DOI: 10.1111/cgf.13493
Eurographics Symposium on Geometry Processing 2018
T. Ju and A. Vaxman
(Guest Editors)

Volume 37 (2018), Number 5

Effective Characterization of Relief Patterns

A. Giachetti

Department of Computer Science, University of Verona, Italy

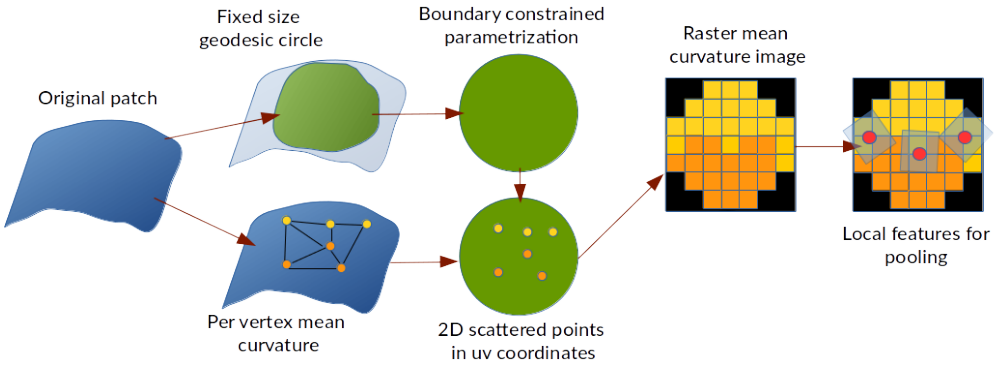

Figure 1: *The proposed mapping of surface patches onto raster images. Original meshes are automatically cropped to geodesic circles of fixed size. Geodesic circles are mapped on the plane with boundary constrained embedding. Vertex attributes (in our tests mean curvature) are interpolated in a fixed grid in the u-v space. State of the art texture analysis methods can then be applied to characterize the relief pattern.*

Abstract

In this paper, we address the problem of characterizing relief patterns over surface meshes independently on the underlying shape. We propose to tackle the problem by estimating local invariant features and encoding them using the Improved Fisher Vector technique, testing both features estimated on 3D meshes and local descriptors estimated on raster images created by encoding local surface properties (e.g. mean curvature) over a surface parametrization. We compare the robustness of the obtained descriptors against noise and surface bending and evaluate retrieval performances on a specific benchmark proposed in a track of the Eurographics Shape REtrieval Contest 2017. Results show that, with the proposed framework, it is possible to obtain retrieval results largely improving the state of the art and that the image-based approach is still effective when the underlying surface is heavily deformed.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computational Geometry and Object Modeling—Curve, surface, solid, and object representations

1. Introduction

Recognition and retrieval of relief patterns are very important problems in shape analysis, still not well addressed in the literature. The necessity of characterizing with a descriptor a high-frequency motif relieved on a low-frequency background shape is now arising in several practical applications, due to the availability of high-resolution 3D models acquired by different kinds of scanners. Examples of such applications are the analysis of artworks' styles [PWM*16], the recognition of natural structures like trees [OVSP13], the retrieval of tissue motifs [BTA*17], but also material roughness analysis [GKM*02]. Figure 2 shows examples of artworks (left) and natural structures (tree barks, right) where it may be useful to localize and recognize different kinds of relief patterns.

This problem can be considered as a particular kind of texture characterization, and can indeed be addressed with classical 2D texture analysis algorithm but, in this case, it first requires to decouple the 2D relief texture from the underlying base shape in a sort of reversed bump mapping procedure.

The subtraction of the background shape and the subsequent analysis of the relief pattern as a 2D texture is a procedure commonly used in material science in order to characterize the so-called

Figure 2: *Example objects with relief patterns superimposed to non planar background surfaces. Left: artworks. Right: natural structures (tree barks).*

surface roughness. In this field, specific procedures, defined in ISO standards (e.g. ISO 25178), are applied to depth maps/point clouds (typical output of microprofilometric acquisitions), in order to subtract the background shape. Areal descriptors (that are actually sets of 2D texture features) are then extracted to characterize the roughness pattern.

In this case, the use of a 2D technique seems clearly the best solution as the background shape is typically approximated at the patch scale with a plane or a low-order polynomial. This is due to the fact that the material roughness is estimated on very high-resolution scans and is characterized by spatial frequencies much higher than those characterizing the underlying shapes.

But when we need to retrieve or recognize surface regions based on a relieved decorative pattern, and the scale of the characterizing elements of the patterns is not negligible with respect to the spatial frequencies of the underlying surface, the problem may be challenging and different strategies may be employed to solve it.

In this paper we propose and compare different approaches to address this problem, trying to leverage on recent advances in image texture analysis methods, adapted to the specific domain.

After a clear statement of the research problem addressed, we propose different solutions based on local feature pooling with Improved Fisher Vectors, compare their effectiveness on a recent shape retrieval benchmark characterizing object classes with the overlaid relief patterns, and analyze the main properties of the relief pattern descriptors obtained.

2. Problem statement and related work

Let us define precisely the problem we want to address with our work. We have surface patches of an approximately constant size, encoded as polygonal meshes, characterized by different relief patterns overlaid on different base shapes. Our goal is to classify with the same label all the instances of patches with the same pattern or to retrieve all the shapes with a similar pattern, independently on the base shape, given an example.

The base shapes are not approximately flat at the scale of the relieved patterns so that it is not obvious that the problem can be solved by fitting a plane and using 2D texture analysis on the depth map measuring the distance from the plane.

In this case, we may consider two different strategies to estimate a descriptor of the texture of the patch. The first one consists in globally separating the high-frequency decoration pattern from the underlying shape.

An example of this approach can be found in [OVSP13]. Here the authors tried to characterize trees species from 3D bark texture by estimating images representing the 3D deviation over a cylinder and using then a wavelet-based descriptor to represent the relief texture. If the base geometry is more complex, it is still possible to try to encode the high-frequency components in 2D images by means of local parametrization, as proposed, for example in [PMCS11] to have a better visual analysis of artworks.

