https://doi.org/10.2352/ISSN.2470-1173.2018.18.3DIPM-423

A blind mesh visual quality assessment method based on convolutional neural network

Ilyass Abouelaziz; LRIT, Associated Unit to CNRST (URAC No 29)- Faculty of sciences, Mohamed V University in Rabat, B.P.1014 RP, Rabat, Morocco.
Aladine Chetouani; University of Orleans PRISME laboratory - Orleans, France.
Mohammed El Hassouni; LRIT, Associated Unit to CNRST (URAC No 29)- Faculty of sciences, Mohamed V University in Rabat, B.P.1014 RP, Rabat, Morocco.
Hocine Cherifi; LE2I UMR 6306 CNRS, University of Burgundy, Dijon, France

Abstract

This paper presents a new method for no reference mesh visual quality assessment using a convolutional neural network. To do this, we first render 2D images from multiple views of the 3D mesh. Then, each image is split into small patches which are learned to a convolutional neural network. The network consists of two convolutional layers with two max-pooling layers. Then, a multilayer perceptron (MLP) with two fully connected layers is integrated to summarize the learned representation into an output node. With this network structure, feature learning and regression are used to predict the quality score of a given distorted mesh without the availability of the reference mesh. Experiments have been successfully conducted on LIRIS/EPFL general-purpose database. The obtained results show that the proposed method provides good correlation and competitive scores comparing to some influential and effective full and reduced reference methods.

Introduction

The interest on perceptual visual quality assessment (MVQ) of 3D meshes has become an important issue in the past few years due to the fact that 3D models have been widely used in several applications. 3D meshes are usually subject to different lossy geometry processing operations, which inevitably introduce variable distortions on the 3D shape of the object and may alter the visual quality of the model. Indeed, the performances of a computer vision application are often sensitive to the quality of the data. Thus, it is crucial to evaluate how much the original 3D model has been affected by a specific operation. In fact, visual distortions can be measured by human observers (subjective visual quality assessment), but this evaluation is too expensive, laborious and time-consuming compared to the objective visual quality assessment [1] that seems to be a good solution to overcome these problems.

To estimate the visual quality, objective visual quality assessment methods are required, they can be classified according to the availability of the reference object: Full-Reference (FR) when the reference is fully available, No-Reference (NR) or Blind when no information about the reference is available, and Reduced-Reference (RR) when only a part of the reference is available i.e features extracted from the reference. Several perceptually driven method have been proposed, however, they are limited to full reference and reduced reference methods, that is to say, they require the presence of the reference in order to evaluate the perceived quality. 3D mesh quality assessment methods can be used in several computers graphics applications such as benchmarking 3D mesh processing algorithms, optimizing and assessing performances of compression, 3D TV and so forth. The subject of this work is to propose a no-reference mesh visual quality assessment method to objectively estimate the perceived visual quality of a given distorted mesh.

In the literature, several mesh visual quality assessment methods have shown good performances and good correlation with mean opinion scores provided by subjects. However, the available methods are limited in full reference [2, 3, 4, 5, 6] and reduced reference methods [7, 8]. In practical situations, the reference is not always available, and the proposition of no reference methods become essential. To overcome this problem, our purpose is to investigate the use of machine learning technics; specifically deep convolutional networks in order to implement a no-reference mesh visual quality assessment method and ensure a good correlation with human judgments. This is a very challenging issue since the reference is not considered, and the quality estimation is based on a learning problem.

The remainder of this paper is organized as follows: a brief description of the state of the MVQ metrics is presented in Section. II. Experimental setup including the used database and the validation protocol is provided in Section. III. Experimental comparisons and discussion are given in Section. IV. Finally, we draw in Section. V some concluding remarks and perspectives.

The proposed method
Overview

Compared to the existing mesh visual quality assessment methods, the proposed method is no-reference, it needs only the distorted mesh to predict the quality score without having the reference mesh. In addition, it integrates the use of a deep learning network in the field of mesh visual quality assessment using 2D views rendered from the 3D object. The flowchart of the proposed no-reference mesh visual quality assessment method is depicted in Fig. 1. Given a distorted mesh, the first step is to render 2D images from multiple views of the 3D mesh. In the second step, each image is split into small patches of size 32×32 in order to fit the requirement of the convolutional neural network which is used for the training and the quality prediction.

