices are assumed to be diagonal, so that the GMM is fully specified by $(2D+1)K$ scalar parameters. Soft data-to-cluster assignments extend the binary assignments to k-means in basic BoVW and can be defined as

$$q_{ki} = \frac{p(\mathbf{x}_i|\mu_k, \Sigma_k)\pi_k}{\sum_{j=1}^{K} p(\mathbf{x}_i|\mu_j, \Sigma_j)\pi_j}, \quad k = 1, \ldots, K \tag{6}$$

Fisher encoding [10] captures the average first- and second-order differences between the local descriptors and the centers of a GMM, which can be thought of as a visual codebook. For the kth GMM, where $k = 1, \ldots, K$, the following vectors are defined:

$$\mathbf{u_k} = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik} \Sigma_k^{-1/2}(\mathbf{x}_i - \mu_k) \tag{7}$$

$$\mathbf{v_k} = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik}[(\mathbf{x}_i - \mu_k)\Sigma_k^{-1}(\mathbf{x}_i - \mu_k) - 1] \tag{8}$$

In $\mathbf{u_k}$ and $\mathbf{v_k}$, the approximate location of the descriptors in each region is encoded, relatively to the mean and the variance, respectively. The division by $\sqrt{2\pi_k}$ can be interpreted as a BoVW inverse document frequency term: the weights of frequent descriptors are reduced [11].

The Fisher encoding of the set of local feature vectors is then given by the concatenation of $\mathbf{u_k}$ and $\mathbf{v_k}$ for all K components, giving an encoding of size $2DK$

$$\mathbf{f} = [\mathbf{u}_1^T, \mathbf{v}_1^T, \ldots \mathbf{u}_K^T, \mathbf{v}_K^T] \tag{9}$$

In the case of the basic BoVW, $2D$ times fewer visual words are required to obtain a vector of the same length.

Both vectors, $\mathbf{u}_k$ and $\mathbf{v}_k$, have size equal to the size of the local feature vector, i.e. $6b = 66$, considering that $b = 11$ binning subdivisions are used. It can be derived from Eq. (9) that the resulting Fisher vector $\mathbf{f}$ has size equal to $2 \times 66 \times K = 132 \times K$.

Intuitively, the originating complete object of Q and its most similar complete object T are expected to have similar Fisher vectors. Therefore, it is natural to assume that the most dissimilar pairs of Fisher components between Q and T are associated with the GMMs that are over-represented in those parts of T that are missing from Q. Starting from this consideration, we define the weighted Fisher vector distance dF_w, in a similar fashion to dm_w, as:

$$dF_w(Q,T) = \left((1/K)\sum_{j=1,2,\ldots,K} wf_j L_{f1}(Q(j), T(j)\right) \tag{10}$$

where $L_{f1}(Q(j), T(j))$ is the L_1 distance of the respective Fisher vectors $\mathbf{f}_Q(j)$, $\mathbf{f}_T(j)$. The pairs $(\mathbf{f}_Q(j), \mathbf{f}_T(j))$ are sorted in increasing order with respect to their distance. The weights $wf_j = (1-(j/K))$ are linearly decreasing, starting from the pair with the smaller distance. This weighting reduces the influence of the more dissimilar pairs of