The second one consists of estimating 3D texture descriptors directly estimated on meshes or point clouds. Examples of this kind are meshLBP, extending the Local Binary Pattern texture description for images to triangle meshes [WBDB15] or meshHOG [ZBVH09], adapting the Histogram of Gradients image descriptors to meshes. In this case, the advantage is to directly work on the surface data, but the risk is to have methods strongly depending on the mesh quality and not well suited to describe relief patterns on generic triangulations.

Many existing techniques may be adapted for the task following these approaches, but, in order to understand and compare them, it is necessary to evaluate their performances on specific testbeds. A relief pattern dataset suitable for this task has been proposed in the Eurographics SHREC 2017 track on retrieval of surfaces with similar relief patterns [BTA*17]. In this contest, a dataset with 720 mesh patches belonging to 15 different decoration classes has been proposed 3. Patches were obtained by scanning relief decorations on real textiles differently bent and proposing different remeshings of the original scans.

In this contest, participants tested both 2D and 3D descriptors to characterize the relieved decorations, but, despite the relatively easy visual classification of the patches, the proposed methods were not particularly successful.

The best 3D approach consisted in the computation of spectral components of 3D meshes using a curvature-based kinetic term, which should remove the influence of shapeâĂŹs articulations. The best 2D approach was based on mapping principal curvatures onto a mesh parametrization and extracting a simple texture descriptor on the images obtained.

Our idea is to tackle the problem by leveraging on recent results in the field of image texture analysis [LCF*18]. Recent works

demonstrated that a reasonable approach to recognizing well characterized statistical patterns is to estimate local invariant features, e.g. SIFT or pre-trained CNN based features, and then use bag-of-features or more advanced pooling approaches (e.g. Improved Fisher Vector or Bilinear feature pooling) to derive texture descriptors in the region of interest [CMK*14, CMKV16, LM16, SLF*16]. These methods outperform texture analysis methods based on statistics of local filters or LBP features in several benchmarks due to their ability to characterize pattern components and their co-occurrence. This scheme can be applied to our case as well.

What it is interesting to investigate is if, for this kind of patterns, it may be more effective to pool features estimated directly on meshes or point clouds, like depth/mesh SIFTs [DK12, SKVS13], or it may be better to rely on 2D image features (e.g. the original SIFT transform [Low99]). These features can be computed on images obtained mapping distance to reference shape or local descriptors (e.g. mean curvature) estimated on the mesh on locally approximating surface or a 2D mesh parametrization. In our work, we analyze the behavior of these descriptors on surface patches and evaluate the retrieval performances on the SHREC 2017 benchmark.

The contribution of our work is therefore twofold: we clarify many aspects hidden in the relief pattern characterization problem and propose practical retrieval methods largely outperforming the state-of-the-art.

3. The SHREC'17 dataset

A public dataset of low quality scans of different base surfaces with classified relief patterns has been recently proposed in the Eurographics Shape REtrieval Contest track on Retrieval of surfaces with similar relief patterns [BTA*17]. This dataset has been realized by acquiring low-resolution meshes from samples of textiles differently bent and stretched using an Intel RealSense depth sensor. Classes correspond to patches coming from the same kind of textile.

Participants were asked to provide dissimilarity matrices, used to evaluate classical retrieval scores, namely Nearest Neighbor (NN, the percentage of most similar retrieval of the same class of the query), First Tier (FT, the average percentage of the N-1 most similar patches of the same class of the query, where N is the number of elements per class), Second Tier (ST), e-measure (e), Discounted Cumulative Gain (DCG) and Mean Average Precision (mAP). The complete description of these classic retrieval scores can be found in [GLD*10].

Examples meshes of the dataset are shown in Figure 3. It is possible to see that background surfaces can present relevant curvature and complex bending. Different classes of decorative patterns are clearly distinguished by humans, but the methods proposed by contest participants obtained poor retrieval results. On the dataset without same scans with different meshing, the best method obtained only a 63% of Nearest Neighbor correct retrieval and a quite poor Mean Average precision of 0.39.

This dataset is quite interesting for evaluating the performances of relief patterns characterization method, as the underlying shape deformations are not negligible. For this reason, we exploited it as the main testbed for validating the proposed algorithms.

A couple of notes on this dataset are, however, mandatory. The full dataset is composed of 720 patches, but for it includes four different remeshings for the same acquired shape. This means that, on this dataset, retrieval scores, especially the Nearest-Neighbor retrieval may be biased by the ability of a method to capture similarity of global shapes instead of pattern similarity. For the same reason, supervised classification tests on the 720 patches dataset are also not significant if training and tests sets include remeshings of the same original patch. and should not be performed on this full set. A reduced dataset without remeshing has been, however, used by the authors to present an unbiased evaluation of shape retrieval based on surface pattern only. We will exploit the single meshing dataset as well.

The second note is related to the fact that, as the different patches of a class are obtained from the same textile, the results will not evaluate the ability of relief encoding to capture style similarity or perceptual texture similarity, so that specific benchmarks should be realized to capture and model different semantic grouping of patterns. We plan to realize specific benchmarks for practical purposes (e.g. artworks' style characterization) in the near future.

4. The proposed approach

Our idea is to solve the relief pattern similarity estimation problem by leveraging on recent results of texture analysis research, trying also to answer a basic question: do we need to rely on 3D features, or it is better to rely on 2D feature estimated on a "background corrected" planar surface?

We propose to exploit for relief texture characterization to the pooling of local invariant features based on the Improved Fisher Vectors, as the method has been successfully applied to several image texture contexts.

Using the same feature encoding scheme, we pool different features derived from a smart mapping of surface features onto 2D images or directly estimated on 3D meshes.