2D views rendering and patch preparation

The first step of the proposed method is to render 2D projections of the 3D object from multiple views. Projections are

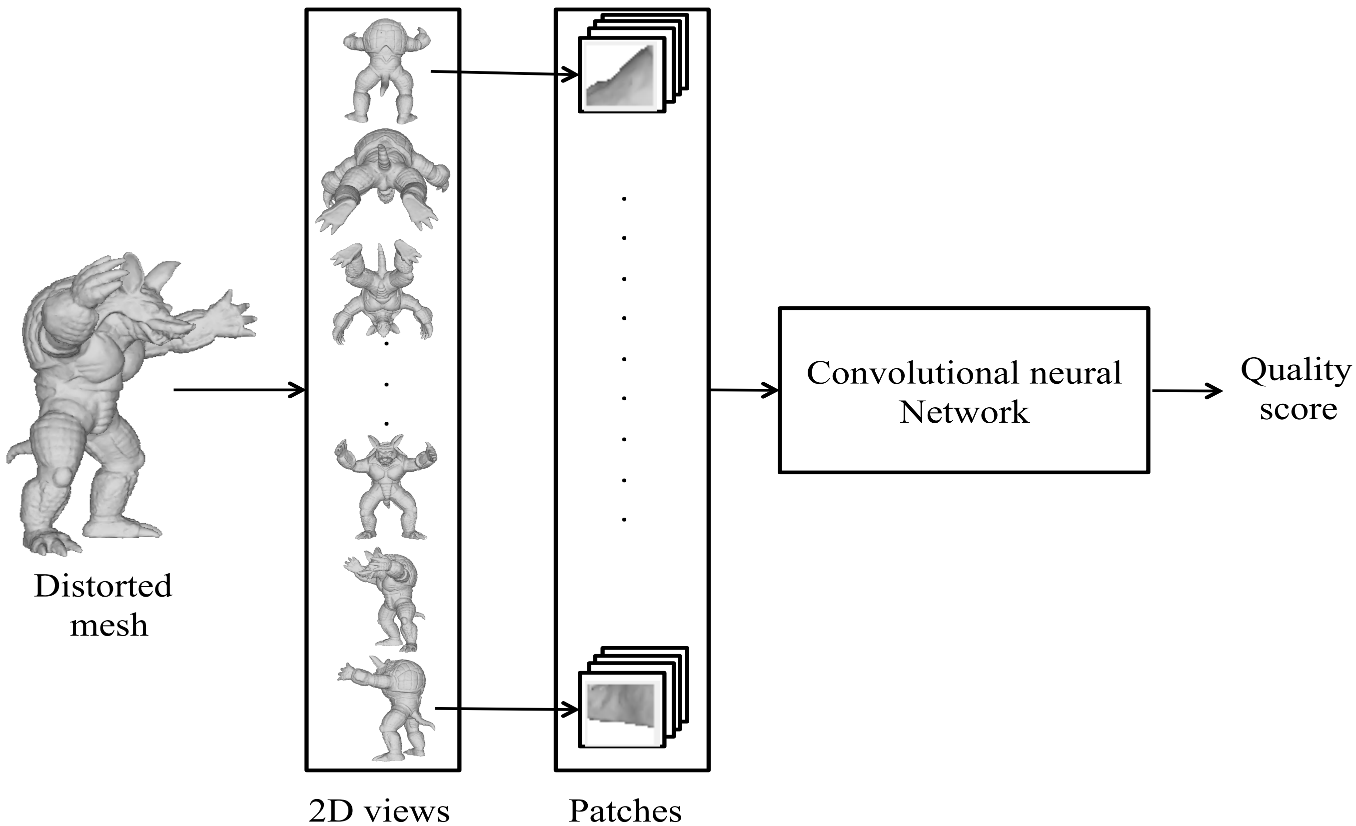

Figure 1. The overall scheme of the proposed no reference mesh quality assessment method.

obtained by fixing a virtual camera and rotate the 3D mesh by an angle of 60 degrees according to the X and Y axes. In total, 11 views are obtained for each distorted mesh. Fig. 2 shows an example of Venus model and its corresponding views. Once the 2D views are rendered, the next step is to split each view into small patches of size 32×32 to feed the convolutional neural network which the different layers are described in the next section.