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

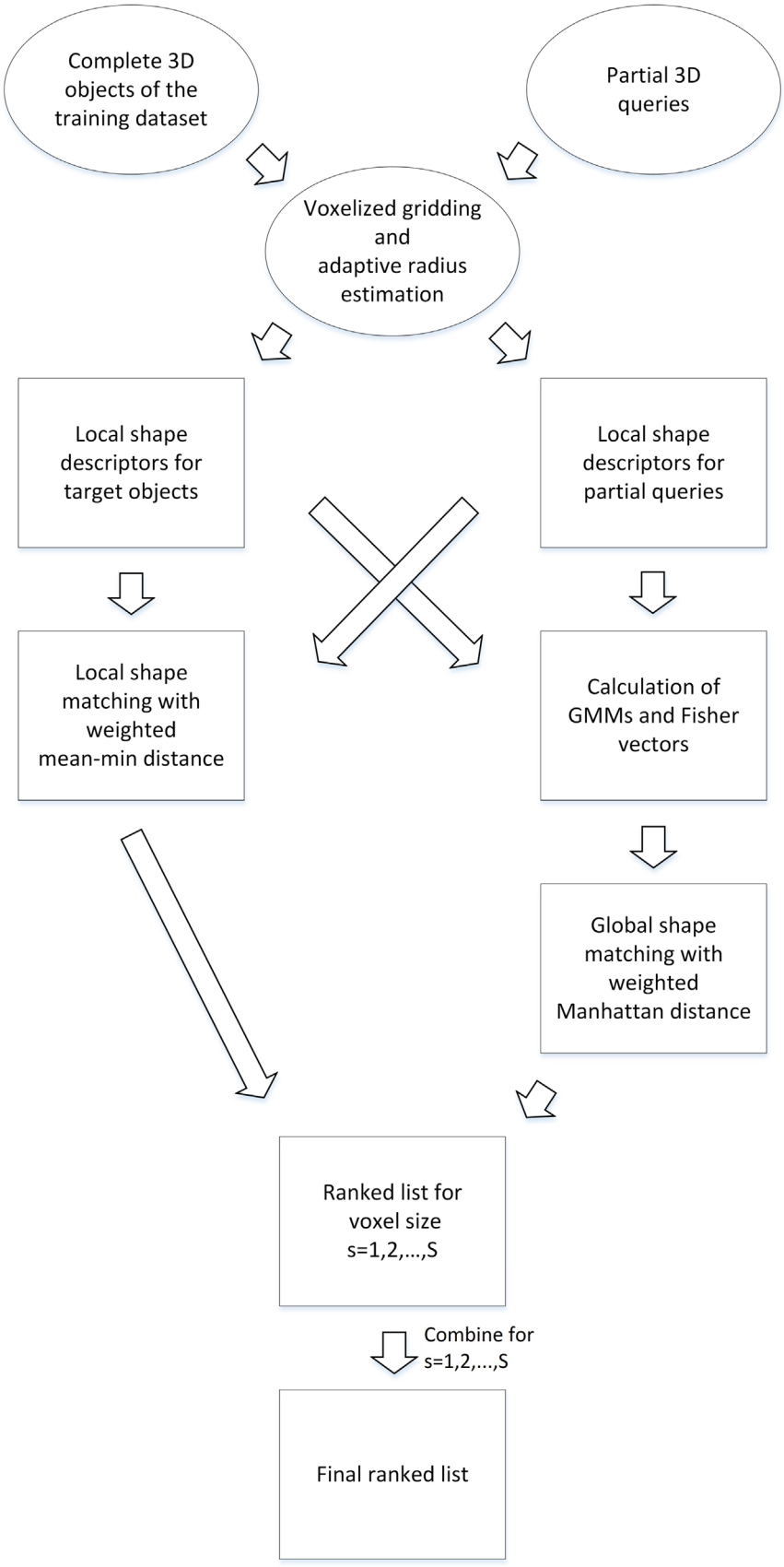

Fig. 3. The pipeline of the proposed method.

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

Fisher components in the distance calculation. As is the case with Eqs. (2) and (3), the use in Eq. (10) of both L_1 and linearly decreasing weighting is supported by preliminary experimentation.

4.3. Hybrid shape similarity assessment

For each voxel size v_s considered in the voxelized gridding filtering step described in Section 3 (where $s = 1, 2\ldots, S$, with S being the number of voxel sizes considered), the hybrid distance $d_{hybrid}(Q, T, s)$ is a weighted sum of $dm_w(s)$ and $dF_w(s)$, defined according to Eqs. (2) and (10) by substituting dm_w and dF_w with $dm_w(s)$ and $dF_w(s)$, respectively:

$$d_{hybrid}(Q, T, s) = w_o dm_w(s) + dF_w(s) \tag{11}$$

where w_o adjusts the relative influence of local and global shape matching distances. The overall multiscale distance $d_{multiscale}$, which is used to obtain a ranked list of complete 3D objects, is a weighted sum:

$$d_{multiscale}(Q, T) = \sum_{s=1,2,\ldots,S} w_s d_{hybrid}(Q, T, s) \tag{12}$$

where the weights w_s adjust the relative influence of each scale s considered.

Fig. 3 shows the distinct components of the proposed pipeline for partial 3D object retrieval.

5. Evaluation

Experiments are performed on two publicly available benchmark datasets. The first dataset has been used in SHREC 2013 track for large scale partial 3D object retrieval [16]. The target set has been created from 360 shapes, organized into 20 classes of 18 objects per class. On the other hand, the process of range scan acquisition from the objects of the target set has been simulated in order to obtain a set of partial views. This process results in 7200 queries, associated with varying levels of partiality. Fig. 4 shows some samples from the target set of SHREC 2013. Recently, a more extensive comparison of five state-of-the-art methods has been performed on the same dataset [2].