4.1. Pooling local invariant features

In order to derive a texture descriptor from a sparse collection of local features we employed the Improved Fisher Vector [PSM10] encoding. Fisher Vectors (FV) store the mean and the covariance deviation vectors per component of a Gaussian-Mixture-Model (GMM) fitted on the unlabeled data, together with the elements of the local feature descriptors. The improved Fisher Vector (IFV) improves the performance of the representation by applying a nonlinear kernel to the vector components and applying $l-2$ norm normalization. The IFVs are therefore vectors including the deviations of the distribution of features of each instance from the model trained on the unlabeled data. The dimensionality of the encoding is $2kd$, where k is the number of Gaussians in the GMM, and d is the dimensionality of the patch feature vector (e.g. 128 in the case of a SIFT).

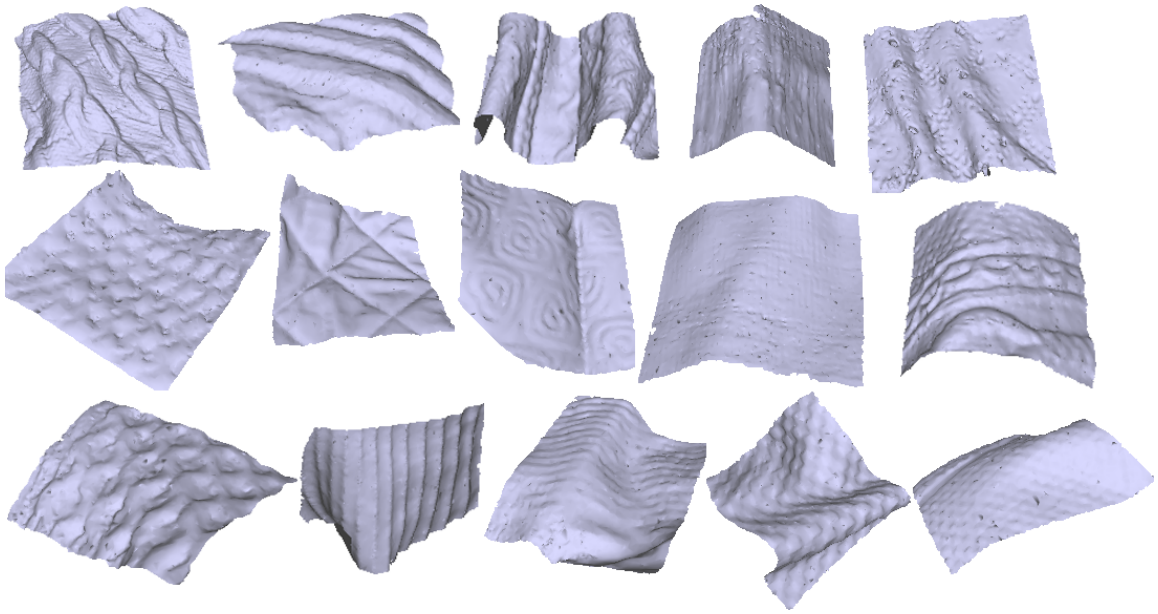

Figure 3: *Example patches from the SHREC 2017 dataset. It is possible to see the different bendings and stretchings applied to the tissues' surfaces, that are often quite relevant.*

4.2. Image based feature encoding

Using the IFV encoding we first created patch descriptors using 2D image features estimated on 2D raster images derived from the 3D meshes. Following the ideas used by a few participants of the SHREC'17 track [BTA*17], we mapped curvature values evaluated on the mesh vertexes (we decided to use only the mean curvature) onto a planar surface and interpolated the values on a regular grid, thus being able to estimate on the resulting texture descriptors on the resulting images.

In our implementation, the surface patches have been first uniformly remeshed to simplify the estimates and regularizing the meshes. This step has been done using the volumetric approach derived from Curless method and implemented in Meshlab [CCC*08], obtaining models with approximately 20K vertices.

We tested two different ways to map the relief pattern onto the plane. The first is a simple projection on the best plane fitting the patch area, used as baseline. This method is straightforward but introduces large distortions in non-planar patches. However, we will see that it can provide reasonable results on the challenging SHREC dataset.

The second is based on Tutte embedding [FH05], mapping the patch boundary onto a circle on the plane. In order to reduce distortions, we cropped the original patches onto a geodesic circle of fixed size, centered in the vertex with maximum distance from the borders before the mapping, as shown in Figure 1, representing the basic steps of the procedure. For the SHREC patches, we used a geodesic radius of 5cm.

The use of circular geodesic regions has been already exploited in shape analysis, for example in [MBBV15]. In that case, however, it was used to characterize vertex neighborhood at a small scale, while we crop large geodesic areas that should include all the pattern elements characterizing the texture class.

With both the mappings, projected points with associated mean curvature are then resampled on a square grid of fixed size (267x267), so that the pixel size correspond to a geodesic length of about 3.75*mm*. Texture features have been estimated on the resulting raster image. Figure 4 shows example images obtained from SHREC'17 meshes with the two methods.

We have chosen the mean curvature as the mesh feature representing the local relief. We also tested the use of principal curvatures, yielding quite similar results. The range of the curvature values on the full dataset was mapped onto an 8-bit depth image range.

Following the approach proposed in [CMK*14], we estimated local features on the resulting raster images using the SIFT transform [Low99]. Our rationale is that, being the textures possibly distorted by the parametrization, this effect could be effectively compensated by analyzing them with invariant features. SIFT are sufficiently robust against scale and orientation changes and even if the parameterizations can create relevant distortions in some part of the domain, we expect that most of the local features that are not strongly affected.

The SIFT descriptors are estimated in all the peaks detected by a multiscale Difference of Gaussian detector. Each keypoint comes with a reference scale and orientation, and these are used to define a local frame where the descriptor, an histogram of gradients in the neighborhood, is estimated.

In our tests, we used the SIFT implementation provided in the VLFeat library [VF08]. SIFT descriptors are vectors with 128 components.