CNN architecture

The architecture of the proposed CNN is composed of seven layers as depicted in Fig. 3. The different layers are presented as follows:

- **Input:** 2-D patches of size 32×32.
- **Convolutional layer 1:** The first layer is a convolutional layer which filters the input patch with 32 kernels. Each kernel is of size (5×5). This layer provides 32 feature maps of size 28×28.
- **Max-pooling layer 1:** The second layer is a max-pooling layer which applies the max-pooling process on each feature map with a local window of size 2×2. This layer produces 32 feature maps with a lower dimension of 14×12.
- **Convolutional layer 2:** The third layer is another convolutional layer which filters the output of the max-pooling layer with 32 kernels of size 5×5. This layer produces 32 feature maps of size 10×10.
- **Max-pooling layer 2:** The fourth layer is another max-pooling layer with a local window of size 10×10. as a result, this layer produces a feature vector of size 1×32.
- **Fully connected layers:** The fifth and sixth layers are two fully connected layers of 250 nodes each.
- **Output layer:** The last layer is a simple linear regression with a one-dimensional output that provides the quality score.

Training

The proposed method consists of two phases: the training phase and the test phase. For the training process, we conduct a leave-one-out cross validation (LOOCV): The training model is built using the patches from all the existing objects in the repository except one object and its degraded version. As the used database is composed of 4 objects and their corresponding degraded version, we thus dexompose the database as follows : 75% for the training and the rest for the test. The proposed CNN architecture is trained on non-overlapping patches of size 32×32. Thanks to the homogeneous distortions in the training meshes, we assign for each input patch a score, which is correspond to the mean opinion score of the source mesh. The parameters of the convolutional neural network are learned using stochastic gradient descent (SGD) and back propagation. We adopt the training objective function defined as follows:

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

$$\hat{\omega} = \min_{\omega} L$$

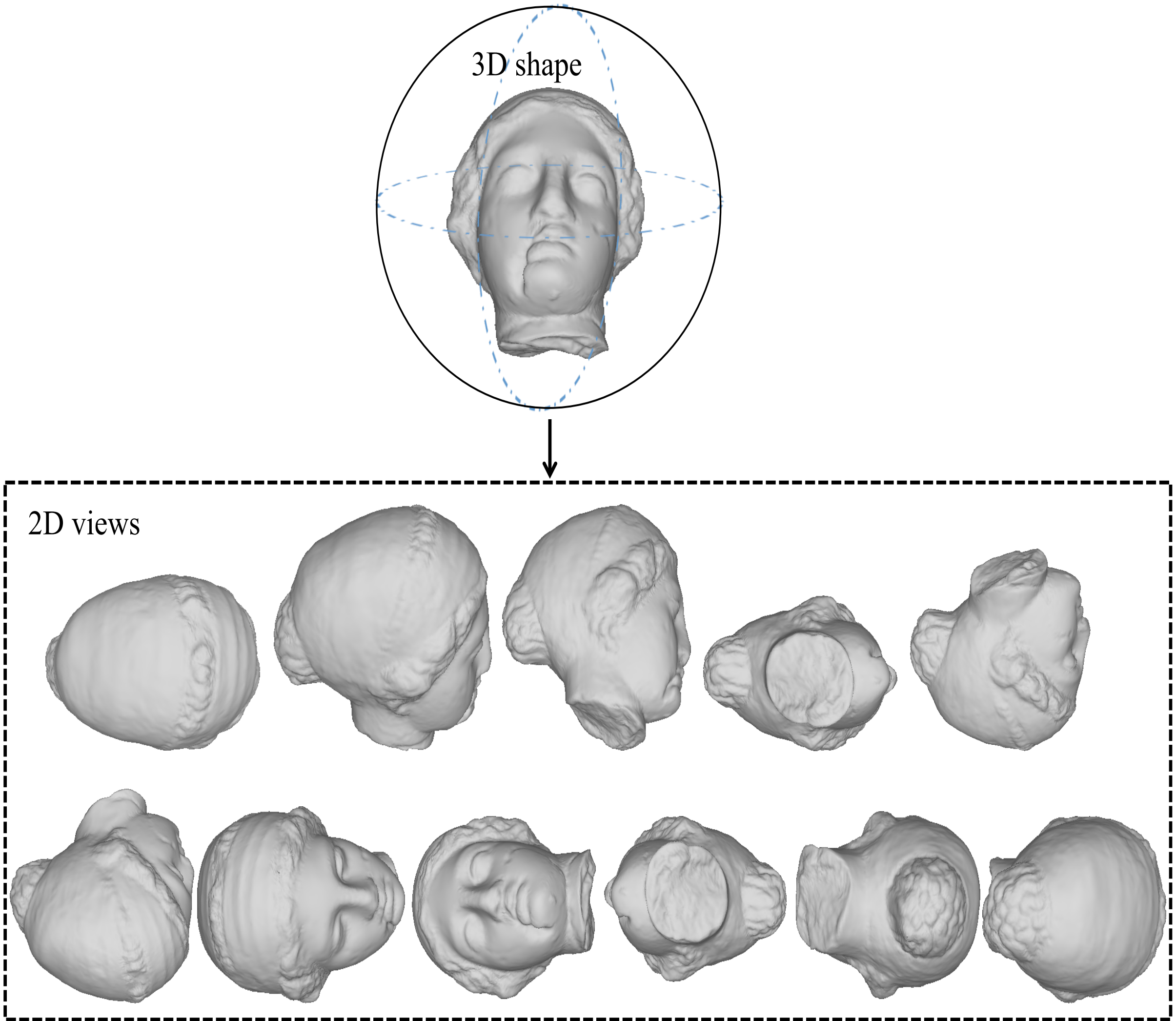

Figure 2. *Example of 2D views rendering from Venus 3D shape.*

Where MOS_n is the mean opinion score assigned to a given input patch p_n and $S(p_n;\omega)$ is the predicted score of p_n with network weights ω. In our experiments, we perform the training for 20 epochs.

During the test phase, the patches of the excluded object are used (test set), and the trained network estimates a quality score for each patch of the mesh. Finally, the overall quality score of the test mesh is obtained by averaging all the predicted patch scores. This process is done until all the meshes in the repository are tested and we obtain a quality score for each distorted mesh. The obtained scores are then compared with the subjective scores using correlation measurements.

Experimental results

Database and test criteria

The performance of our proposed method is evaluated on the general-purpose database. This database was created at the EPFL, Switzerland. It contains 4 reference meshes, Armadillo, Dyno, Venus and RockerArm, from which 84 distorted models are generated (88 models total). Two types of distortion are applied, smoothing and noise addition, either locally or globally on the reference mesh. The subjective evaluation was done by 12 observers. The given scores are between 0 (good quality) and 10 (bad quality), and for each model, a normalized mean opinion score (MOS) is computed by averaging all the scores given by the 12 observers. Fig. 4 shows the 4 reference meshes of the general-purpose database.

The correlation between the estimated quality scores and the mean opinion scores is used as criteria to evaluate the performance of a given quality assessment method. Two types of correlation are usually used:

The Pearson linear correlation coefficient (r_p) which employed to measure the prediction accuracy.

$$r_p = \frac{Cov(PMOS,MOS)}{Var(PMOS)Var(MOS)},$$

where MOS and PMOS are two vectors of the subjective and objective scores respectively.