The second benchmark dataset used for the evaluation is related to the CH domain and consists of 3D pottery models originating from the Virtual Hampson Museum collection (http://hampson.cast.uark.edu). It is publicly available and has already been used for the evaluation of two state-of-the-art methods [17,18]. The dataset consists of 384 models classified to 6 distinct geometrically defined classes (bottle, bowl, jar, effigy, lithics and others), which can further be divided in 23, more precisely defined, sub-classes. 21 partial queries have been artificially created by slicing and cap filling complete 3D models. The partial queries used in our experiments have a reduced surface compared to the original 3D object, which is associated with 25% partiality. Fig. 5 shows some examples of pottery models used in this dataset. Apart from experiments with artificially created queries, additional

Fig. 4. Samples of the SHREC 2013 benchmark dataset [2].

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

Fig. 5. Example 3D models of the pottery dataset used (http://hampson.cast.uark.edu).

experiments are performed with real queries, obtained with Breuckmann Optoscan scanner. For the creation of the real queries, derivative vessels were constructed by a professional potter, using the Hampson models as a template.

Experimental evaluation is based on precision–recall (P–R) plots and five quantitative measures: nearest neighbor (NN), first tier (FT), second tier (ST), discounted cumulative gain (DCG) and mean average precision (MAP). More details on these measures can be found in [16,18].

The proposed method has been developed on a hybrid Matlab/C++ architecture. The experiments have been performed on an Intel Core i7 workstation, operating at 3.5 GHz with 16 GB of RAM.

Interestingly, it has been observed that by separately employing local and global shape similarity, the retrieval performance in SHREC 2013 is significantly lower (FT approximately equal to 15% and 18%, respectively) than the one obtained by the proposed hybrid approach (FT 28%), verifying that complementary information is derived from these distinct processes.

Based on preliminary experiments, parameter settings have been determined as follows: the linear coefficients adaptively associating the radii of the concentric spheres of dFPFH to the mean point distance are $r=2.7$, $r_{outer}=13.6$ and $r_{inner}=13.1$, respectively (Eq. (1)). In addition, $k=3$ (Eq. (3)) and $K=10$ GMMs were found to be sufficient for the construction of the visual codebook, leading to Fisher vectors of $2\times 66\times 10=1320$ components. k-means pre-clustering by means of Lloyds' variant [43] has been used to initialize GMM construction. The signed square root function has been applied to the resulting Fisher vectors, followed by L_2 normalization. The weight w_o (Eq. (11)) has been set to 0.4. Finally, the number of scales considered is $S=3$, with respective voxel sizes equal to 0.1, 0.3 and 0.5 and weights $w_s=0.4,\ 1.0,$ and 0.4, respectively (Eq. (12)).

Based on our preliminary experiments, we also performed a parameter sensitivity analysis. Fig. 6 presents the FT obtained for experiments on both SHREC 2013 and Hampson, when K, k, w_o and r_o vary. From Fig. 6 it can be derived that the optimal settings, as well as the pattern of dependency, are rather consistent between the two datasets. In the case of K and k, the optimal values are almost identical (the slight differences in the optimal value of K are associated with marginal effects on retrieval performance). It can also be derived that the proposed method is not sensitive in the exact value of w_o and r_o, since the FT obtained in both SHREC 2013 and Hampson, for a wide range of values, allows the proposed method to outperform state-of-the-art, as it can be seen by comparing: (i) the FT illustrated in Fig. 6e and g, with the FT presented in Table 1 (comparisons with state-of-the-art in SHREC 2013), (ii) the FT illustrated in Fig. 6f and h, with the FT presented in Table 2 (comparisons with state-of-the-art in Hampson). Moreover, as is the case for w_o, we found that for triples of w_s in the range [0.1,1.0], the FT may change up to approximately 2%. Similarly, r_{outer} and r_{inner} follow the same pattern with r_o. Finally, we found that weighting as defined in Eq. (3) and (10) allows a performance boost of approximately 1.5–2.5% with respect to FT, when compared with uniformly weighted L_1-based retrieval. Overall, it can be derived that when working with new datasets, the parameter settings used here are expected to lead to nearly optimal retrieval performance.

Table 1 presents the retrieval performance, as quantified by NN, FT, ST and MAP, which was obtained by the proposed method and five state-of-the-art methods on SHREC 2013 benchmark dataset. It can be noticed that the proposed method achieves the highest performance with respect to all metrics. Fig. 7 illustrates the average P–R scores for all retrieval methods. The proposed method and the data-aware method obtain the highest precision values, with each of these two methods having the advantage for different recall ranges.