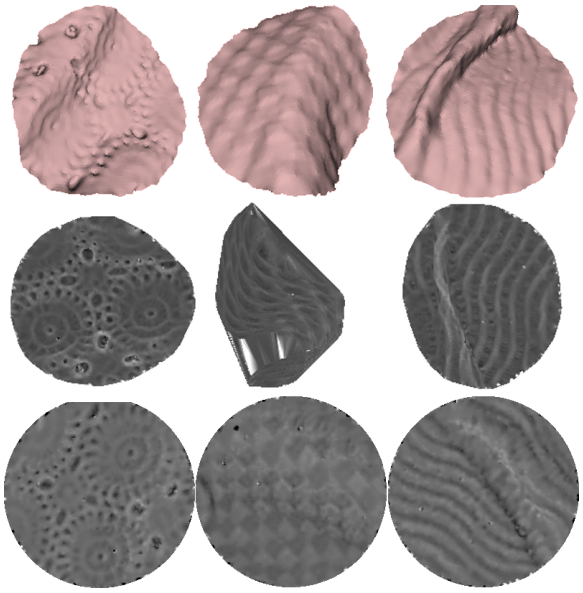

Figure 4: *Images obtained mapping the mean curvature evaluated on patch vertexes onto the parametrization coordinates. Top row: mesh renderings. Middle row: images obtained with best fitting plane. Bottom row: images obtained on boundary constrained parametrization (after geodesic circle cropping).*

On the whole set of features estimated on the dataset we fitted a Gaussian Mixture Model and then encoded the features estimated on each patch using the Improved Fisher Vector encoding given the Gaussians parameters, still exploiting the VLFeat implementation. To obtain a dissimilarity matrix for the patch set we estimated the Eucliedean distance of the IFV (we also tested different metrics with similar results).

The number of Gaussian components used in the Fisher kernel can be a critical choice. Research works have been dedicated to the estimation of an ideal number of components [FJ02]. In our application we started by using 50 components, a number comparable to those employed in other texture analysis works [CMK*14], but we performed experimental tests to assess the effect of varying the number of Gaussians (see Section 5).

4.3. Encoding CNN-based feature

Most state-of-the-art texture descriptors for 2D images are based on transfer learning exploiting pre-trained deep networks [CMKV16]. The idea is to take the output of the last convolutional layers of Convolutional Neural Networks trained on large image datasets as local feature. In our implementation, we used the VGG-16 network trained on more than a million images from the ImageNet data (`http://www.image-net.org`). The pre-trained network is available in the Matlab Neural Network toolbox. We extracted as local features the outputs of the last convolutional layer that means 196 descriptors with 512 elements. The network inputs are the curvature images replicated in the three channels and resized to match the dimensionality of the pre-trained VGG-16 input.

The activation vectors are pooled similarly to the SIFT vectors, using the same number of Gaussian components for the IFV encoding.

For further comparison, we also considered the possible use of the last or penultimate Fully Connected layers outputs of the pre-trained network. This choice has been proposed as well in recent papers [CMK*14] as an effective 2D texture patch feature. In this case, the unique layer output is directly used as the final global descriptor, so the pooling step is avoided.

4.4. Encode 3D local features

To compare the SIFT/IFV method with a "similar" 3D approach, we estimated on the mesh patches the local depth SIFT (LD-SIFT) features proposed in [DK12]. The method is based on detecting points of interest by applying a multiscale difference of Gaussians (DoG) function and then evaluating scale invariant features on locally oriented and scaled depth images around the keypoints.

In this method, similarly to the original image-based technique, a Gaussian filter is applied on the mesh geometries a set of filtered meshes, the mesh octaves are computed. On these meshes, a density-invariant DoG detector, demonstrating high repeatability on specific benchmarks is applied to extract salient points. Similarly to the image case, the detector gives also scale and orientation estimations, that are used to define a local plane and a frame to derive a localized depth map, used for the SIFT-like histogram evaluation.

We used the authors' implementation to evaluate the features on the remeshed patches, and used the same IFV encoding of the feature vectors with a Gaussian mixture.

5. Experimental results

5.1. Tests on SHREC'17 data

With all the proposed methods we estimated the dissimilarity matrices for both the 720 and the 180 patches datasets of the SHREC'17 contest. Table 1 and 2 show the retrieval scores obtained with the proposed methods, compared with the ones collected in the contest.

The retrieval scores achieved on the full and non-remeshed datasets with the proposed technique are rather good and much higher than those obtained in the contest.

These large improvements are evident also looking at the Precision-Recall plots represented in Figure 5 and 6. Precision is the fraction of retrieved instances that are relevant (e.g. of the same class of the query), while recall is the fraction of relevant instances that are retrieved. An optimal retrieval technique would yield curves close to the constant value of 1.

The plots in Figures 5 and 6 demonstrate the effectiveness of the IFV pooling method, that, despite the lack of encoding of spatial relationships between the feature components, is known to outperform other pooling methods like bag-of-words schemes.

However, unlike what happens in other 2D texture classification benchmarks, the pooling method seems not effective when combined with pre-trained CNN based features. Retrieval scores and

	NN	1-Tier	2-Tier	e	DCG	mAP
KLBO-FV [BTA*17]	0.986	0.333	0.449	0.332	0.759	0.339
CMC-2 [BTA*17]	0.763	0.272	0.389	0.261	0.686	0.271
Plane/meanC/SIFT/FV	0.950	0.636	0.765	0.593	0.892	0.650
Tutte/meanC/SIFT/FV	0.993	**0.712**	**0.850**	**0.647**	**0.929**	**0.739**
Tutte/VGGConvFV	0.957	0.532	0.669	0.490	0.843	0.549
Tutte/VGGFC7	0.982	0.548	0.701	0.511	0.860	0.573
MeshLDSift+FV	**0.997**	0.459	0.559	0.464	0.833	0.482