The Spearman rank-order correlation coefficient (r_s) which employed to measure the prediction monotonicity [10], it is com-

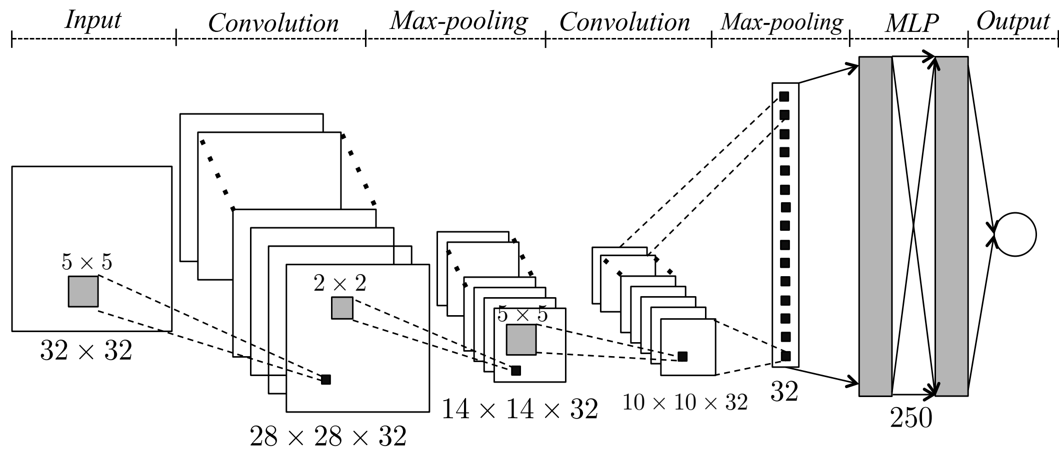

Figure 3. *The architecture of the proposed convolutional neural network.*

Figure 4. *Reference Meshes from the general-purpose database.*

puted suing the same equation but replacing the raw scores by their ranks. The Spearman rank order correlation coefficient depends only on the rank of the objective scores of the models; if the rank of objective scores is similar as the rank of MOSs, a high value will be obtained, regardless of the distance between the objective score and the corresponding MOS. For both criteria, a higher value indicates a better prediction performance.

Comparison and discussion

Contrary to the 2D and stereoscopic image quality metrics, there is a lack of no-refrence methods in mesh visual quality assessment. In the experiment results, we only compared our method to several full reference, reduced reference methods and only one recent no-reference method [9]. Table . 1 presents the (r_p) and (r_s) values of our method from the general-purpose database as well as the values from the existing methods. Note that the correlations on the whole corpus are computed between the objective scores of all objects in the corpus and their corresponding MOSs.

the proposed method provide high correlation scores that contend the most effective methods on the general-purpose database. Particularly, the highest r_s and r_p for armadillo and Venus model. In addition, the proposed method has a competitive correlation concerning Dyno and Rocker model as well as on the whole repository with a correlation of ($r_s = 81.75\%$) and ($r_p = 82.54\%$) . Our method outperforms the BMQI method, which is a no-reference metric for all meshes except the Rocker. However, the global performances are higher than the latter. Comparing to the other databases, the general purpose database contains an important number of distorted models (21 distorted version for each model as well as a variety of distortion types). The high correlation scores provided by the proposed method in this database appear to be a good indicator for the forcefulness of our method in mesh visual quality assessment.
According to the above evaluation and comparison, we can see that the proposed method is quite robust and effective in term of predicting the objective quality score of a distorted mesh, as reflected by the high and competitive scores comparing to existing well-known methods. In addition, we recall that our method is no-reference, and do not require the reference meshes for the quality estimation. Accordingly, it can be appropriate for practical situations and real-time applications.

Conclusion

In this paper, we presented a convolutional neural network based method for blind mesh visual quality assessment. The network is trained using small patches extracted from images of multiple views of the 3D mesh. The proposed architecture is composed of multiple layers of convolution and max-pooling. In addition, the MLP layer with two fully connected layers is integrated to summarize the representation and produce the final quality score. The comparison with existing full reference and reduced reference mesh visual quality assessment methods shows that the proposed no-reference method provides good correlations with human judgment. In addition, it is noteworthy that the proposed method is blind and does not require the reference mesh. Future work will concern other representative features and network configuration.