Table 2 presents the retrieval performance, as quantified by NN, FT, ST and DCG, which was obtained by the proposed method and two state-of-the-art methods on the Hampson pottery dataset. It should be noted that in this case we use DCG instead of MAP, since this measure was used for the evaluation of the Panoramic [18] and Global Fisher [1] methods. In addition, an accuracy of three decimal digits is maintained, as in these works. Finally, we present results obtained on queries associated with 25% partiality. The proposed method achieves the highest retrieval performance with respect to all measures considered. This is verified in Fig. 8, which illustrates the average P–R scores for all retrieval methods.

Fig. 9 illustrates example ranked lists obtained in the case of the Hampson pottery dataset.

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

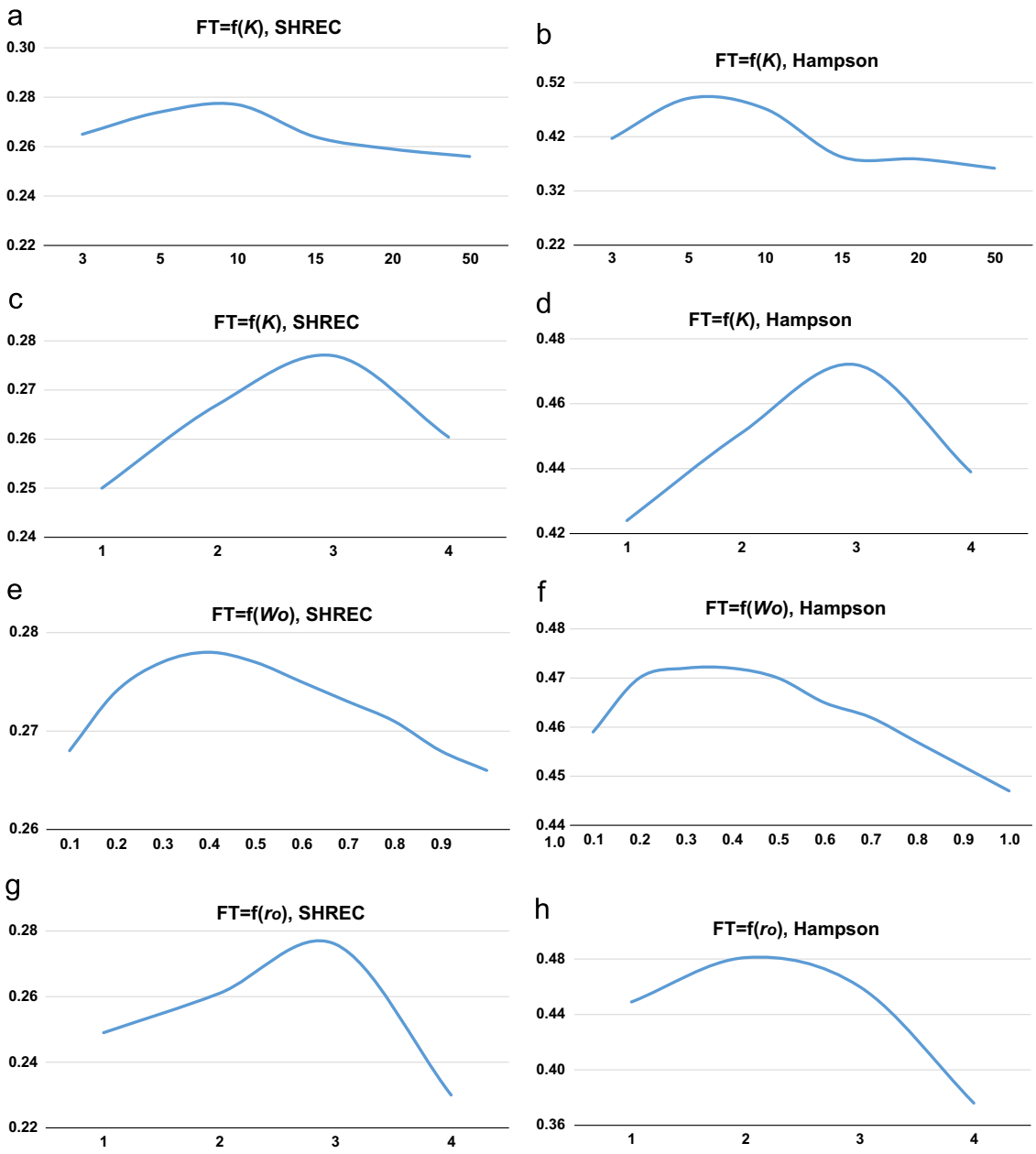

Fig. 6. Retrieval performance (FT) obtained on SHREC 2013 and Hampson datasets, for varying values of: (a and b) K, (c and d) k, (e and f) w_o and (g and h) r_o.

