Pattern Recognition 100 (2020) 107174

Contents lists available at ScienceDirect

Pattern Recognition

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

No-reference mesh visual quality assessment via ensemble of convolutional neural networks and compact multi-linear pooling

Ilyass Abouelaziz[a,*], Aladine Chetouani[b], Mohammed El Hassouni[e], Longin Jan Latecki[c], Hocine Cherifi[d]

[a] Mohammed V University in Rabat, Morocco
[b] University of Orleans PRISME laboratory Orleans, France
[c] Department of Computer and Information Sciences, Temple University, Philadelphia, USA
[d] LE2I, University of Burgundy, UMR 6306 CNRS, Dijon, France
[e] FLSH, Mohammed V University in Rabat, Morocco

ARTICLE INFO

Article history:
Received 17 June 2019
Revised 4 October 2019
Accepted 15 December 2019
Available online 25 December 2019

Keywords:
Blind mesh quality assessment
Convolutional neural network
Fine-tuning
Compact multi-linear pooling
Visual saliency

ABSTRACT

Blind or No reference quality evaluation is a challenging issue since it is done without access to the original content. In this work, we propose a method based on deep learning for the mesh visual quality assessment without reference. For a given 3D model, we first compute its mesh saliency. Then, we extract views from the 3D mesh and the corresponding mesh saliency. After that, the views are split into small patches that are filtered using a saliency threshold. Only the salient patches are selected and used as input data. After that, three pre-trained deep convolutional neural networks are employed for feature learning: VGG, AlexNet, and ResNet. Each network is fine-tuned and produces a feature vector. The Compact Multi-linear Pooling (CMP) is used afterward to fuse the retrieved vectors into a global feature representation. Finally, fully connected layers followed by a regression module are used to estimate the quality score. Extensive experiments are executed on four mesh quality datasets and comparisons with existing methods demonstrate the effectiveness of our method in terms of correlation with subjective scores.

1. Introduction

The digital representation of 3D shapes is widely depicted by 3D triangular meshes, and recently have been used in a broad range of computer vision applications [1]. Usually, 3D triangular meshes describing a 3D shape go through some lossy operations in order to make the transmission more sophisticated [2,3], store and render 3D meshes, or to protect the models by a copyright [4,5]. However, the perceived quality of the mesh is influenced because of the distortions introduced by these operations. To quantify the degree of distortion, subjective evaluations can be conducted by directly asking human subjects to give their estimation of the perceived visual quality. Obviously, this evaluation is unpractical in most real-world applications as it is generally costly in terms of time and human resources. Thus, it is crucial to adopt objective quality metrics that try to mimic an ideal human observer and accurately predict the subjective assessment scores [6].

As for 2D images and videos, we can classify the objective methods depending on the accessible data about the reference: a metric is called no reference (NR) or blind when the evaluation is done without having the reference data. a metric is called reduced-reference (RR) or full reference (FR) respectively when the evaluation is done using the reference data, partially or totally.

The classical methods Root Mean Squared error (RMS) [7] and the Hausdorff distance [8] use simple geometric distances to compare a distorted mesh and its reference with the same connectivity. These methods neglect the perceptual information describing the Human Visual System (HVS) since they compute a pure geometric distance. Thus, the prediction is not always reliable as proven by the moderate correlation with human perception [9,10]. In order to include the perceptual information, many researchers have recently developed perceptually driven quality methods for Mesh Visual Quality (MVQ) assessment.

Although tremendous advance has been achieved in the last decade in objective image visual quality (IVQ) assessment, the

* Corresponding author.
E-mail addresses: ilyass.abouelaziz@um5s.net.ma (I. Abouelaziz), aladine.chetouani@univ-orleans.fr (A. Chetouani), mohamed.elhassouni@um5.ac.ma (M. El Hassouni), latecki@temple.edu (L.J. Latecki), hocine.cherifi@u-bourgogne.fr (H. Cherifi).

https://doi.org/10.1016/j.patcog.2019.107174
0031-3203/

researches on objective MVQ assessment are still in the early stage, with very few metrics proposed, especially blind metrics.

In this paper, we present a novel method for blind MVQ assessment. The proposed method is based on the combination of three automatic learned feature vectors provided by three Convolutional Neural Networks (CNN) models. The combination step is based on the Compact Bi-linear Pooling (CBP) that has been extended to Compact Multi-linear Pooling (CMP) in order to consider more than two feature vectors with a multiplicative interaction between all the extracted features [11]. The organization of the paper is as follows: we present in Section 2 the most relevant objective metrics for MVQ assessment and the motivation behind our contribution. A detailed description of the proposed method is illustrated in Section 3. Then, we provide experiments, results and discussions in Section 4. Finally, Section 5 is dedicated to some concluding points and perspectives.

2. Related work

In the literature, several methods have been proposed for the MVQ prediction. In 2000, Karni and Gotsman [12] developed an MVQ metric for the evaluation of a mesh compression algorithm. The evaluation is performed by computing a distance between the model and its distorted version. The distance is obtained by computing the weighted sum of the vertex root mean squared error and the vertex Laplacian coordinate error. An improved version has been proposed by Sorkine et al. [13]. This method gives more weight to the Laplacian values. The strain energy field (SEF) is introduced by a specific mesh deformation in the method of Bian et al. [14]. The visual deformation is considered as a level of energy and the perceptual distance is computed as the level of strain energy of the normalized triangular faces. Based on the well-known structural 2D quality metric, so-called SSIM (Structural SIMilarity) [15], Lavoué et al. [16] proposed a metric named Mesh Structural Distortion Measure (MSDM). The latter extended the principle of SSIM for images to 3D meshes. It uses the mesh mean curvature as an alternative for the pixel intensities in the SSIM index. This method supposes that the compared meshes have the same connectivity, which is considered a limitation in MVQ assessment since it is not always the case. An improved version so-called MSDM2 has been proposed in order to overcome the connectivity issue [17]. For that, the authors introduced a vertex correspondence processing step. Multi-scale analysis has also considered to evaluate the visual difference which leads to a considerable amelioration in predicting the perceived visual quality. Torkhani et al. [18] proposed a quality method called Tensor-based Perceptual Distance Measure (TPDM). This method computes a distance between curvature tensors of the compared meshes. The curvature amplitudes and the principal curvature directions, which are obtained from the tensor eigenvalues and eigenvectors respectively, are used to compute a perceptually-oriented tensor distance. Recently, Yildiz et al. [19] proposed a full-reference perceptual quality metric for animated meshes to predict the visibility of local distortions on the mesh surface. Feng et al. [20] proposed a method called TPDMPW based on percentile weighting strategy. Chetouani [21] proposed a quality measure based on the fusion of selected features and the support vector regression (SVR). The mesh roughness is used by Corsini et al. [22] and Gelasca et al. [23]. In their methods, they consider the relationship between the perceived visual quality and the roughness on the surface. The former used dihedral angles computed as the angle between normals of two adjacent faces, while the latter compute the difference between the original model and the smoothed version. Dihedral angles also used in the work of Vasa and Rus [24]. They proposed an MVQ assessment metric that computes the difference between the original model and its deformation based on oriented angles.

