https://doi.org/10.2352/ISSN.2470-1173.2018.12.IQSP-298

A Neural-Based Stereoscopic Image Quality Assessment with Reference

Aladine Chetouani, PRISME Laboratory, University of Orleans, Orleans, France
aladine.chetouani@univ-orleans.fr

Abstract

Image quality is an important aspect for several applications (biometrics, tracking, object detection and so on). Several methods have been proposed in the literature to estimate it. These methods are able to predict subjective judgments according to different characteristics. The goal of this paper is to present a framework for stereoscopic image quality metric with reference based on Neural Networks (CNN & ANN). The proposed CNN model is composed of 3 convolutional layers and two Fully Connected (FC) layers and it is used to identify the degradation type in the image. The quality is then estimated using an ANN model. Its inputs are some computed features, selected according to the identified degradation type. The results obtained through two common datasets show the relevance of the proposed approach.

Introduction

Image quality estimation is an important process in several computer vision applications. Indeed, the performances of those applications are often affected by the quality of the data (even in the case where the best method is used). To quantify the quality, several metrics have been proposed in the literature with different approaches: Full Reference (FR) where the pristine image is supposed accessible [1], Reduced Reference (RR) where only some characteristics of the pristine image are available [2] and No Reference (NR) where only the degraded image is used [3].

Most of the existing metrics in the literature are FR and are often used to estimate the quality of all kinds of degraded image. However, there is no true universal metric that can be used in all cases and that can provide similar performance whatever the degradation type. In this study, we propose to focus on the non-universality of those metrics for image stereoscopic images. The goal is here to improve the global estimation quality process by integrating degradation identification and features fusion steps. The former step is based on a CNN model, while in the latter step an ANN model is used. Two well-known datasets have been used to evaluate the performance of our method in terms of degradation identification and prediction of subjective judgments.

Our paper is organized as follows: We first present the method by describing both steps (degradation identification and subjective quality prediction). Then, the experimental results are shown for the considered datasets and compared to the state-of-the-art. We finish by a brief conclusion.

Proposed Method

Fig. 1 illustrates the flowchart of the proposed method based on two main steps:

- **Degradation identification:** The first step aims to identify the degradation type through a Convolutional Neural Network (CNN) model. The proposed CNN model is composed of 3 convolutional layers and two Fully Connected (FC) layers.

- **Features fusion:** The second step permits to predict the subjective quality score by combining some degradation-based features from the Cyclopean Image of the pristine and the degraded stereoscopic images. The combination step has been here realized using an Artificial Neural Network (ANN) model. The inputs of our ANN model are some features, selected according to the previous step (identified degradation type).

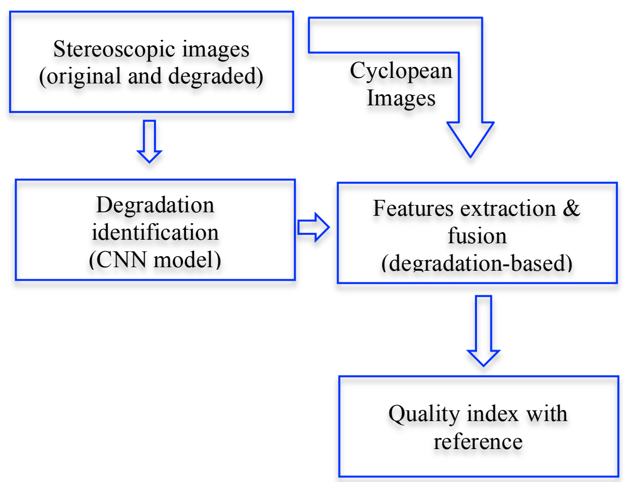


Figure 1. Flowchart of the proposed method

So, the degradation identification is given by both the right and the left views, while the features extraction is applied on the Cyclopean Images. Both steps are described in this section

Degradation identification

The architecture of our CNN model is presented in Fig. 2 (128x128x1 → 61x61x16 → 27x27x16 → 10x10x16 → 2x2x16 → 200 → 4). Its input is a 128x128x1-normalized patch. The first four layers are convolutional layers followed by a pooling step without overlap (stride=2 & padding=0). Each convolutional layer is composed of 16 kernels with a size 7x7 (i.e. 16 feature maps). There are one fully connected and one output layers after the last pooling operation. The former is composed of 200 neurons and the latter is a logistic regression layer with five outputs corresponding to the considered degradation types. The activation function used here is a ***tanh*** and the Stochastic Gradient Descent (SGD) has been used as optimization method. Our model has been developed and tested using Torch [4]. The target (output) corresponds to the detected degradation type of the corresponding input (patch). So, for a given stereoscopic image, we first identify the degradation type of its patches and we then keep the maximum occurrence as the degradation type of the entire stereoscopic image.