Table 1
The results of the proposed method, along with 5 state-of-the-art methods on SHREC 2013 benchmark dataset.

Method	NN	FT	ST	MAP
Proposed	**0.3856**	**0.2772**	**0.2135**	**0.2851**
SBR-2D-3D	0.3535	0.2290	0.1808	0.2455
SBR-VC	0.3218	0.2065	0.1638	0.2199
Data-aware	0.3457	0.2495	0.2088	0.2836
Polar spin images	0.0931	0.0809	0.0768	0.0968
SQFD	0.3108	0.2043	0.1576	0.1978

Experiments were also performed with 25 real queries obtained by using the Breuckmann Optoscan scanner, in order to demonstrate the applicability of the proposed method on a real digitization scenario.[1] Table 3 presents the retrieval performance and Fig. 10 illustrates the average P–R scores obtained for both

[1] This set will be publicly available.

Table 2
The results of the proposed method, along with two state-of-the-art methods on the Hampson pottery dataset.

Method	NN	FT	ST	DCG
Proposed	**0.952**	**0.460**	**0.642**	**0.778**
Global Fisher	**0.952**	0.320	0.461	0.694
Panoramic	0.619	0.416	0.626	0.721

23-class and 6-class classifications. These results show that although the proposed method outperforms state-of-the-art in two benchmark datasets, still cannot obtain high retrieval performance in a certain real-world scenario. In this respect, the remark in the comparative study of Sipiran et al. [2], that P3DOR is very challenging and open to future solutions, is still valid.

The time required for the offline calculation of Gaussian mixture models depends on dataset size. In the case of Hampson pottery dataset, this time is approximately 1 min. For each 3D model, local descriptor calculation requires 781 ms, Fisher vectors calculation

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

requires 2749 ms, distance matrix calculation requires 482 ms and the calculation of the ranked list requires 29 ms. In total, the online process for each partial query is approximately 4 s.

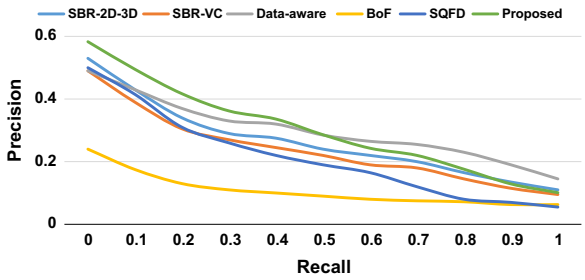

Fig. 7. Average P–R for all retrieval methods applied on SHREC 2013.

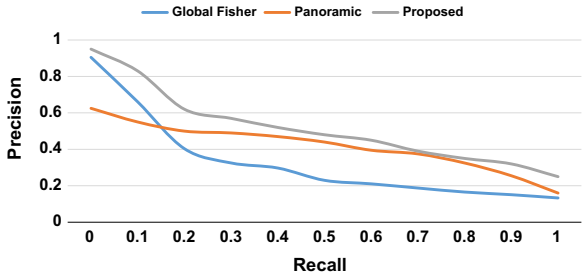

Fig. 8. Average P–R for all retrieval methods applied on the Hampson pottery dataset, for queries of 25% partiality.

It should be stressed that unlike the methods applied in SHREC 2013 and the Panoramic method, which are mesh-based, Global Fisher and the proposed method require only raw point cloud information.

6. Conclusions

This work presents a P3DOR method, applicable on both point clouds and structured 3D models, which is based on a hybrid shape matching scheme, incorporating both local and global shape similarities for multiple scales. The main contributions of the proposed method address both local shape descriptor and partial retrieval aspects:

- The newly proposed dFPFH descriptor, which extends the well-known FPFH, in order to more accurately capture local geometric transitions,
- The use of an adaptive radius for dFPFH neighborhood, which depends on point cloud density,
- The use of a hybrid shape matching scheme, which incorporates local information directly derived from local shape descriptors, as well as global shape information derived from Fisher vectors,
- The definition of similarity that relies upon a weighted mean-minimum distance, as well as upon a weighted Fisher vector distance, both addressing the partiality of the 3D object query,
- The first application of Fisher encoding, instead of the popular *k*-means-based BoVW, in 3D object retrieval.