FR and RR methods are mostly employed for guiding mesh compression and watermarking. These approaches successfully estimate the quality scores as proven by the high correlations. However, in practice, the reference is not always available.

NR or blind quality assessment is an alternative solution to overcome this issue. In this context, machine learning becomes trending in score predictions and it is successfully used for image quality assessment (IQA) [25–30].

In the same vein, several NR metrics have been developed for MVQ assessment. Abouelaziz et al. [31] proposed a blind method that relies on the mean curvature features and the General Regression Neural Network (GRNN) for the feature learning and the quality prediction. Visual saliency and Support Vector Regression (SVR) are used in the method of Nouri et al. [32]. The authors proposed a NR method called 3D Blind Mesh Quality Assessment Index (BMQI). SVR is also used in [33,34]. In [35], the CNN is fed by perceptual hand-crafted features (i.e dihedral angles and mesh shape) extracted from the 3D mesh and presented into 2D patches of a fixed size. In [36], the CNN is fed by rendered images from the 3D object, the view is changed by rotating the 3D mesh by an angle of 60 degrees according to the X and Y axes. In our primary work [37], we introduced a patch-selection strategy based on mesh saliency in order to give more importance to the attractive regions.

Motivated by the promising result of the above methods, we propose a deep-based method for the quality assessment of 3D models without reference. More precisely, we extract feature vectors from three different CNN models and we extend the Compact Bi-linear Pooling concept to Compact Multi-linear Pooling in order to combine the extracted features. In addition, we apply a saliency-based selection strategy that aims to focus more on the perceptual relevant regions.

The novelties of our method consist of using fine-tuned deep CNNs for quality assessment and the Compact Multi-linear Pooling for the feature combination. In addition we conduct extensive experiments on four databases: the effect of mesh saliency in MVQ assessment and comparison of many combination strategies (concatenation, element-wise multiplication and CMP).

Details about the proposed method are provided in the next section.

3. Proposed method

3.1. Flowchart

The proposed method consists of two major parts as depicted in Fig. 1: learning data preparation and feature learning for the quality score estimation. The data preparation module aims to focus only on the more attractive regions using the saliency information, while the feature learning and quality score estimation module aim to assess the quality by combining three automatic learned features given by three different CNN models.

3.2. Learning data preparation

In order to focus on regions that impact more the subjective quality, we apply a saliency-based patch selection strategy. For that, we first compute the mesh saliency of the 3D model. Then, 2D views are rendered from the 3D model and its mesh saliency. The rendered views are split into small patches and the saliency of each of them is analyzed to select only the more relevant ones. After a local normalization, the selected patches feed the CNN model for the training. In this section, we provide more details for each step.

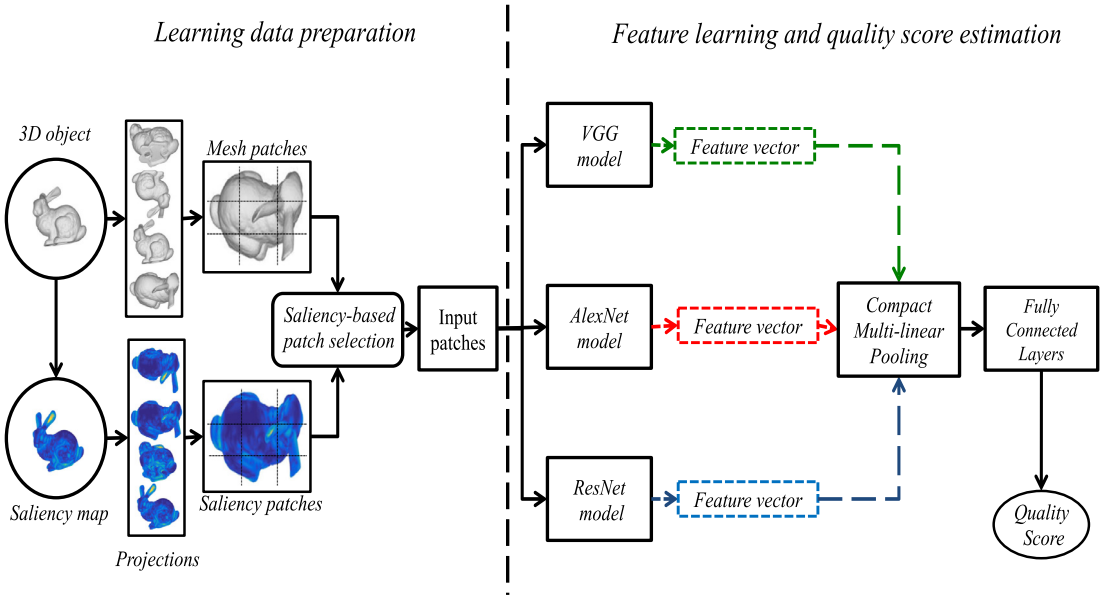

Fig. 1. Flowchart of the proposed blind mesh visual quality assessment method.

3.2.1. Mesh visual saliency

Saliency is a perceptual concept that describes the attention of our HVS to some regions due to its specificities (curvature, orientation and so on). In this work, mesh visual saliency is employed to select patches that are perceptually relevant using the method proposed in [38]. It is based on the mechanism adopted by the well-known method proposed by Itti et al. [39] for 2D images. The first step to obtain the mesh saliency is to compute the mean curvature at mesh vertices. After that, fine and coarse Gaussians are used to filter the mean curvatures and the saliency is obtained by computing the difference between the filtered mean curvatures within different scales. Finally, a non-linear normalization sum of all the multi-scale saliency maps is applied in order to compute the final saliency map. It is noteworthy that our contribution is not to implement a saliency method, but rather to use an existing one to demonstrate the usefulness of mesh visual saliency in quality assessment. The relevance of this step is discussed in Section 4.2.

3.2.2. 2D Projections rendering and patch selection

Once the 3D visual saliency is computed, the next step consists of rendering 2D projections from the 3D mesh and its corresponding mesh saliency for a multiple views representation. To this end, we surround the 3D mesh by several virtual cameras at different angles according to the axes X and Y. The centroid of the 3D object is placed at the origin of the coordinate system. The coordinates (x_i, y_i) of the virtual cameras are obtained by varying the angles $x \in [0, 2\pi]$ and $y \in [0, 2\pi]$ by $\frac{\pi}{6}$ (30°). In total, 144 projections are obtained from each 3D mesh. We note that we use only the axes X and Y since using also the Z axis duplicates the views and provides redundant and useless information. In our projection strategy, the obtained projections describe the 3D object from all important views. Moreover, in view-based 3D shape retrieval also only X- and Y-rotations are used to generate 2D views of 3D objects [40,41]. The obtained views are then split into small patches. After that, a Local level of Saliency (LoS) is computed for each patch and used to select the important patches with a saliency threshold S_t sets experimentally (more details can be found in Section 4.2).