Table 1: *Retrieval scores of the proposed methods compared with the best results of the contest on the full set of patches including multiple remeshings. All the proposed methods largely outperform the methods presented in the contest, with 2D methods performing best.*

	NN	1-Tier	2-Tier	e	DCG	mAP
KLBO-FV [BTA*17]	0.522	0.295	0.412	0.247	0.603	0.307
CMC-2 [BTA*17]	0.633	0.363	0.494	0.293	0.662	0.390
Plane/meanC/SIFT/FV	0.828	0.616	0.761	0.420	0.837	0.645
Tutte/meanC/SIFT/FV	**0.872**	**0.710**	**0.849**	**0.457**	**0.883**	**0.741**
Tutte/VGGConvFV	0.717	0.529	0.665	0.370	0.765	0.554
Tutte/VGGFC7	0.728	0.549	0.688	0.392	0.782	0.577
MeshLDSift+FV	0.778	0.433	0.526	0.305	0.730	0.460

Table 2: *Retrieval scores of the proposed methods compared with the best results of the contest on the patch set without remeshings. All the proposed methods largely outperform the methods presented in the contest, with 2D methods performing best.*

precision vs. recall curves obtained with the pooled convolutional activations of the VGG16 net are not only relevantly worse than those obtained with SIFTs but also slightly worse than those obtained with the direct comparison of the vector extracted from the penultimate Fully Connected layers.

Note that we also tried different choices for layer selection, the combination of features coming from different layers and PCA-based dimensionality reduction, but these solutions did not improve the overall retrieval scores.

This may be due to different reasons. First, convolutional features of this kind are not invariant against distortions. Furthermore, relief patterns (and the differences of decoration classes) are not necessarily well represented by the features of the images used for training (and on their inter-class variations). By training a specific network on differently distorted relief patterns, it could be possible to improve the performances.

SIFT features are then estimated on salient points, while convolutional features are sampled in a fixed grid and this may be a key factor in capturing relevant relief pattern features. To confirm this hypothesis we also tested the encoding of uniformly gridded Dense Sift features [FFP05] instead of the classic ones, getting scores relevantly lower than those obtained with SIFTs.

The results obtained with the SIFT features are remarkably good. Looking at the tables and plots, it is possible to see that the invariance properties of the SIFT descriptors make them really good even when computed on the images obtained with plane projection (scores are higher than those obtained with VGG16 based features and mesh based depth SIFT).

Using mesh-based features, the retrieval performances were still relevantly higher than those obtained by the SHREC contest participants, but poor with respect to the image based ones.

5.2. Effect of Gaussian components' number

In the SHREC retrieval tests, we analyzed the effects of changing the dimensionality of the Gaussian Mixture in the IFV encoding. For the image Sift and LDSift pooling, we varied the number of Gaussian components in the range 10-100. evaluating retrieval scores. Looking at Figure 7, it is possible to see that the mean average precision is already good with 10 components (e.g. relatively low dimensional IFV), and 50 components seem a good choice.

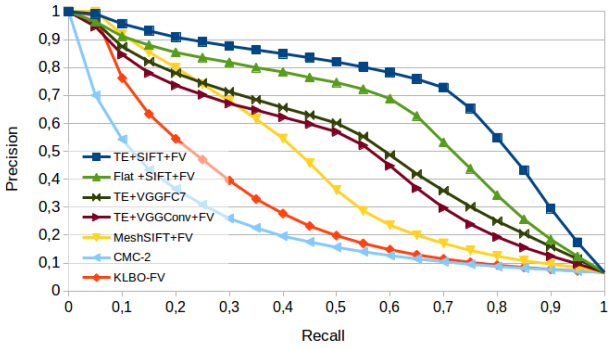

Figure 5: *Precision vs Recall plot showing the retrieval performances of the proposed methods on the full dataset, compared with the techniques performing best in the SHREC'17 contest.*

Clearly we expect that, increasing the number of patterns to be discriminated and their complexity, the ideal number of components could be different.

#PCA comp.	NN	1-Tier	2-Tier	e	DCG	mAP
-	0.993	0.712	0.850	0.647	0.929	0.739
64	**0.997**	0.726	0.857	0.655	0.933	0.750
32	0.996	**0.743**	**0.869**	**0.670**	**0.939**	**0.768**
16	**0.997**	0.716	0.859	0.651	0.933	0.748
#PCA comp.	NN	1-Tier	2-Tier	e	DCG	mAP
-	0.872	0.710	0.849	0.457	0.883	0.741
64	0.883	0.714	0.851	0.457	0.881	0.741
32	0.883	**0.739**	**0.858**	**0.468**	**0.899**	**0.768**
16	**0.889**	0.709	0.849	0.463	0.885	0.747

Table 3: *Retrieval scores obtained on the SHREC benchmarks applying PCA dimensionality reduction before Fisher encoding. Top: results on benchmark including remeshings. Bottom: results on 180 original meshes only.*

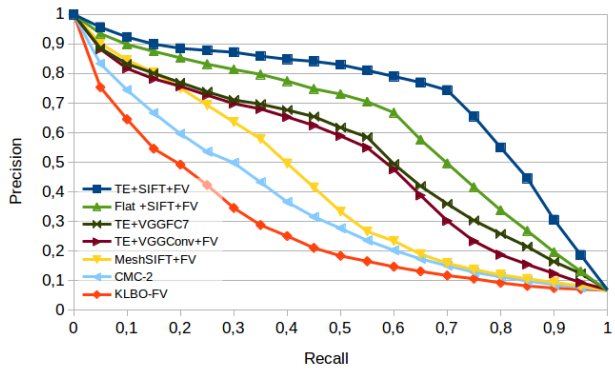

Figure 6: *Precision vs Recall plot showing the retrieval performances of the proposed methods on the dataset without remeshed patches, compared with the techniques performing best in the SHREC'17 contest.*