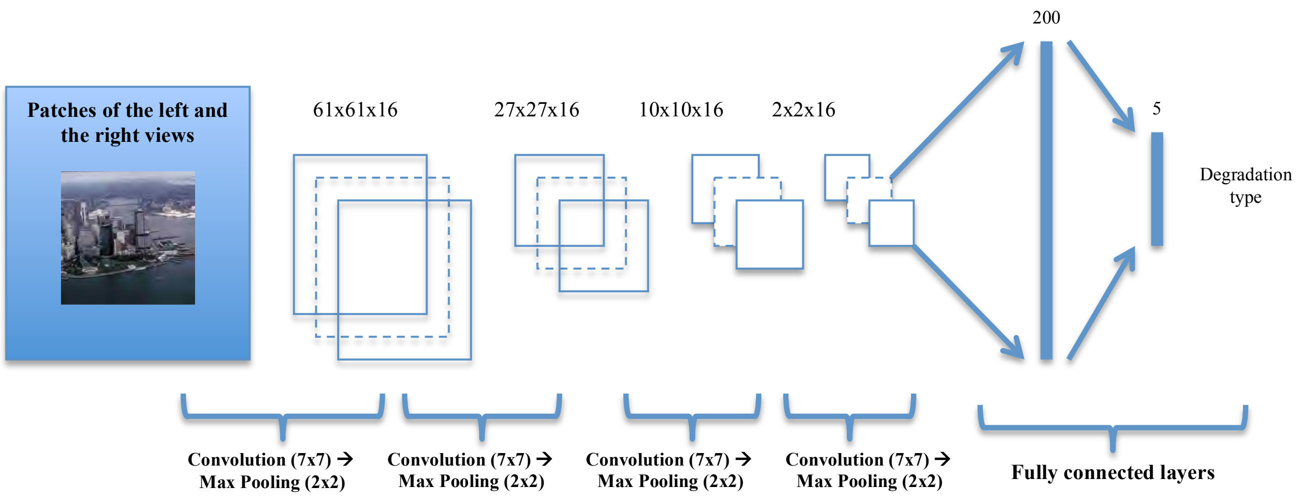


Figure 2: Proposed CNN model used to identity the degradation type

Features Combination

Cyclopean image

Once the degradation is identified, we then propose to predict the quality of the stereoscopic image by extraction some features. These features are extracted here from the Cyclopean Image (CI), which permits to consider the binocular rivalry phenomenon that occurs when two sufficiently different images are presented to the right and the left eyes. The method proposed in [5] has been used in this study and the CI is computed by weighting the left and the right views.

Selected Features

Different experimental tests have been done to select the more adapted features. Table I lists the selected metrics for each of the considered degradation type (the used datasets are described in the next section). During our tests, 15 metrics have been used (SSIM and MSSIM [6], VIFP and VIF [7], UQI [8], WSNR [9], VSNR [10], IFC [11], NQM [12], PSNR, SNR, MSE, FSIM [25], GMSD [26], PSNR_HVS_A [27]) and the best combination, that provides the best results, has been considered. We can note that some of the metrics selected for FF degradation have been also selected for BLUR and JP2K degradations (FSIM & IFC). This can be justified by the fact that BLUR is the common degradation type at certain byte rate.

Table I: Selected features for each degradation type

Degradation type	Selected metric
WN	SSIM, WSNR, NQM
BLUR	UQI, **FSIM**, GMSD
JPEG	UQI, SNR
JP2K	**IFC**, SNR, **FSIM**
FF	WSNR, **IFC**, PSNR_HVS_A

Combination Step

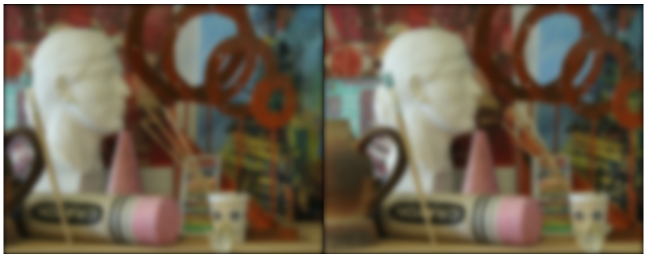
As combination tool, an ANN (Artificial Neural Network) model has been used to predict the quality score. We have one ANN model for each degradation type and the inputs of each of them are the selected features. The number of neurons in the hidden layer is different from a model to another.