The selection is performed by keeping patches with $LoS \geq S_t$ that are considered salient patches. It is noteworthy that the LoS is computed using only the pixels that contain the saliency information, the background pixels at object boundary are not considered and informative patches (with high saliency) at object boundary are not ignored.

3.2.3. Patch normalization

The next step is to apply a simple local contrast normalization on the selected patches. The normalized value $\hat{I}(i, j)$ of a pixel $I(i, j)$ at location (i, j) is computed as follows:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + c} \tag{1}$$

$$\mu(i,j) = \frac{1}{(2M+1)\times(2N+1)} \sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} I(i+m, j+n) \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{m=-M}^{m=M} \sum_{n=-N}^{n=N} (I(i+m, j+n) - \mu(m,n))^2} \tag{3}$$

where c is a constant that prevents instabilities from dividing by zero. M and N are the normalization window sizes. The used normalization is crucial to make the trained networks robust to illumination and contrast variation by decreasing the effect of the saturation problem [27].

3.3. Feature learning and quality score estimation

In this work, we propose to fuse automatically learned features extracted from CNN models. To this end, we fine-tune three well-known pre-trained CNN models (AlexNet, VGG, and ResNet). A feature vector is then extracted for each network and a defined combination is applied to obtain a global feature vector to be used for the quality prediction. In the next, we give more details about the used networks and the combination strategies.

3.3.1. Deep convolutional neural networks

The CNN models used in this work are briefly described as follows:

- **AlexNet** [42]: This CNN model, proposed by Alex Krizhevsky, is the winner of the Image Large Scale Visual Recognition Challenge (ILSVRC) in 2012. It consists of five convolutional layers, max-pooling layers, and three fully connected layers. The dropout regularization method is used in the fully connected

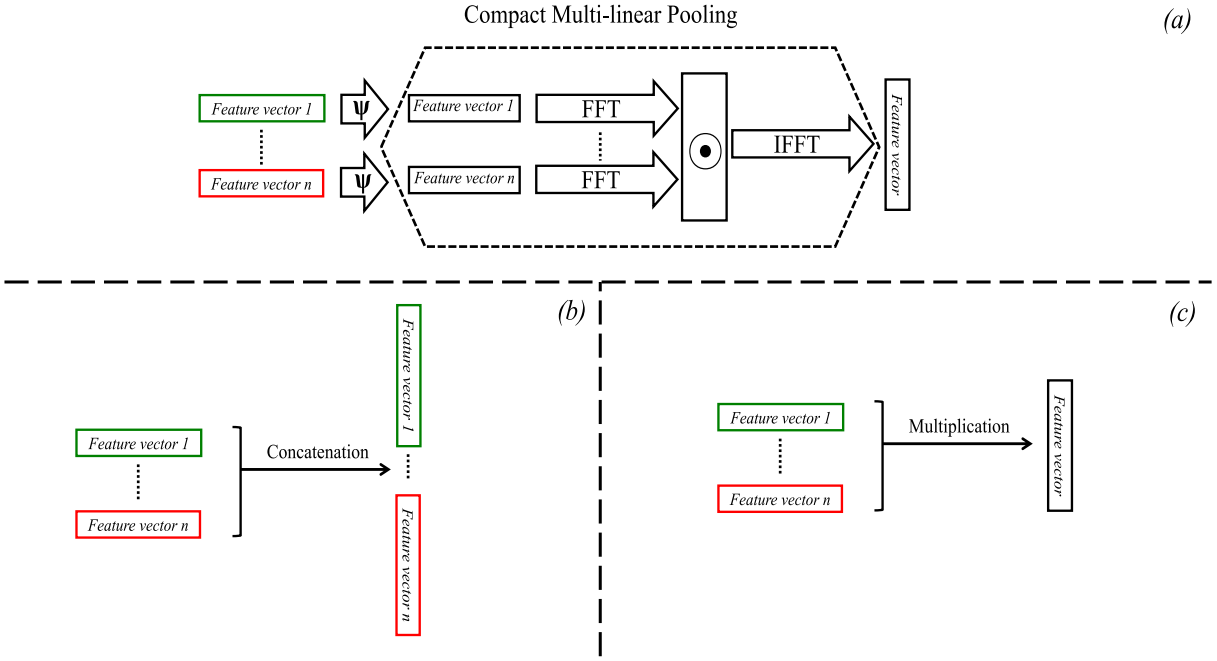

Fig. 2. Combination strategies.

layers to prevent overfitting. In addition, the authors highlight the use of the ReLU function and the overlap in the max-pooling layers.

- **VGG** [43]: VGG network is a deep CNN proposed by the Oxford Visual Geometry Group. The network achieved successful performance in ILSVRC 2014. Several versions of VGG have been developed with different convolutional layers: VGG11, VGG13, VGG16, and VGG19. In our method, we use VGG16. This network consists of 13 convolutional layers with max-pooling and three fully connected layers.
- **ResNet** [44]: The residual Neural Network (ResNet) has been proposed by Kaimimg He et al. in 2015. The network is developed to make the training of deep networks easier by accelerating the speed of the training. Different versions of ResNet have been proposed: ResNet 18, ResNet 34, ResNet 52, and others. The used network consists of 16 convolutional layers with max-pooling and two fully connected layers.

It is worth noting that the last layer is a regression since the quality scores are seen as "continuous values". In addition, the input of the pre-trained network is adjusted to be fed by patches of a fixed size (patch-size $= 32 \times 32$).

3.3.2. Compact multi-linear pooling (CMP)

Once the feature vectors are extracted from the above-described models, we combine them using the concept of Compact Bi-linear Pooling (CBP) [45] that computes the outer product of two feature vectors u and v. The authors demonstrate that this outer product can be seen as a convolution ($\circ ledast$) when the Count Sketch projection function is applied. This latter aims to project the feature vectors into a lower dimensional feature space. Moreover, as the convolution of two vectors is equivalent to the element-wise product in the frequency domain, the outer product $u \circ ledastv$ can be finally rewritten as $FFT^{-1}(FFT(u) \odot FFT(v))$, where $\odot$ refers to the element-wise product and FFT designs the Fast Fourier Transform. This combination is very important because it permits the interaction of all elements of the vectors in a multiplicative way without a high computational cost.

In order to combine more than two vectors, the process is extended to Compact Multi-linear Pooling (CMP) [46] as depicted in Fig. 2a. The CMP combination consists firstly of projecting the feature vectors to a lower dimensional feature space through a Count Sketch projection function. After computing the FFT of each considered feature vector, we multiply the obtained spectra and apply the Inverse Fast Fourier Transform (IFFT) to obtain a single feature vector.