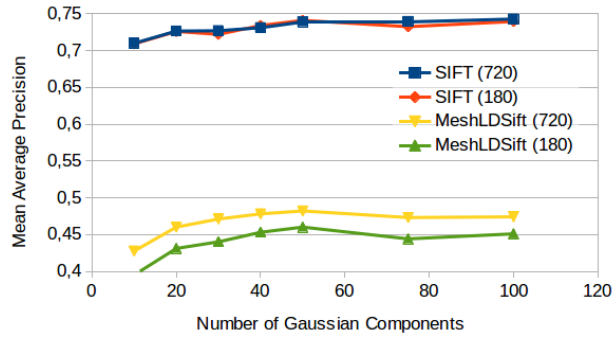

Figure 7: *Mean average precision obtained on the two dataset pooling Image SIFT and LD-SIFT descriptors with Improved Fisher Vectors varying the number of components of the Gaussian mixture model.*

5.3. Dimensionality reduction

The IFV encoding of SIFTs is a high dimensional vector, in our case of size 12800 elements. In some applications, it is proposed make the representation more compact by using dimensionality reduction methods, e.g. PCA mapping of the original feature vectors (SIFT). We checked also the effect of decorrelating with PCA our SIFT vectors before to apply the IFV. It turns out that reducing the dimensionality of the SIFTs is in our case a good choice, slightly improving the retrieval scores, and that 32 components seem optimal, as shown in Table 3.

5.4. Patch size and discretization

A reasonable characterization of a relief patterns over an underlying geometry depends critically on two different scale parameters: the area of a sufficiently large surface patch such that the statistical distribution of the geometrical motifs defining the "texture" classes of interest and the scale of the single elements characterizing the pattern, that should be captured by the local features extraction. These parameters are clearly task-specific and should be carefully tuned or automatically estimated for each potential application. For the SHREC patches, we have patches of similar with coherent patterns, so it is expected we can use the full patch or a large subset for the characterization. When we create the images of the parametrization we rasterize with a pixel size small with respect to the basic elements of the patterns reported in Figure 3. We can actually test the effect of taking smaller geodesic regions or coarser image discretization on the retrieval scores. Table 4 shows that the scores are not relevantly decreased reducing the resolution to 3/4 of the original one, showing that both the parameters are well adapted to the task.

Size	NN	1-Tier	2-Tier	e	DCG	mAP
Full	0.883	0.714	0.851	0.457	0.881	0.741
75%	0.872	0.686	0.820	0.449	0.876	0.724
50%	0.694	0.492	0.633	0.361	0.743	0.507

Res.	NN	1-Tier	2-Tier	e	DCG	mAP
Full	0.883	0.714	0.851	0.457	0.881	0.741
75%	0.939	0.691	0.838	0.453	0.890	0.735
50%	0.783	0.497	0.640	0.366	0.764	0.522

Table 4: *Retrieval scores changes in the 180 original meshes test obtained by reducing geodesic patch size (top) or mean curvature image resolution (bottom)*

5.5. Analysis of features' behavior

To better understand the behavior of the proposed descriptors of relief patterns on bended surfaces, we created a few synthetic patches modifying mesh patches extracted from cylindrical surfaces of different radii by mapping vertex displacements along the normal direction onto the cylindrical parametrization.

Using five decorative patterns encoded as raster images (see Figure 8), we generated 125 patches using different cylinder radii and different levels of added noise (see Figure 9). Bending ranges from a flat patch (level 0) to a patch where the geodesic patch radius is the same of the cylinder's one (level 4), while the noise consists of normally distributed displacements along the normals of the original surface before applying the relief pattern. Level 0 corresponds to zero noise. At the maximum noise level (4) the standard deviation of the displacements is in the same range of the patterns' relief so that patterns are almost obscured at the human eye (Figure 10).

On this synthetic dataset, a retrieval test based on the same scores used before still gives good results for our technique based on IFV pooling of SIFT features estimated on mean curvature images (Table 5) The mesh-based method obtains instead poor scores, probably due to the lack of robustness of the method against vertex noise.

What is more interesting, here, is to see the robustness of the features against bending and noise. Figure 11 shows the average difference of IFV encoded SIFT descriptors estimated on bent patches and the same descriptors evaluated on the planar patch (level 0), with different noise levels. It is possible to see that, when the "acquisition" is not noisy, the difference of the descriptors actually de-

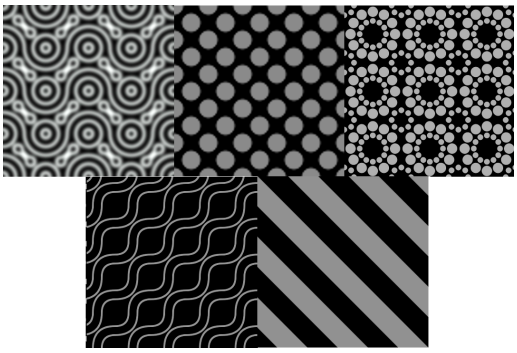

Figure 8: *Image patterns used to generate the bended mesh with relief decoration with displacement mapping.*

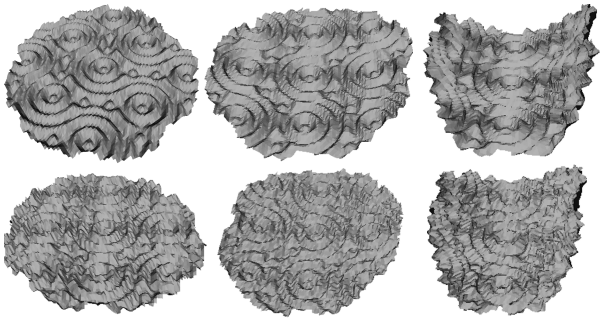

Figure 9: *Examples of synthetic patches generated from the first decorative pattern. Top: Noise level 1 (scale 0-4) and bending levels 0. 2, 4. Bottom: Noise level 3 and bending levels 0.2,4.*

pends on the bending, even if is quite lower than the average inter-class distance.