The experimental evaluation of the proposed method on the large-scale P3DOR benchmark dataset of SHREC 2013, as well as on the Hampson pottery dataset leads to the following conclusions:

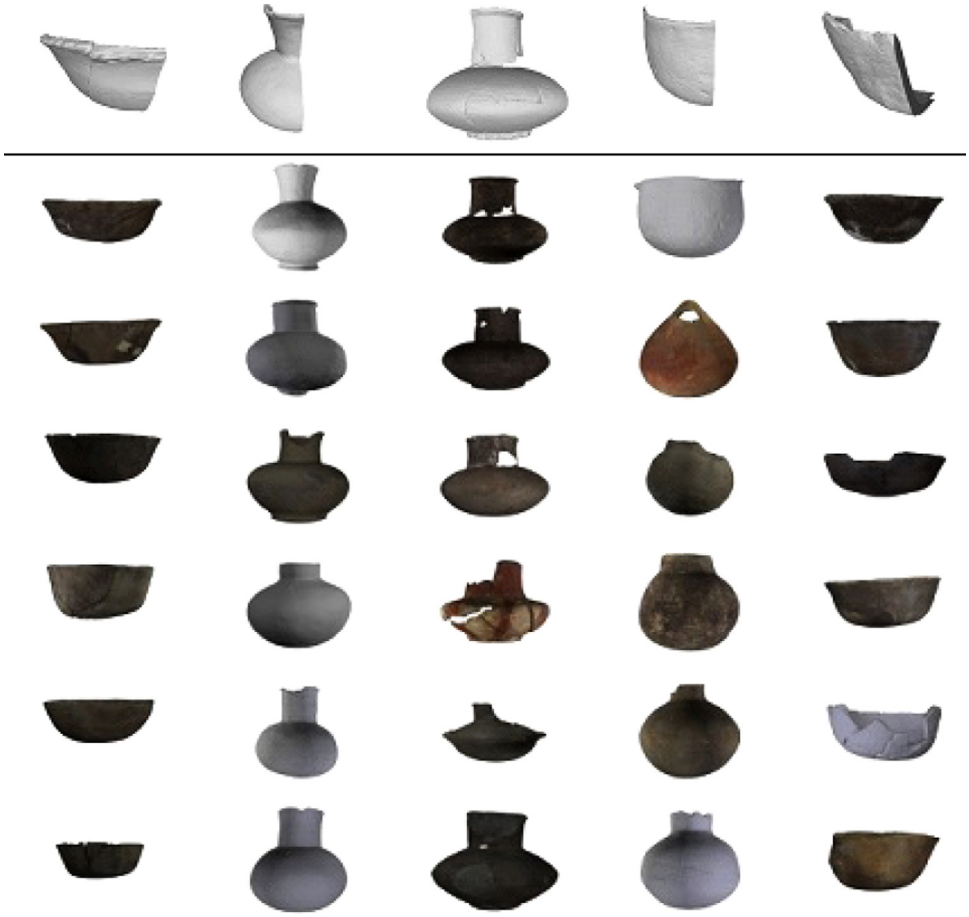

Fig. 9. Example ranked lists obtained by the proposed 3D object retrieval method in the case of the Hampson pottery dataset. Examples of partial queries are shown in the upper row, whereas the corresponding top-6 objects retrieved are shown below.

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

Table 3
Retrieval results of the proposed method on the Hampson pottery dataset with 25 real queries obtained by Breuckmann Optoscan scanner.

Method	NN	FT	ST	DCG
Proposed-6	0.641	0.351	0.554	0.712
Proposed-23	0.524	0.199	0.321	0.562

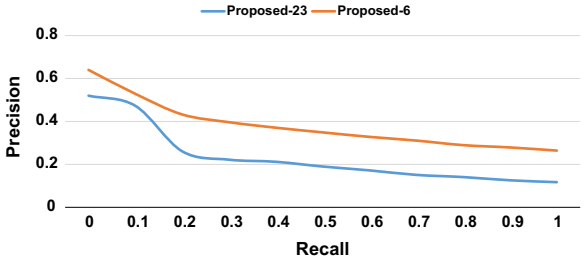

Fig. 10. Average P–R of the proposed method applied on the Hampson pottery dataset, for 25 queries obtained by Breuckmann Optoscan scanner. 'Proposed-23' and 'Proposed-6' correspond to 23-class and 6-class classification, respectively.

- Local and global shape similarity information, derived in multiple scales, act in a complementary fashion, maximizing the achieved retrieval performance when combined,
- The proposed P3DOR method outperforms state-of-the-art in terms of retrieval performance, when applied in SHREC 2013 dataset,
- The proposed method outperforms two recent CH applications of P3DOR, when applied in the Hampson pottery dataset.