In Section 4.3), the performance of CMP strategy is compared to some common combination, described as follows:

- **Concatenation (see** Fig. 2b): The simplest way to combine vectors is to concatenate them. It allows all the elements to interact in the learning, however, the result vector contains more elements and can slow down the prediction time.
- **Element-wise multiplication (see** Fig. 2c): The feature vectors can be combined by multiplying their elements, the result vector is of a smaller size than the concatenated one, however, not all the elements interact together in the learning process and some information can be lost.
- 1×1 **convolution:** another type of combination is to use 1×1 convolution filter. It creates a linear projection of the features and is used to reduce the data size.

CMP is interesting because it allows the interaction of all elements of the vectors in a multiplicative way, which is not the case in the element-wise multiplication. Besides, unlike the bi-linear pooling, the CMP projects the outer product to a lower-dimensional space that leads to a lower computation time. After the CMP strategy is applied, a global feature vector is obtained, and it is fed to fully connected layers followed by a regression layer for the quality score estimation.

3.3.3. Training settings

The final step of our method consists of predicting the perceived quality score. In this section, we describe the followed training settings. For the training process, we use the objective function adopted in [27] defined as follows:

$$L = \frac{1}{N}\sum_{n=1}^{N} ||S(p_n;\omega) - MOS_n||_{l1} \tag{4}$$

$\hat{\omega} = \min_{\omega} L$ where MOS_n denotes the subjective MOS related to a specific input patch p_n and $S(p_n; \omega)$ is the estimated objective score of p_n with network weights ω. The Stochastic Gradient Descent (SGD) and back propagation are used to learn the parameters of the CNN by minimizing the objective function (Eq. (4)). The batch size and the learning rate are fixed respectively to 32 and 0.01. The maximum number of epochs has been fixed to 40.

We applied the leave-one-out cross-validation to train and test our method. A training model is built using all the distorted meshes in the repository except one mesh and its distorted versions. We then use the constructed model on the excluded meshes to estimate the quality. Each patch is labeled by a quality score, the same as the ground truth score of the whole mesh as commonly used in image quality assessment in [47].

4. Experiments

In order to verify its performance, the proposed method has been extensively tested and compared with existing MVQ metrics. We begin by describing the used databases and the criteria used to evaluate our method. Then, we investigate the importance of including the visual saliency-based patch selection technique in our method and how the performances of the networks are affected. We also compared the performance obtained by different combination strategies. Finally, we present the experimental results and comparative analysis on mesh visual quality assessment state of the art.

4.1. Datasets

The experiments and tests are conducted on four subjectively-rated mesh visual quality databases:

- **LIRIS/EPFL general-purpose database** [23]: It is the largest available dataset in MVQ. It has four reference meshes, each one with 21 distortion versions. These latter are obtained by applying smoothing and noise addition in different regions (local and global distortions). Subjective scores were provided by 12 experts.
- **LIRIS masking database** [48]: This database is designed to study the influence of visual masking in MVQ metrics. It has four references with 6 distorted versions obtained by adding a local noise. Subjective scores were provided by 11 experts.
- **IEETA simplification database** [24]: This database contains five references with 30 distorted versions obtained by applying three simplification algorithms. Two different vertex reduction ratios were used to simplify the original meshes. 65 observers participated in the subjective study.
- **UWB compression database** [49]: This database contains five references with 64 distorted models. 13 types of compression are applied for each original model by different algorithms. 69 observers participated in the subjective study.

The coherence between the subjective scores (ground truth scores) and the objective scores (predicted scores by MVQ methods) is computed by correlation measures. In our experiments we adopt two coefficients: the first one called Pearson Linear Correlation Coefficient (denoted r_p), it measures the accuracy of the score estimation. The second coefficient called Spearman Rank-Order Correlation Coefficient (denoted r_s) measuring the monotonicity of the estimations [50]. The compared scores are generally non-linear, thus, it is highly recommended to apply a psychometric fitting function. Similarly to Torkhani et al. [18], Wang et al. [51], a cumulative Gaussian psychometric function [52] is used:

$$p(m, n, X) = \frac{1}{\sqrt{2\pi}} \int_{m+nX}^{\infty} \exp - \left(\frac{t^2}{2}\right) dt \tag{5}$$

Where X denotes the estimated scores by our method, m and n presents two fitting parameters computed for each MVQ dataset using the predicted scores and their corresponding subjective scores.

4.2. Effect of the saliency-based patch selection

As mentioned earlier, the patch selection strategy is based on the mesh saliency obtained from the distorted meshes. The patches are selected by fixing a saliency threshold. To demonstrate the importance of the patch selection strategy used in our method, we conduct an experiment by testing the ability of the fine-tuned networks to predict the perceived visual quality with different saliency threshold on the General-purpose database. In addition, we compare the performance of the networks with and without using the selection strategy ($S_t = 0$) on the four databases. Fig. 3 presents the correlation coefficients obtained by the trained networks with different saliency thresholds on the General purpose database. We note that the tests are conducted only on this database since it contains the greater number of meshes. After that, the saliency threshold is generalized for the other databases.

When $S_t = 0.4$, the correlation scores obtain their best value for all the networks. However, when $S_t > 0.4$, the number of selected patches decreases clearly, especially when $S_t = 0.9$ (only patches with a saliency superior to 0.9 are selected). Thus, the performance, in this case, is the worst. When $S_t < 0.4$, the correlation scores decrease since the selection is less important in this case and this refers to the influence of the patch selection strategy in our method. Comparing to $S_t = 0$ (without selection), it is remarkable that the patch selection strategy significantly improves the performance as proven by the difference between the score regarding $S_t = 0$ and $S_t = 0.4$ (best saliency threshold). The saliency threshold is thus fixed to 0.4

The importance of the saliency selection is confirmed in Table 1 that presents the averaged scores obtained by the fine-tuned networks with and without the selection over all databases.

The correlation scores increase remarkably when using the saliency selection. For all networks and databases, the improvement is between 2.3% (AlexNet r_p on the simplification database) and 7.2% (AlexNet r_s on the general database).

From these results, we conclude that the used patch selection strategy based on visual saliency is very effective, especially on the LIRIS masking and the General-purpose databases.

4.3. Performances with different combination strategies

After the saliency threshold is fixed, we examine in this section the performance of our method according to the different combination strategies (concatenation, element-wise multiplication and CMP discussed in Section 3.3.2). We tried all the possible combinations using the three above described models that lead to four experiments by a combination type. In addition, we compute the mean correlation scores for all the databases. Table 2 presents the correlation coefficients r_s (%) and r_p (%) of the fine-tuned networks on described databases with the different combinations.

In general, the three combination types provide excellent performances on all the databases as proven by the high correlation scores.