Correlation coefficients r_s (%) and r_p (%) of different objective metrics on 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[2]	69.5	30.2	30.9	22.6	1.6	0.8	18.1	5.5	13.8	1.3
	RMS [3]	62.7	32.3	0.3	0.0	90.1	77.3	7.3	3.0	26.8	7.9
	MSDM2 [4]	81.6	85.3	85.9.4	85.7	89.3	87.5	89.6	87.2	80.4	81.4
	DAME [5]	60.3	76.3	92.8	88.9	91.0	83.9	85.0	80.1	76.6	75.2
	TPDM [6]	84.5	78.8	92.2	89.0	90.6	91.0	92.2	91.4	89.6	89.2
Reduced Reference	3DWPM1 [7]	65.8	35.7	62.7	35.7	71.6	46.6	87.5	53.2	69.3	38.4
	3DWPM2 [7]	74.1	43.1	52.4	19.9	34.8	16.4	37.8	29.9	49.0	24.6
	FMPD [8]	75.4	83.2	89.6	88.9	87.5	83.9	88.8	84.7	81.9	83.5
No Reference	BMQI [9]	20.1	–	83.5	–	88.9	–	92.7	–	78.1	–
	Our method	93.44	95.62	86.22	84.34	94.09	90.35	80.45	82.16	81.75	82.54

References

[1] Corsini, M., Larabi, M. C., Lavoué, G., Petrik, O., Vasa, L., & Wang, K. (2013, February). Perceptual metrics for static and dynamic triangle meshes. In Computer Graphics Forum (Vol. 32, No. 1, pp. 101-125). Blackwell Publishing Ltd.

[2] Aspert, N., Santa Cruz, D., & Ebrahimi, T. (2002, August). MESH: measuring errors between surfaces using the Hausdorff distance. In ICME (1) (pp. 705-708).

[3] Cignoni, P., Rocchini, C., & Scopigno, R. (1998, June). Metro: measuring error on simplified surfaces. In Computer Graphics Forum (Vol. 17, No. 2, pp. 167-174). Blackwell Publishers.

[4] Lavoué, G. (2011, August). A multiscale metric for 3D mesh visual quality assessment. In Computer Graphics Forum (Vol. 30, No. 5, pp. 1427-1437). Blackwell Publishing Ltd.

[5] Vasa, L., & Rus, J. (2012, August). Dihedral angle mesh error: a fast perception correlated distortion measure for fixed connectivity triangle meshes. In Computer Graphics Forum (Vol. 31, No. 5, pp. 1715-1724). Blackwell Publishing Ltd.

[6] Torkhani, F., Wang, K., & Chassery, J. M. (2014). A curvature-tensor-based perceptual quality metric for 3D triangular meshes. Machine Graphics and Vision, 23(1-2), 59-82.

[7] Corsini, M., Gelasca, E. D., Ebrahimi, T., & Barni, M. (2007). Watermarked 3-D mesh quality assessment. IEEE Transactions on Multimedia, 9(2), 247-256.

[8] Wang, K., Torkhani, F., & Montanvert, A. (2012). A fast roughness-based approach to the assessment of 3D mesh visual quality. Computers & Graphics, 36(7), 808-818.

[9] Nouri,A., Charrier, C., & Lezoray, 0. (2017). 3D Blind Mesh Quality Assessment Index. IS&T International Symposium on Electronic Imaging - 3D Image Processing, Measurement (3DIPM), and Applications, 9-14.

[10] Z. Wang and A. C. Bovik, ”Modern image quality assessment,” Synthesis Lectures on Image, Video, and Multimedia Processing, vol. 2, no. 1, pp. 1156, 2006.

Author Biography

Ilyass Abouelaziz *is a Ph.D. student at the Mohammed V University in Rabat - Morocco. His research is about 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. His research focuses on image and video processing, visual saliency, machine learning, and mainly on mesh visual quality assessment.*

Aladine Chetouani *received his Masters degree in computer science, from the University Pierre & Marie CURIE, France in 2005. In 2010, he obtained his PhD degree in image processing from the University of Paris 13, France. From 2010 to 2011, he worked as a postdoctoral researcher at L2TI Laboratory of the University of Paris 13. Currently, he works as a teacher and researcher at Laboratory PRISME. His present interests are in image quality, perceptual analysis, visual attention and image processing for cultural heritage.*

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 since 2012. He also received the Habilitation from Mohammed V University in 2012. He was a visitor of several universities (Bordeaux, Orlans, 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.*

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*