Experimental Results

Used Datasets

In this study, two well-known databases have been used and are briefly described below:

- **3D LIVE image database – Phase I (3DLIVE-P1) [13]:** 5 degradation types have been considered (JPEG2000, JPEG, White Noise, Gaussian Blur and Fast Fading) derived from 20 reference images (365 degraded images). For each degraded image, the DMOS (Differential Mean Opinion Score), the disparity and depth maps are provided. Only symmetric distribution has been considered (see Fig. 3.a).
- **3D LIVE image database – Phase II (3DLIVE-P2) [14]:** in this dataset, the same five degradation types have been considered. From 8 pristine images, 360 degraded images with co-registered DMOS values are available. Symmetric and asymmetric distributions have been considered (see Fig. 3.b). The corresponding disparity and depth maps are also given.



a)

b)

Figure 3: Stereoscopic images a) with symmetric distribution and b) asymmetric distribution

Obtained Results

The obtained results have been compared to the state-of-the-art in order to better show its efficiency and its relevance in terms of degradation classification and quality estimation. In this section, we evaluate our method for both steps.

Degradation Identification

In order to evaluate our degradation identification step, we first split the 3DLIVE-P1 dataset into training (80%) and test (20%) sets randomly 20 times without overlap. Then, a cross-dataset evaluation has been done in order to generalize the result of our method. The mean good classification rate has been used to evaluate the performance of this step. Table II shows the obtained results for the 3DLIVE-P1 dataset. Our method obtained high performance (94.74%) with a small standard deviation.

Table II: Mean good classification rate obtained for the 3DLIVE-P1 dataset

Mean percentage of good classification rate	**94.74%**
Standard deviation	**2.3915**

We then test the generalization of our method by using the 3DLIVE-P2 dataset. The entire 3DLIVE-P1 dataset has been used to train the CNN model. Table III shows the obtained results for the entire database and we compute also the performance for symmetric distribution. As we can see, we obtain 92.50% as good classification rate when the whole dataset is used, while the performance are higher when only symmetric distribution is considered. This is only due to the fact that the training dataset (3DLIVE-P1) is composed only of symmetric distribution.

Table III: Mean good classification rate obtained for the 3DLIVE-P2 dataset using the 3DLIVE-P1 for the training.

Distribution type	Percentage of good classification
ALL (Asymmetric/Symmetric)	92.50%
Symmetric	97.50%

Quality Prediction

Table IV: PCC correlations for the 3DLIVE-P1 dataset

LCC					
	JP2K	**JPEG**	**WN**	**BLUR**	**FF**
***Benoit* [15]**	0.939	0.641	0.925	0.949	0.747
***You* [16]**	0.878	0.488	0.941	0.920	0.730
***Gorley* [17]**	0.485	0.312	0.796	0.852	0.369
Hewage [18]	0.904	0.531	0.896	0.798	0.670
Wang [19]	0.916	0.570	0.913	0.957	0.783
Ma [20]	0.918	0.722	0.913	0.925	0.807
Akhter [21]	0.905	0.729	0.904	0.617	0.503
Shao [22]	0.872	**0.897**	0.916	0.923	------
Chen [23]	0.907	0.695	0.917	0.917	0.735
StereoCNN [24]	0.926	0.740	0.944	0.930	0.845
Our method	**0.969**	0.796	**0.972**	**0.972**	**0.890**

Once the degradation is well identified, we then evaluate the quality prediction step using the well-known correlation coefficients (Pearson and Spearman). For that, we first split the 3DLIVE-P1 dataset (training 80% and test 20%) randomly 5 times without overlap. Tables IV and V show the obtained results for each considered degradation type. The results are also compared to some recent methods. We show in gray-background the two best metrics and in bold the best metric.

Table V: SROCC correlations for the 3DLIVE-P1 dataset

SROCC					
	JP2K	**JPEG**	**WN**	**BLUR**	**FF**
***Benoit* [15]**	0.910	0.602	0.929	0.931	0.699
***You* [16]**	0.860	0.439	0.939	0.882	0.583
***Gorley* [17]**	0.420	0.015	0.741	0.750	0.366
Hewage [18]	0.856	0.500	0.896	0.690	0.548
Wang [19]	0.883	0.542	0.907	0.925	0.655
Ma [20]	0.887	0.616	0.912	0.879	0.696
Akhter [21]	0.866	0.675	0.914	0.555	0.640
Shao [22]	0.900	0.607	0.903	0.926	------
Chen [23]	0.863	0.617	0.919	0.878	0.652
StereoCNN [24]	**0.931**	**0.693**	**0.946**	0.909	**0.834**
Our method	0.920	0.680	0.942	**0.936**	0.811

As we can see, our method outperforms the compared metric for most of the considered degradation types (in terms of PCC), except for the JPEG degradation. The best improvement has been obtained for JP2K and WN degradations.