- The multiplication combination provides the lowest scores, especially when combining the three networks. Multiplying three vectors amplifies their values and leads to considerable modifications on the extracted features. Thus, the estimation is less reliable comparing to the other combinations.
- The concatenation allows obtaining an extended feature vector. This combination provides high scores and outperforms

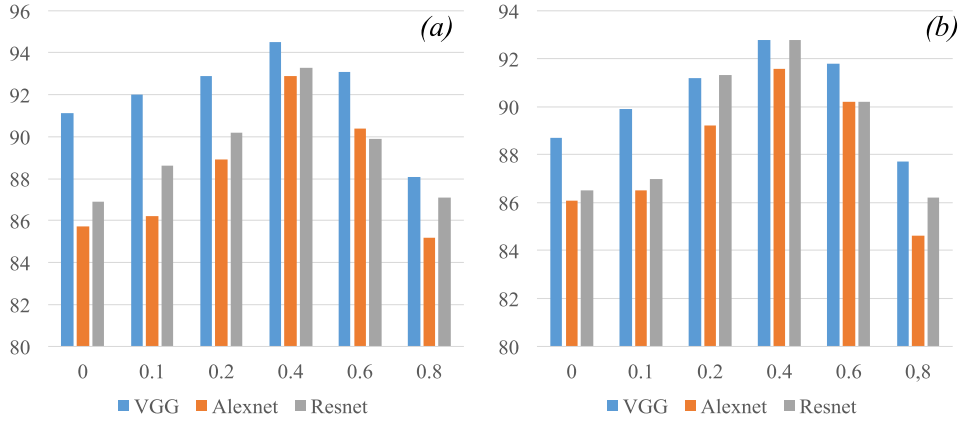

Fig. 3. Correlation coefficients r_s (%) (a) and r_p (%) (b) of the fine-tuned networks using different saliency threshold with patch size 32 × 32 on the LIRIS/EPFL general-purpose database.

Table 1
Correlation coefficients r_s (%) and r_p (%) of the proposed method with and without the patch selection strategy on the four tested databases.

		Masking		General-Purpose		Compression		Simplification	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Without patch selection	VGG	92.2	91.0	91.1	88.7	85.3	84.2	84.1	83.8
	AlexNet	89.6	88.6	85.7	86.1	89.1	88.4	90.1	89.1
	ResNet	88.9	89.7	86.9	86.5	86.1	86.6	88.2	87.3
With patch selection	VGG	96.2	94.2	94.5	92.8	89.5	86.7	89.8	88.3
	AlexNet	95.4	93.3	92.9	91.6	92.2	91.4	92.5	91.4
	ResNet	93.4	92.2	93.3	92.8	90.4	89.9	91.6	90.4

Table 2
Correlation coefficients r_s (%) and r_p (%) of the fine-tuned networks on LIRIS masking database, LIRIS/EPFL general-purpose database, the UWB compression database and the IEETA simplification database using different combination strategies.

Combination type	Networks	Masking		General-Purpose		Compression		Simplification		Mean scores	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
No combination	VGG	96.2	94.2	94.5	92.8	89.5	86.7	89.8	88.3	92.5	90.5
	AlexNet	95.4	93.3	92.9	91.6	92.2	91.4	92.5	91.4	93.2	92.0
	ResNet	93.4	92.2	93.6	92.8	90.4	89.9	**91.6**	90.4	93.1	91.5
Concatenation	VGG + AlexNet	95.9	95.3	93.0	91.2	89.9	88.6	87.6	85.4	91.6	90.1
	VGG + ResNet	94.2	93.9	92.6	91.6	88.5	88.1	86.5	86.1	90.4	89.9
	AlexNet + ResNet	93.2	91.6	91.9	90.8	90.1	88.9	88.1	88.3	90.8	89.9
	All networks	**96.3**	95.1	93.6	91.9	90.5	89.3	90.0	90.2	92.6	91.6
Multiplication	VGG + AlexNet	90.1	88.9	88.1	86.9	85.2	84.2	86.3	85.6	87.4	86.4
	VGG + ResNet	91.5	90.0	87.6	88.3	84.3	84.8	89.2	88.6	88.1	87.9
	AlexNet + ResNet	89.6	90.6	88.6	87.3	86.1	84.0	84.5	83.0	87.2	86.2
	All networks	88.6	86.9	86.0	85.2	84.9	82.6	83.5	81.9	85.7	84.1
1 × 1 Convolution	All networks	90.3	90.1	91.0	89.6	85.4	86.5	86.2	85.8	88.2	88.0
	VGG + ResNet	93.7	94.3	88.9	87.8	86.7	88.8	89.7	89.0	89.8	90.0
	AlexNet + ResNet	90.6	90.4	91.0	92.6	85.8	86.3	87.5	88.9	88.7	89.5
	All networks	91.2	92.8	89.8	91.2	86.3	89.0	85.9	88.3	88.3	90.3
Compact multi-linear pooling	VGG + AlexNet	94.8	95.0	92.6	93.2	91.3	92.6	90.7	89.1	92.3	92.5
	VGG + ResNet	94.5	93.1	91.9	92.6	90.8	90.3	90.1	89.6	91.8	91.4
	AlexNet + ResNet	94.5	93.9	93.6	93.5	**93.2**	92.2	91.0	90.8	93.0	92.6
	All networks	95.8	**95.9**	**94.4**	**94.8**	92.7	**93.8**	91.0	**91.1**	**93.3**	**93.8**

the multiplication with a considerable correlation scores improvement reaches 8.3% (All networks r_p on the simplification database).
- 1 × 1 convolution performs better than the multiplication with an improvement up to 6.2% (All networks r_p on the mean score) but not as good as the concatenation.
- The CMP combination provides the highest scores in most situations: the highest performance on the General-purpose database, the highest r_p score and the second r_s score with a slight difference on the other databases.

From the above observations, we conclude that the best combination strategy is the CMP as proven also by the high scores when averaging over all the databases. In the following, we adopt the CMP using the three networks for the comparison with the state-of-the-art.

4.4. Evaluation and comparison with the state-of-the-art

In this section, we conduct a comparative study of the proposed method with the state of the art including FR (HD [8], RMS [7], MSDM2 [17], TPDM [18], Yildiz et al. [19], TPDMPW [20], Chetouani [21]), RR (3DWPM1 [23], 3DWPM2 [23], FMPD [51], DAME [24]) and NR (NR-SVR [33], NR-GRNN [31], NR-CNN1 [35], NR-CNN2 [36], BMQI [32]). The correlation coefficients values r_s and r_p on

Table 3
Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the LIRIS/EPFL general-purpose database.