When the noise level is higher, however, the difference no longer depends on the amount of bending and clearly grows, but it is still sufficiently low to ensure the correct retrieval.

Note that the inter-class distance standard deviation is quite low (3.5% of the average) and this means that, even at the higher level of noise, the intra-class and inter-class mean differences are statistically significant. This fact is confirmed by the near-optimal retrieval scores obtained.

5.6. Estimating Gaussians on different patterns

The small synthetic dataset of decorated patches has been used for another simple experiment. We just wanted to show that it is not mandatory to estimate the Gaussian Mixture parameters on the same dataset of the test to have good results, as it is just the type of patterns involved that determines the basic elements of the description. If we use the Gaussian Mixture parameters estimated on our small and noisy synthetic dataset to encode IFV for the SHREC dataset, the resulting retrieval scores are still very good (Table 6).

	NN	1-Tier	2-Tier	e	DCG	mAP
Tutte/MeanC/SIFT/FV	**1.00**	**0.79**	**0.89**	**0.71**	**0.96**	**0.84**
MeshLDSift+FV	0.05	0.17	0.36	0.20	0.53	0.23

Table 5: *Retrieval scores obtained on the synthetic patches with bending and noise. The image based method provides very good scores.*

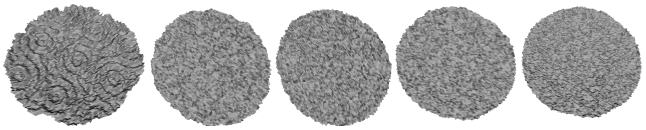

Figure 10: *At the maximum level of noise patterns are hardly visible on the patches' surfaces.*

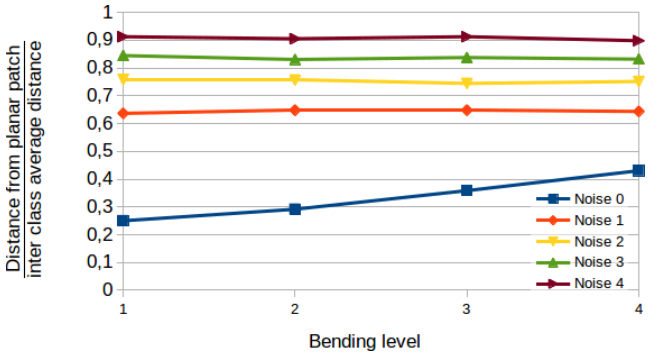

Figure 11: *Average distance between the 2D SIFT/IFV descriptors computed on bended meshes (with different noise added) and on the corresponding flat mesh divided by the average inter-class descriptors distance. The distance depends on bending only for meshes without added vertex noise.*

	NN	1-Tier	2-Tier	e	DCG	mAP
720 patches GM SHREC	0.993	0.712	0.850	0.647	0.929	0.739
720 patches GM Synth	0.985	0.636	0.787	0.580	0.895	0.657
180 patches GM SHREC	0.872	0.710	0.849	0.457	0.883	0.741
180 patches GM Synt	0.844	0.621	0.780	0.427	0.835	0.658

Table 6: *Retrieval scores are not much decreased if we encode IFV with the Gaussian Mixture parameters estimated on the small synthetic relief pattern dataset.*

6. Discussion

The results obtained in our experiments reveal some interesting facts:

- The Improved Fisher Vector encoding of local features is particularly effective in this specific pattern recognition task. This is due to smart use of the generative model that is effective in capturing the relevant texture properties with respect to the simple Bag of Features approach.
- Image based features obtained mapping discrete curvatures onto a 2D parametrization of the whole patch seem more effective than mesh based descriptors. The deformations of the base

shapes in the SHREC dataset are often relevant, inducing a strong distortion of the decorative pattern in the resulting images (see Figure 3). These deformations are actually larger than those typically visible in the application scenarios of our interest. Despite this fact, the use of the 2D SIFT on the mean curvature image was the best option tested for the SHREC'17 models' retrieval. Of course this result must be considered with care, as many implementation details (mesh discretization, resolution, smoothing, etc.) can be changed ad may influence the ability of the method to capture differences in the patterns.

- The use of convolutional layer activations of pre-trained networks as local image texture features, successful in other texture analysis applications, was not optimal in our case. This does not mean that the NN approach would not be able to improve the methods based on hand-crafted features, but just the fact that it is necessary to modify the approach by using fine-tuned networks (requiring a sufficient amount of training data) and/or exploiting different networks schemes. We plan to investigate custom CNN based relief texture characterizations in the near future.
- It is also worth noting that, using the same pooling approach, many other recently proposed vertex neighborhoods' descriptors could be tested as well, e.g. CNN features trained on local neighborhoods [MBBV15], coefficients of local shape decompositions [BDC18].
- It is not necessary to estimate the Gaussian models on the same patterns used for testing, we have shown that Gaussian parameters evaluated on a few synthetic patterns can be used as well with good results.

The geodesic patches have been chosen, in our case, also to have comparable 2D images for the texture analysis with relevant information in the same region of interest. This is a good choice for the data analyzed, but it can be limiting, in general, as it it may be necessary to analyze decorative patterns that may be overlaid on surface patches with largely variable support. However, a variety of methods can be used as well for the parametrization and the IFV descriptor does not depend on the patch/image shape and size. We plan to test and compare more parametrization options in the near future.

The Fisher Vector encoding is quite successful in texture analysis, but actually it does not capture second order properties (e.g. co-occurrence) of the basic texture elements. This can be considered a limitation of the method as it is possible to imagine decorative patterns with the same first order statistics with different spatial arrangement.

As previously pointed out, the choice of the ideal method for characterizing relief patterns also depends on the semantic grouping characterizing the relief classes to be recognized. The common shape features of the elements of the classes can be quite different depending on the specific task. This means that each classification/retrieval solution should be specifically designed for the applicative domain, as it happens for classical texture analysis methods applied to real-world tasks (e.g. medical diagnosis, material science, art, etc.).