We then evaluate the generalization capacity of our method by using the 3DLIVE-P1 dataset to train our model and 3DLIVE-P2 dataset to test it. Tables VI and VII present the obtained performances whatever the distribution type (Symmetric and Asymmetric). As we can see, we obtained high correlations for all degradation types.

Table VI: PCC correlations for the 3DLIVE-P2 dataset by using the 3DLIVE-P1 dataset for the training

	PCC				
	WN	**BLUR**	**JPEG**	**JP2K**	**FF**
Our method	0.980	0.989	0.821	0.906	0.927

Table VII: SROCC correlations for the 3DLIVE-P2 dataset by using the 3DLIVE-P1 dataset for the training

	SROCC				
	WN	**BLUR**	**JPEG**	**JP2K**	**FF**
Our method	0.968	0.946	0.764	0.911	0.936

Tables VIII-XI show the obtained correlation coefficients for asymmetric and symmetric distributions. As we can see, our method is always among the two best metrics for asymmetric distribution (Tables VIII and IX). The best improvement has been obtained for JP2K degradation type. For symmetric distribution (Tables X and XI), we also obtained competitive results for all degradation types, especially for BLUR degradation. We generally obtained better performances for symmetric distribution, except for BLUR degradation. For this latter degradation type, StereoCNN's method achieved also best result for Asymmetric distribution.

Table VIII: PCC correlations for the 3DLIVE-P2 dataset by using the 3DLIVE-P1 dataset for the training (Asymmetric distribution)

	PCC				
	WN	**BLUR**	**JPEG**	**JP2K**	**FF**
SSIM [6]	0.823	0.840	0.685	0.676	---
FSIM [25]	0:941	0.888	0.796	0.785	---
Chen [22]	0.945	0.692	0.564	0.722	---
Shao [23]	0.924	0.855	0.705	0.789	---
StereoCNN [24]	0.796	0.924	0.583	0.782	---
Our method	**0.966**	**0.989**	**0.801**	**0.910**	0.936

Table IX: SROCC correlations for the 3DLIVE-P2 dataset by using the 3DLIVE-P1 dataset for the training (Asymmetric distribution)

	SROCC				
	WN	**BLUR**	**JPEG**	**JP2K**	**FF**
SSIM [6]	0.8821	0.807	0.714	0.724	---
FSIM [25]	**0.9521**	0.850	**0.805**	0.806	---
Chen [22]	0.9292	0.691	0.636	0.722	---
Shao [23]	0.9235	0.803	0.696	0.789	---
StereoCNN [24]	0.7797	0.865	0.581	0.793	---
Our method	0.949	**0.967**	0.771	**0.905**	0.917

Table X: PCC correlations for the 3DLIVE-P2 dataset by using the 3DLIVE-P1 dataset for the training (Symmetric distribution)

	PCC				
	WN	**BLUR**	**JPEG**	**JP2K**	**FF**
SSIM [6]	**0.975**	0.833	0.677	0.816	---
FSIM [25]	0.963	0.864	0.846	0.818	---
Chen [22]	0.946	0.918	0.601	0.670	---
Shao [23]	0.917	0.977	0.873	0.903	---
StereoCNN [24]	0.957	0.899	**0.927**	**0.921**	---
Our method	0.973	**0.920**	0.844	0.894	0.943

Table XI: SROCC correlations for the 3DLIVE-P2 dataset by using the 3DLIVE-P1 dataset for the training (Symmetric distribution)

	SROCC				
	WN	**BLUR**	**JPEG**	**JP2K**	**FF**
SSIM [6]	0.945	0.770	0.718	0.726	---
FSIM [25]	0.937	0.850	0.841	0.824	---
Chen [22]	0.907	0.845	0.630	0.662	---
Shao [23]	0.937	**0.911**	0.910	**0.904**	---
StereoCNN [24]	0.941	0.490	**0.942**	0.897	---
Our method	**0.970**	0.870	0.642	0.838	0.897

Conclusion

In this paper, a stereoscopic image quality metric has been proposed. The integration of the degradation type using a CNN model and the features fusion are the main contributions of this paper. It is worth noting that the degradation type is often not explicitly used to improve the quality estimation process. Both steps have been evaluated and the achieved results show its relevance. As perspective, we will try to combine other metrics and better exploit the CNN model.