Type	Metric	Armadillo		Dyno		Venus		Rocker		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full Reference	HD [8]	69.5	30.2	30.9	22.6	1.6	0.8	18.1	5.5	13.8	1.3
	RMS [7]	62.7	32.2	0.3	0.0	90.1	77.3	7.3	3.0	26.8	7.9
	MSDM2 [17]	81.6	72.8	85.9	73.5	89.3	76.5	89.6	76.1	80.4	66.2
	TPDM [18]	84.5	78.8	92.2	89.0	90.6	91.0	92.2	91.4	89.6	86.2
	Yildiz et al. [19]	-	86.0	-	79.0	-	89.0	-	88.0	-	-
	TPDMPW [20]	-	-	-	-	-	-	-	-	87.2	87.7
	Chetouani [21]	75.7	86.1	90.6	90.0	94.9	95.5	91.4	92.1	88.1	90.9
Reduced Reference	3DWPM1 [23]	65.8	35.7	62.7	35.7	71.6	46.6	87.5	53.2	69.3	38.4
	3DWPM2 [23]	74.1	43.1	52.4	19.9	34.8	16.4	37.8	29.9	49.0	24.6
	FMPD [51]	75.4	83.3	89.6	88.9	87.5	83.9	88.8	84.7	81.9	83.5
	DAME [24]	60.3	76.3	92.8	88.9	91.0	83.9	85.0	80.1	76.6	75.2
No-Reference	NR-SVR [33]	76.8	91.5	78.6	84.1	85.7	88.6	86.2	86.6	81.5	87.8
	NR-GRNN [31]	87.1	**97.3**	91.2	**94.1**	86.3	85.0	78.6	74.8	86.2	88.7
	NR-CNN1 [35]	87.2	84.3	86.4	86.2	92.2	85.6	91.3	85.2	83.6	82.7
	NR-CNN2 [36]	93.4	95.6	86.2	84.3	**94.1**	90.3	80.4	82.2	81.8	82.5
	BMQI [32]	20.1	-	83.5	-	88.9	-	92.7	-	78.1	-
	Our method	**95.8**	95.6	**93.6**	92.9	93.4	**91.3**	**94.5**	**95.2**	**94.4**	**94.8**

Table 4
Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the LIRIS masking database.

Type	Metric	Armadillo		Lion		Bimba		Dyno		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full Reference	HD [8]	48.6	37.7	71.4	25.1	25.7	7.5	48.6	31.1	26.6	4.1
	RMS [7]	65.7	44.6	71.4	23.8	71.4	21.8	71.4	50.3	48.8	17.0
	MSDM2 [17]	88.6	65.8	94.3	87.5	**100**	93.7	**100**	91.7	89.6	76.2
	TPDM [18]	88.6	91.4	82.9	88.4	**100**	97.2	**100**	**97.1**	90.0	88.6
	PDMPW [20]	-	-	-	-	-	-	-	-	94.2	91.7
	Chetouani [21]	**99.0**	**99.0**	83.0	94.0	99.0	99.0	93.0	98.0	93.9	97.8
Reduced Reference	3DWPM1 [23]	58.0	41.8	20.0	9.7	20.0	8.4	66.7	45.3	29.4	10.2
	3DWPM2 [23]	48.6	37.9	38.3	22.0	37.1	14.4	71.4	50.1	37.4	18.2
	FMPD [51]	94.2	88.6	93.5	94.3	98.9	**100**	96.9	94.3	80.8	80.2
	DAME [24]	94.3	96.0	**100**	**99.5**	97.7	88.0	82.9	89.4	68.1	58.6
No-Reference	NR-SVR [33]	89.5	84.7	**100**	96.3	94.2	93.6	94.4	89.7	90.4	91.2
	NR-GRNN [31]	82.3	80.5	94.1	97.0	90.2	94.3	78.2	82.3	90.2	82.4
	NR-CNN1 [35]	95.2	97.6	89.4	91.6	93.4	98.7	96.3	89.9	88.2	85.4
	BMQI [32]	94.3	-	94.3	-	**100**	-	83.0	-	78.1	-
	Our method	96.2	95.5	93.1	92.4	92.5	92.8	94.2	94.0	**95.8**	**95.5**

Table 5
Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the UWB compression database.

Type	Metric	Bunny		James		Jessy		Nissan		Helix		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full Reference	HD [8]	34.1	52.2	-16.8	6.8	-23.6	12.5	14.4	23.6	45.1	46.4	10.6	28.3
	RMS [7]	34.2	20.9	14.0	10.8	0.0	14.8	17.8	29.7	46.9	44.6	22.0	24.1
	MSDM2 [17]	**97.4**	90.1	82.6	69.2	84.3	63.1	84.4	73.1	98.1	94.7	89.3	78.0
	TPDM [18]	95.1	**96.5**	90.8	73.6	85.8	75.8	82.7	73.4	**98.7**	95.0	91.5	82.9
	TPDMPW [20]	-	-	-	-	-	-	-	-	-	-	91.3	**96.4**
Reduced Reference	3DWPM1 [23]	94.7	93.4	77.3	72.3	87.2	**89.5**	63.6	59.3	98.0	**95.2**	84.1	81.9
	3DWPM2 [23]	96.0	91.2	76.9	65.3	86.9	85.9	56.3	67.6	95.5	94.3	82.3	80.9
	FMPD [51]	94.2	89.6	95.3	91.2	63.3	60.0	92.4	77.5	98.4	90.8	88.8	81.8
	DAME [24]	96.8	93.4	**95.7**	**93.4**	84.4	70.5	**93.9**	75.3	96.6	**95.2**	**93.5**	85.6
No-Reference	Our method	95.6	94.8	**92.5**	90.6	92.5	87.1	88.7	**89.0**	90.7	90.4	92.7	93.8

the LIRIS masking, LIRIS/EPFL General-purpose, UWB compression and the IEETA simplification databases are listed respectively in Tables 3–6. The values of the state-of-the-art metrics are obtained from Lavoué [17] for Tables 3–5 and from Torkhani et al. [18] for Table 6.

As shown in Tables 3–6, the geometric measures HD, and RMS performs the worst. One reason is that these methods do not include the main operations of the HVS and the visual quality is computed by a simple geometric distance. For the other FR measures, MSDM2 and TPDM incorporate the perceptual information, represented in the mesh curvature. As such, the perceptual information is included and better prediction is achieved compared to the geometric measures as proven by the obtained correlation coefficients. The RR method FMPD also provides good correlations compared to MSDM2 and TPDM. This method (FMPD) includes a roughness measure which is an important feature in mesh processing. The proposed method shows excellent performance on all the available subjectively-rated MVQ databases, as proven by its high scores on the individual models as well as on the whole repositories.

- The General-purpose database (see Table 3) is the largest MVQ database so far, it comprises the highest number of distorted meshes among all the other databases (i.e 84 distorted

Table 6
Correlation coefficients r_s (%) and r_p (%) of different objective metrics on the IEETA simplification database.

Type	Metric	Bones		Bunny		Head		Lung		Strange		All models	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
Full Reference	HD [8]	94.3	84.8	39.5	14.3	88.6	53.0	88.6	64.9	37.1	27.4	49.4	25.5
	RMS [7]	94.3	71.1	77.1	79.2	42.9	23.1	94.3	71.3	94.3	92.4	64.3	34.4
	MSDM2 [17]	77.1	**96.7**	94.3	96.3	88.6	79.0	65.7	85.3	**100**	**98.1**	86.7	79.6
	TPDM [18]	**99.0**	94.3	**98.0**	94.3	63.1	65.7	**98.6**	94.3	98.7	94.3	86.9	88.2
Reduced Reference	FMPD [51]	88.6	96.0	94.3	**98.0**	65.7	70.4	88.6	**95.5**	65.7	96.0	87.2	89.3
No-Reference	Our method	91.3	88.9	91.1	92.4	**91.8**	**91.5**	95.3	89.4	91.1	88.9	**91.0**	**91.1**

meshes and a variety of distortion types). On this database, the proposed method shows good performance and provide the highest correlation coefficients ($r_s = 94.4\%$ and $r_p = 94.8\%$).