We believe, however, that the outcomes of our analysis can be helpful for the design of successful relief patterns retrieval and classification tools for different kind of pattern groupings and applicative domains.

Acknowledgements

This work was partially supported by the Scan4Reco project funded by European Union's Horizon 2020 Framework Programme for Research and Innovation under grant agreement no 665091 and by the DSURF (PRIN 2015) project funded by the Italian Ministry of University and Research.

References

[BDC18] BÉARZI Y., DIGNE J., CHAINE R.: Wavejets: A local frequency framework for shape details amplification. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 13–24. 9

[BTA*17] BIASOTTI S., THOMPSON E. M., AONO M., HAMZA A. B., BUSTOS B., DONG S., DU B., FEHRI A., LI H., LIMBERGER F., ET AL.: Shrec'17 track: Retrieval of surfaces with similar relief patterns. In *10th Eurographics Workshop on 3D Object Retrieval* (2017). 1, 2, 3, 4, 6

[CCC*08] CIGNONI P., CALLIERI M., CORSINI M., DELLEPIANE M., GANOVELLI F., RANZUGLIA G.: Meshlab: an open-source mesh processing tool. In *Eurographics Italian Chapter Conference* (2008), vol. 2008, pp. 129–136. 4

[CMK*14] CIMPOI M., MAJI S., KOKKINOS I., MOHAMED S., VEDALDI A.: Describing textures in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on* (2014), IEEE, pp. 3606–3613. 3, 4, 5

[CMKV16] CIMPOI M., MAJI S., KOKKINOS I., VEDALDI A.: Deep filter banks for texture recognition, description, and segmentation. *International Journal of Computer Vision 118*, 1 (2016), 65–94. 3, 5

[DK12] DAROM T., KELLER Y.: Scale-invariant features for 3-d mesh models. *IEEE Transactions on Image Processing 21*, 5 (2012), 2758–2769. 3, 5

[FFP05] FEI-FEI L., PERONA P.: A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (2005), vol. 2, IEEE, pp. 524–531. 6

[FH05] FLOATER M. S., HORMANN K.: Surface parameterization: a tutorial and survey. In *Advances in multiresolution for geometric modelling*. Springer, 2005, pp. 157–186. 4

[FJ02] FIGUEIREDO M. A. T., JAIN A. K.: Unsupervised learning of finite mixture models. *IEEE Transactions on pattern analysis and machine intelligence 24*, 3 (2002), 381–396. 5

[GKM*02] GADELMAWLA E., KOURA M., MAKSOUD T., ELEWA I., SOLIMAN H.: Roughness parameters. *Journal of Materials Processing Technology 123*, 1 (2002), 133–145. 1

[GLD*10] GODIL A., LIAN Z., DUTAGACI H., FANG R., VANAMALI T., CHEUNG C. P.: Benchmarks, performance evaluation and contests for 3d shape retrieval. In *Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop* (2010), ACM, pp. 42–47. 3

[LCF*18] LIU L., CHEN J., FIEGUTH P., ZHAO G., CHELLAPPA R., PIETIKAINEN M.: A survey of recent advances in texture representation. *arXiv preprint arXiv:1801.10324* (2018). 2

[LM16] LIN T.-Y., MAJI S.: Visualizing and understanding deep texture representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2791–2799. 3

[Low99] LOWE D. G.: Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (1999), vol. 2, Ieee, pp. 1150–1157. 3, 4

[MBBV15] MASCI J., BOSCAINI D., BRONSTEIN M. M., VANDERGHEYNST P.: Shapenet: Convolutional neural networks on non-euclidean manifolds. *CoRR abs/1501.06297* (2015). URL: http://arxiv.org/abs/1501.06297, arXiv:1501.06297. 4, 9

[OVSP13] OTHMANI A., VOON L. F. L. Y., STOLZ C., PIBOULE A.: Single tree species classification from terrestrial laser scanning data for forest inventory. *Pattern Recognition Letters 34*, 16 (2013), 2144–2150. 1, 2

[PMCS11] PIETRONI N., MASSIMILIANO C., CIGNONI P., SCOPIGNO R.: An interactive local flattening operator to support digital investigations on artwork surfaces. *IEEE transactions on visualization and computer graphics 17*, 12 (2011), 1989–1996. 2

[PSM10] PERRONNIN F., SÁNCHEZ J., MENSINK T.: Improving the fisher kernel for large-scale image classification. In *European conference on computer vision* (2010), Springer, pp. 143–156. 3

[PWM*16] PHILLIPS S. C., WALLAND P. W., MODAFFERI S., DORST L., SPAGNUOLO M., CATALANO C. E., OLDMAN D., TAL A., SHIMSHONI I., HERMON S.: Gravitate: geometric and semantic matching for cultural heritage artefacts. In *Proceedings of the 14th Eurographics Workshop on Graphics and Cultural Heritage* (2016), Eurographics Association, pp. 199–202. 1

[SKVS13] SMEETS D., KEUSTERMANS J., VANDERMEULEN D., SUETENS P.: meshsift: Local surface features for 3d face recognition under expression variations and partial data. *Computer Vision and Image Understanding 117*, 2 (2013), 158–169. 3

[SLF*16] SONG Y., LI Q., FENG D., ZOU J. J., CAI W.: Texture image classification with discriminative neural networks. *Computational Visual Media 2*, 4 (2016), 367–377. 3

[VF08] VEDALDI A., FULKERSON B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008. 4

[WBDB15] WERGHI N., BERRETTI S., DEL BIMBO A.: The mesh-lbp: a framework for extracting local binary patterns from discrete manifolds. *IEEE Transactions on Image Processing 24*, 1 (2015), 220–235. 2

[ZBVH09] ZAHARESCU A., BOYER E., VARANASI K., HORAUD R.: Surface feature detection and description with applications to mesh matching. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (2009), IEEE, pp. 373–380. 2