Although the proposed method achieves state-of-the-art retrieval performance, the results remain far from perfect in absolute numbers, especially in the case of partial scans obtained with Breuckmann Optoscan scanner. In this respect, P3DOR remains a very challenging problem, which is open to future solutions. Hybrid retrieval methods combining multiple techniques for shape similarity assessment provide a promising direction for P3DOR [2].

Conflict of interest

None declared.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under Grant agreement no. 600533 PRESIOUS.

References

[1] M.A. Savelonas, I. Pratikakis, K. Sfikas, An overview of partial 3D object retrieval methodologies, Multimed. Tools Appl. (2014) 1–26.
[2] I. Sipiran, R. Meruane, B. Bustos, T. Schreck, B. Li, Y. Lu, H. Johan, A benchmark of simulated range images for partial shape retrieval, Vis. Comput. (2014).
[3] A. Bronstein, M. Bronstein, L. Guibas, M. Ovsjanikov, Shape Google: geometric words and expressions for invariant shape retrieval, ACM Trans. Graph. 30 (2011) 1–20.
[4] G. Lavoué, Combination of bag-of-words descriptors for robust partial shape retrieval, Vis. Comput. 28 (2012) 931–942.
[5] M.M. Bronstein, I. Kokkinos, Scale-invariant heat kernel signatures for non-rigid shape recognition, In: Proceedings of the CVPR, 2010, pp. 1704–1711.
[6] B. Li, A. Godil, H. Johan, Hybrid shape descriptor and meta similarity generation for non-rigid and partial 3D model retrieval, Multimed. Tools Appl. 72 (2014) 1531–1560.
[7] X. Hu, R.C. Eberhart, Human tremor analysis using particle swarm optimization, In: Proceedings of the Congress on Evolutionary Computation, 1999, pp. 1927–1930.
[8] M. Schmitt, R. Wanka, Particle swarm optimization almost surely finds local optima, Theoret. Comput. Sci. 561 (Part A) (2015) 57–72.
[9] F. Perronin, C. Dance, Fisher kernels on visual vocabularies for image categorization, In: Proceedings of the CVPR, 2007.
[10] J. Sánchez, F. Perronnin, T. Mensink, J. Verbeek, Image classification with the fisher vector: theory and practice, Int. J. Comput. Vis. 105 (2013) 222–245.
[11] H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, C. Schmid, Aggregating local image descriptors into compact codes, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1704–1716.
[12] R. Tao, E. Gavves, C.G.M. Snoek, A.W.M. Smeulders, Locality in generic instance search from one example, In: Proceedings of the CVPR, 2014, pp. 2099–2106.
[13] K. Chatfield, V. Lempitsky, A. Vedaldi, A. Zisserman, The devil is in the details: an evaluation of recent feature encoding methods, In: Proceedings of the BMVC, 2011, pp. 1–12.
[14] X. Zhou, K. Yu, T. Zhang, T. Huang, Image classification using super-vector coding of local image descriptors, In: Proceedings of the ECCV, 2010.
[15] R.B. Rusu, N. Blodow, M. Beetz, Fast point feature histograms (FPFH) for 3D registration, In: Proceedings of the ICRA, 2009, pp. 3212–3217.
[16] I. Sipiran, R. Meruane, B. Bustos, T. Schreck, H. Johan, B. Li, Y. Lu, SHREC'13 track: large-scale partial shape retrieval using simulated range images, In: Proceedings of the 3DOR, 2013.
[17] M. Savelonas, I. Pratikakis, K. Sfikas, Fisher encoding of adaptive fast persistent feature histograms for partial retrieval of 3D pottery objects, In: Proc. 3DOR, 2014, pp. 61–68.
[18] K. Sfikas, I. Pratikakis, A. Koutsoudis, M. Savelonas, T. Theoharis, Partial matching of 3D cultural heritage objects using panoramic views, Multimed. Tools Appl. (2014).
[19] M. Savelonas, I. Pratikakis, K. Sfikas, Partial 3d object retrieval combining local shape descriptors with global Fisher vectors, In: Proceedings of the 3DOR, 2015.
[20] A. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3D scenes, IEEE Trans. Pattern Anal. Mach. Intell. 21 (1999) 433–449.
[21] B. Steder, G. Grisetti, M. Van, L.W. Burgard, Robust online model-based object detection from range images, In: Proceedings of the IROS, 2009, pp. 4739–4744.