- On the LIRIS masking database (see Table 4), our method provides the highest Spearman and Pearson correlation coefficients on the whole corpus ($r_s = 95.8\%$ and $r_p = 95.5\%$) and outperforms the NR methods (BMQI, NR-SVR, and NR-GRNN) as well as the most effective FR and RR methods.
- On the UWB compression database (see Table 5), the proposed method performs the best in terms of PLCC score ($r_p = 93.8\%$) outperforming the most effective methods. In addition, it provides the second higher r_s score on the whole repository ($r_s = 92.7\%$) against $r_s = 93.5\%$ for the RR method DAME.
- On the IEETA simplification database (see Table 6), the proposed method provides the highest correlation coefficients ($r_s = 91.0\%$ and $r_p = 91.1\%$). The perceptual methods MSDM2, TPDM, and FMPD also perform well in this database.

5. Conclusion

We have proposed in this paper an effective blind objective MVQ method for the assessment of the perceived mesh visual quality. Feature vectors are first extracted using three fine-tuned CNN models and the compact multi-linear pooling is then used to fuse the extracted feature vectors into a global feature representation. In addition, 3D visual saliency is adapted to select the most relevant patches taking into account that distortions are more important in salient regions. Several tests have been conducted, in particular, we show that the patch selection strategy is very effective. Moreover, different combination strategies are tested and compared, and we show that combining multiple DCNNs increases the performances and we can derive an effective blind MVQ method. Through comparisons with the state of the art on prominent MVQ databases, it is shown that the proposed method provides high correlations with subjective scores and overcomes effective existing full reference and reduced reference methods.

References

[1] M. Botsch, L. Kobbelt, M. Pauly, P. Alliez, B. Lévy, Polygon Mesh Processing, CRC press, 2010.
[2] P. Alliez, C. Gotsman, Recent advances in compression of 3d meshes, in: Advances in Multiresolution for Geometric Modelling, Springer, 2005, pp. 3–26.
[3] H. Lee, c Dikici, G. Lavoué, F. Dupont, Joint reversible watermarking and progressive compression of 3d meshes, Visual Comput. 27 (6) (2011) 781–792.
[4] K. Wang, G. Lavoué, F. Denis, A. Baskurt, A comprehensive survey on three-dimensional mesh watermarking, IEEE Trans. Multimed. 10 (8) (2008) 1513–1527.
[5] Y.P. Wang, S.M. Hu, A new watermarking method for 3d models based on integral invariants, in: IEEE Transactions on Visualization and Computer Graphics, vol. 15, 2009, pp. 285–294.
[6] A. Bulbul, T. Capin, G. Lavoué, M. Preda, Assessing visual quality of 3-d polygonal models, IEEE Signal Process. Mag. 28 (6) (2011) 80–90.
[7] P. Cignoni, C. Rocchini, R. Scopigno, Metro: Measuring error on simplified surfaces, Computer Graphics Forum 17 (1998) 167–174. Wiley Online Library
[8] N. Aspert, D. Santa-Cruz, T. Ebrahimi, MESH: Measuring errors between surfaces using the hausdorff distance, in: Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on, 1, 2002, pp. 705–708. IEEE
[9] Z. Wang, A.C. Bovik, Mean squared error: love it or leave it? A new look at signal fidelity measures, IEEE Signal Process. Mag. 26 (1) (2009) 98–117.
[10] G. Lavoué, M. Corsini, A comparison of perceptually-based metrics for objective evaluation of geometry processing, IEEE Trans. Multimed. 12 (7) (2010) 636–649.
[11] Y. Gao, O. Beijbom, N. Zhang, T. Darrell, Compact bilinear pooling, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 317–326.
[12] Z. Karni, C. Gotsman, Spectral compression of mesh geometry, in: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co, 2000, pp. 279–286.
[13] O. Sorkine, D. Cohen-Or, S. Toledo, High-Pass Quantization for Mesh Encoding, Vol. 42, 2003, June.
[14] Z. Bian, S.M. Hu, R.R. Martin, Evaluation for small visual difference between conforming meshes on strain field, J. Comput. Sci. Technol. 24 (1) (2009) 65–75.
[15] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[16] G. Lavoué, E.D. Gelasca, F. Dupont, A. Baskurt, T. Ebrahimi, Perceptually driven 3d distance metrics with application to watermarking, SPIE Optics+ Photonics, International Society for Optics and Photonics, 2006. 63120L–63120L
[17] G. Lavoué, A multiscale metric for 3d mesh visual quality assessment, Comput. Graph. Forum 30 (2011) 1427–1437.
[18] F. Torkhani, K. Wang, J.M. Chassery, A curvature tensor distance for mesh visual quality assessment, Comput. Vis. Graph. (2012) 253–263.
[19] Z.C. Yildiz, T. Capin, A perceptual quality metric for dynamic triangle meshes, EURASIP J. Image Video Process. 2017 (1) (2017) 12.
[20] X. Feng, W. Wan, R.Y.D. Xu, S. Perry, P. Li, S. Zhu, A novel spatial pooling method for 3d mesh quality assessment based on percentile weighting strategy, Comput. Graph. 74 (2018) 12–22.
[21] A. Chetouani, Three-dimensional mesh quality metric with reference based on a support vector regression model, J. Electron. Imaging 27 (4) (2018) 43048.
[22] M. Corsini, M.-C. Larabi, G. Lavoué, O. Petřík, L. Váša, K. Wang, Perceptual metrics for static and dynamic triangle meshes, Comput. Graph. Forum 32 (2013) 101–125.
[23] M. Corsini, E.D. Gelasca, T. Ebrahimi, M. Barni, Watermarked 3d mesh quality assessment, IEEE Trans. Multimed. 9 (2) (2007) 247–256.
[24] L. Váša, J. Rus, Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes, Comput. Graph. Forum 31 (2012) 1715–1724.
[25] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17 (5) (2010) 513–516.
[26] C. Li, A.C. Bovik, X. Wu, Blind image quality assessment using a general regression neural network, IEEE Trans. Neural Netw. 22 (5) (2011) 793–799.
[27] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[28] J. Gu, G. Meng, S. Xiang, C. Pan, Blind image quality assessment via learnable attention-based pooling, Pattern Recognit. 91 (2019) 332–344.
[29] W. Zhang, C. Qu, L. Ma, J. Guan, R. Huang, Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network, Pattern Recognit. 59 (2016) 176–187.
[30] P. Chen, L. Li, X. Zhang, S. Wang, A. Tan, Blind quality index for tone-mapped images based on luminance partition, Pattern Recognit. 89 (2019) 108–118.
[31] I. Abouelaziz, M.E. Hassouni, H. Cherifi, A curvature based method for blind mesh visual quality assessment using a general regression neural network, in: Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on, IEEE, 2016, pp. 793–797.
[32] A. Nouri, C. Charrier, O. Lézoray, 3D blind mesh quality assessment index, Electron. Imaging 2017 (20) (2017) 9–26.
[33] I. Abouelaziz, M.E. Hassouni, H. Cherifi, No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression, in: International Conference on Image and Signal Processing, Springer, 2016, pp. 369–377.
[34] I. Abouelaziz, M.E. Hassouni, H. Cherifi, Blind 3d mesh visual quality assessment using support vector regression, Multimed. Tools Appl. 77 (18) (2018) 24365–24386.
[35] I. Abouelaziz, M.E. Hassouni, H. Cherifi, A convolutional neural network framework for blind mesh visual quality assessment, in: IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 755–759. 2017
[36] I. Abouelaziz, A. Chetouani, M.E. Hassouni, H. Cherifi, A blind mesh visual quality assessment method based on convolutional neural network, Electronic Imaging 2018 (18) (2018). 423–1