References

[1] M. Carnec, P. Le Callet and D Barba, "Objective quality assessment of color images based on a generic perceptual reduced reference", Signal Processing: Image Communication, Volume 23, Issue 4, pp. 239-256, 2008.

[2] A. Chetouani, A. Beghdadi, M. A. Deriche and A. Bouzerdoum, "A Reduced Reference Image Quality Metric based on feature fusion and neural networks", EUSIPCO, pp. 589-593, 2011

[3] A. Chetouani, A. Beghdadi, S. Chen and G. Mostafaoui, "A free reference image quality measure using neural networks", International Workshop on Video Processing and Quality Metrics, 2010.

[4] http://torch.ch/

[5] M.J. Chen, C.C. Su, D.K. Kwon, L.K. Cormack and A.C. Bovik, "Full-reference quality assessment of stereopairs accounting for rivalry," Signal Processing: Image Communication, Vol. 28, pp. 1143 – 1155, 2013.

[6] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, Vol. 13, issue 4, pp. 600-612, 2004

[7] H.R. Sheikh, A.C. Bovik and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics,"

IEEE Transactions on Image Processing, Vol.14, no.12, pp. 2117-2128, 2005

[8] Z. Wang, A. Bovik, “A universal image quality index,” IEEE Signal Processing Letters, Vol. 9, pp. 81-84, 2002

[9] T. Mitsa and K. Varkur, “Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms”, International Conference on Acoustics, Speech, and Signal Processing, pp. 301- 304, 1993.

[10] D.M. Chandler, S.S. Hemami, “VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images,” IEEE Transactions on Image Processing, Vol. 16, pp. 2284-2298, 2007

[11] H.R. Sheikh.and A.C. Bovik, “Image information and visual quality,” IEEE Transactions on Image Processing, Vol.15, no.2, pp. 430-444, 2006.

[12] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans and A.C. Bovik, “Image Quality Assessment Based on a Degradation Model”, IEEE Transactions on Image Processing, Vol. 9, pp. 636-650, 2000

[13] http://live.ece.utexas.edu/research/quality/live_3dimage_phase1.html

[14] http://live.ece.utexas.edu/research/quality/live_3dimage_phase2.html

[15] Benoit A. Benoit, P. Le Callet, and P. and Campisi, “Quality assessment of stereoscopic images,” EURASIP journal on image and video processing, Vol. 2008, 2009.

[16] You J. You, L. Xing, A. Perkis, and X. Wang, “Perceptual quality assessment for stereoscopic images based on 2d image quality metrics and disparity analysis,” International Workshop on Video Processing and Quality Metrics, 2010.

[17] P. Gorley, N. Holliman, “Stereoscopic image quality metrics and compression,” SPIE 6803, Stereoscopic Displays and Applications XIX, 2008.

[18] C.T.E.R. Hewage and M.G. Martini, “Reduced-reference quality metric for 3d depth map transmission,” 3DTV-CON, pp. 1–4, 2010.

[19] Z. Wang, E. P. Simoncelli and A. C. Bovik, “Multiscale structural similarity for image quality assessment”, Asilomar Conference on Signals, Systems, and Computers, vol. 2, pp. 1398–1402, 2003.

[20] L. Ma, X. Wang, Q. Liu, and K. N. Ngan, “Reorganized DCT-based image representation for reduced reference stereoscopic image quality assessment”, Neurocomputing, 2016.

[21] R. Akhter, Z.M.P. Sazzad, Y. Horita, and J. Baltes, “No reference stereoscopic image quality assessment,” IS&T/SPIE Electronic Imaging, Vol. 7524, 2010.

[22] F. Shao, W. Lin, S. Wang, G. Jiang, and M. Yu, “Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook”, IEEE Transactions on Broadcasting, 2015.

[23] Chen M.J. Chen, L. K. Cormack and A. C. Bovik, “No-reference quality assessment of natural stereopairs”, IEEE Transactions on Image Processing, pp. 3379–3391, 2013.

[24] W. Zhang, C. Qu, L. Ma, J. Guan and R. Huang, “Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network”, Pattern Recognition, Vol. 59, pp. 176-187, 2016.

[25] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: a feature similarity index for image quality assessment", IEEE Transactions on Image Processing, vol. 20, no. 8, pp. 2378-2386, 2011.

[26] W. Xue, L. Zhang, X. Mou and A.C. Bovik,"Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index", ???

[27] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, M. Carli, “Modified Image Visual Quality Metrics for Contrast Change and Mean Shift”, Proceedings of CADSM, pp. 305 – 311, 2011.

Author Biography

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. Since 2012, 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.*