[22] R. Ohbuchi, M. Nakazawa, T. Takei, Retrieving 3d shapes based on their appearance, In: Proceedings of the ACM SIGMM MIR, 2003, pp. 39–45.
[23] B. Steder, R.B. Rusu, K. Konolige, W. Burgard, Point feature extraction on 3D range scans taking into account object boundaries, In: Proceedings of the IEEE ICRA, 2011.
[24] L. Bo, X. Ren, D.Fox, Depth kernel descriptors for object recognition, In: Proceedings of the IROS, 2011, pp. 821–826.
[25] R.B. Rusu, N. Blodow, Z. Marton, M. Beetz, Aligning point cloud views using persistent feature histograms, In: Proceedings of the IROS, 2008, pp. 3384–3391.
[26] G. Stavropoulos, P. Moschonas, K. Moustakas, D. Tzovaras, M. Strintzis, 3D model search and retrieval from range images using salient features, IEEE Trans. Multimed. 12 (2010) 692–704.
[27] P. Daras, A. Axenopoulos, A 3D shape retrieval framework supporting multimodal queries, Int. J. Comput. Vis. 89 (2009) 229–247.
[28] P. Li, H. Ma, A. Ming, Combining topological and view-based features for 3D model retrieval, Multimed. Tools Appl. 65 (2013) 335–361.
[29] J. Tierny, J. Vandeborre, M. Daoudi, Partial 3D shape retrieval by reeb pattern unfolding, Comput. Graph. Forum 28 (2009) 41–55.
[30] A. Agathos, I. Pratikakis, P. Papadakis, S. Perantonis, P. Azariadis, S. Sapidis, 3D articulated object retrieval using a graph-based representation, Vis. Comput. 26 (2010) 1301–1319.
[31] T. Furuya, R. Ohbuchi, Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features, In: Proceedings of the ACM ICIVR, 2009.
[32] B. Li, H. Johan, Sketch-based 3D model retrieval by incorporating 2D–3D alignment, Multimed. Tools Appl. 65 (2013) 363–385.
[33] B. Li, Y. Lu, H. Johan, Sketch-based 3D model retrieval by viewpoint entropy-based adaptive view clustering, In: Proceedings of the 3DOR, 2013, pp. 49–56.
[34] I. Sipiran, B. Bustos, T. Schreck, Data-aware 3D partitioning for generic shape retrieval, Comput. Graph. 37 (2013) 460–472.
[35] I. Sipiran, B. Bustos, Harris 3D: a robust extension of the Harris operator for interest point detection on 3d meshes, Vis. Comput. 27 (2011) 963–976.
[36] D.V. Vranic, DESIRE: a composite 3d-shape descriptor, In: Proceedings of the IEEE, 2005, pp. 962–965.
[37] A. Johnson, Spin-Images: a representation for 3-D surface matching (Ph.D. thesis), Robotics Institute, Carnegie Mellon University, 1997.
[38] A. Frome, D. Huber, R. Kolluri, T. Blow, J. Malik, Recognizing objects in range data using regional point descriptors, In: Proceedings of the ECCV, Lecture Notes in Computer Science, vol. 3023, 2004, pp. 224–237.
[39] W.K. Leow, R. Li, The analysis and applications of adaptive-binning color histograms, Comput. Vis. Image Underst. 94 (2004) 67–91.
[40] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2004) 91–110.
[41] R.B. Rusu, Semantic 3D object maps for everyday manipulation in human living environments (Ph.D. thesis), Computer Science Department, Technische Universitaet Muenchen, Germany, 2009.
[42] R.B. Rusu, Z.C. Marton, N. Blodow, M. Beetz, I.A. Systems, T.U. Mnchen, Persistent point feature histograms for 3D point clouds, In: Proceedings of the ICIAS, 2008.
[43] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (1982) 129–136.

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003

Michalis Savelonas is a research fellow in Athena Research and Innovation Center, Greece, as well as in the Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece.

Ioannis Pratikakis is currently an Assistant Professor in the Department of Electrical and Computer Engineering, Democritus University of Thrace, Greece, and an adjunct researcher in Athena Research and Innovation Center, Greece.

Konstantinos Sfikas is a research fellow in Athena Research and Innovation Center, Greece, and Department of Computer and Information Science, NTNU, Trondheim, Norway.

Please cite this article as: M.A. Savelonas, et al., Fisher encoding of differential fast point feature histograms for partial 3D object retrieval, Pattern Recognition (2016), http://dx.doi.org/10.1016/j.patcog.2016.02.003