[37] I. Abouelaziz, A. Chetouani, M.E. Hassouni, L.J. Latecki, H. Cherifi, Convolutional neural network for blind mesh visual quality assessment using 3d visual saliency, in: 2018 25th IEEE International Conference on Image Processing (ICIP), IEEE, 2018, pp. 3533–3537.
[38] C.H. Lee, A. Varshney, D.W. Jacobs, Mesh saliency, ACM Trans. Graph. (TOG ACM) 24 (2005) 659–666.
[39] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach.Intell. 20 (11) (1998) 1254–1259.
[40] S. Bai, X. Bai, Z. Zhou, Z. Zhang, Q. Tian, L.J. Latecki, Gift: towards scalable 3d shape retrieval, IEEE Trans. Multimed. 19 (6) (2017) 1257–1271.
[41] Z. Zhu, Cong Rao, Song Bai, L. Longin Jan, Training convolutional neural network from multi-domain contour images for 3D shape retrieval, Pattern Recognition Letters, 2017.
[42] A. Krizhevsky, One weird trick for parallelizing convolutional neural networks, 2014. arXiv preprint arXiv: 1404.5997.
[43] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. arXiv preprint arXiv: 1409.1556.
[44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[45] A. Fukui, D.H. Park, D. Yang, A. Rohrbach, T. Darrell, M. Rohrbach, Multimodal compact bilinear pooling for visual question answering and visual grounding, 2016. arXiv preprint arXiv: 1606.01847.
[46] F.M. Algashaam, K. Nguyen, M. Alkanhal, V. Chandran, W. Boles, J. Banks, Multispectral periocular classification with multimodal compact multi-linear pooling, IEEE Access 5 (2017) 14572–14578.
[47] L. Kang, P. Ye, Y. Li, D. Doermann, Convolutional neural networks for no-reference image quality assessment, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1733–1740.
[48] G. Lavoué, M.C. Larabi, L. Váša, On the efficiency of image metrics for evaluating the visual quality of 3d models, IEEE Trans. Visualization Comput.Graph. 22 (8) (2016) 1987–1999.
[49] S. Silva, B.S. Santos, C. Ferreira, J. Madeira, A perceptual data repository for polygonal meshes, in: Visualisation, 2009. VIZ'09. Second International Conference in, IEEE, 2009, pp. 207–212.
[50] Z. Wang, A.C. Bovik, Modern image quality assessment, synthesis lectures on image, Video Multimed. Process. 2 (1) (2006) 1–156.
[51] K. Wang, F. Torkhani, A. Montanvert, A fast roughness-based approach to the assessment of 3d mesh visual quality, Comput. Graph. 36 (7) (2012) 808–818.
[52] P.G. Engeldrum, Psychometric scaling: avoiding the pitfalls and hazards, PICS (2001) 101–107.

Ilyass Abouelaziz is a Ph.D. student at the Mohammed V University in Rabat - Morocco. His main research interests include 3D mesh visual quality assessment and machine learning. He started his research in December 2014 in the Computer Science & Telecommunications Laboratory Research "Laboratoire de recherche en informatique et telecommunicatins (LRIT)". He received his master's degree in computer science and telecommunications in 2014 at the same University. He obtained the Fulbright scholarship that allows him to pursue a year of study and research in the United States at Temple University, Philadelphia - Pennsylvania. He was also a visiting researcher in PRISME laboratory, University of Orleans - France. His research focuses on image and video processing, visual saliency, machine learning, and mainly on mesh visual quality assessment.

Aladine Chetouani received his master's degree in computer science from the University Pierre and Marie Curie, France, in 2005, and his Ph.D. degree in image processing from the University of Paris 13, France, in 2010. From 2010 to 2011, he was a postdoctoral researcher with the L2TI Laboratory, Paris 13 University. He is currently an associate professor with the Laboratory PRISME, Orleans, France. His present interests are in image quality, perceptual analysis, visual attention, and image processing for cultural heritage using convolutional neural networks.

Mohammed El Hassouni received the Ph.D. in image and video processing from the University of Burgundy in 2005. He joined Mohammed V University of Rabat as assistant professor since 2006 and associate professor in 2012. He also received the Habilitation from Mohammed V University in 2012. Since 2018, he is a full professor at Mohammed V Unversity in Rabat. He was a visitor of several universities (Bordeaux, Orleans, Dijon and Konstanz). He is member of IEEE, IEEE Signal Processing Society and several conferences program committee. He is also cochair of QUAMUS workshop. His research focuses on image analysis, quality assessment and mesh processing.

Longin Jan Latecki is a professor of computer science at Temple University, Philadelphia. His main research interests include shape representation and similarity, object detection and recognition in images, robot perception, machine learning, and digital geometry. He received the PhD degree in 1992 and the habilitation in 1996, both from University of Hamburg, Germany. He has published over 225 research papers and books. He is an editorial board member of journals: Pattern Recognition, Computer Vision and Image Understanding, and International Journal of Mathematical Imaging. He received 2010 College of Science and Technology Research Excellence Award, the annual Pattern Recognition Society Award together with Azriel Rosenfeld for the best article published in the journal Pattern Recognition in 1998. He is also the recipient of the 2000 Olympus Prize, the main annual award, from the German Society for Pattern Recognition (DAGM).

Hocine Cherifi has been Professor of computer science at the University of Burgundy, France, since 1999. Prior to moving to Dijon, he held faculty positions at Rouen University and Jean Monnet University, in the disciplines of signal processing and mathematics. He has held visiting positions at Yonsei, Korea, University of Western Australia, Australia, National Pintung University, Taiwan, and Galatasaray University, Turkey. He has been elected to a variety of leadership positions in the French Classification Society. He has been an Associate Editor of a variety of image processing and computer vision journals. More recently, he joined the editorial board of the Computational Social Networks Journal published by Springer. He received a PhD in Signal Processing from the Grenoble Institute of Technology, France in 1984. His research focuses on the fields of computer vision and complex networks.