

# Local Binary Patterns and Neural Networks for No-Reference Image and Video Quality Assessment

Marko Panić, Dubravko Ćulibrk, Srdjan Sladojević, and Vladimir Crnojević

University of Novi Sad, Faculty of Technical Sciences,  
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia  
{mpanic,dculibrk,sladojevic,crnojevic}@uns.ac.rs  
<http://www.ftn.uns.ac.rs>

**Abstract.** *In the modern world where multimedia is predicted to form 86% of traffic transmitted over the telecommunication networks in the near future and providers are looking to shift towards Quality of Experience, rather than Quality of Service in multimedia delivery, no-reference image quality assessment and the related video quality assessment remaining open research problem, with significant market potential. In this paper we describe a study focused on evaluating the applicability of Local Binary Patterns (LBP) as features and neural networks as estimators for image quality assessment. We focus on blockiness artifacts, as a prominent effect in all block-based coding approaches and the dominant artifact in occurring in videos coded with state-of-the-art video codecs (MPEG-4, H.264, HVEC). In this initial study we show how an LBP-inspired approach, tuned to this particular effect can be efficiently used to predict the MOS of JPEG coded images. The proposed approach is evaluated on a well-known public database and against widely-used features. The results presented in the paper show that the approach achieves superior performance, which forms a sound basis for future research aimed at video quality assessment and precise blocking artifact detection with sub-frame precision.*

**Keywords:** Local Binary Patterns, Neural Networks, Multimedia Quality, Image Quality, Assessment

## 1 Introduction

Video traffic is predicted to form approximately 86 percent of global consumer traffic by 2016 [7]. Every second, 1.2 million minutes of video content will cross the network in 2016. Video traffic is becoming the dominant application over 3G/4G mobile systems. Fuelled by proliferation of smartphones, netbooks and tablets, mobile data traffic is estimated to grow 18 times from 2011-2016, with the surge caused mainly by mobile video services. Within this landscape an important question for service providers is the quality assurance of the service provided to the consumers of multimedia content.

Video Quality Assessment (VQA) algorithms attempt to automatically assess perceptual degradations introduced by signal processing and transmission operations performed on video sequences. Unfortunately, despite recent advances in video processing and communication technology, their performance leaves something to be desired and there is considerable room for improvement [12]. This is especially true for no-reference methods, which have no notion of the original, non-compressed multimedia content.

To estimate video quality one usually first needs to be able to derive metrics of impairments at the level of single frames. Once the quality is estimated at frame level, diverse methods exist to aggregate this information to create a quality score for the whole video. Overall degradation of the quality of frames(images) is a compound effect of different coding artifacts. Three types of artifacts are typically considered pertinent to DCT block (JPEG, MPEG and H.264) coded data: blocking, ringing and blurring. Blocking appears in all block-based compression techniques due to coarse quantization of frequency components [16]. It can be observed as surface discontinuity (edge) at block boundaries. These edges are perceived as abnormal high-frequency components in the spectrum. Ringing is observed as periodic pseudo edges around original edges [9]. It is due to improper truncation of high frequency components. This artifact is also known as the Gibbs phenomenon or Gibbs effect. In the worst case, the edges can be shifted far away from the original edge locations, observed as false edge. Blurring, which appears as edge smoothness or texture blur, is due to the loss of high frequency components when compared with the original image. Blurring causes the received image to be smoother than the original one [5].

Broadly, the different artifacts can be considered to add texture not present in original content, to the images. However, general-purpose texture descriptors are rarely, if ever, used as basis for deriving the metrics of the level of artifacts introduced into multimedia content. In the study presented here, we explore the applicability of Local Binary Patterns (LBP) to the problem of creating quality metrics. LBPs represent texture descriptors which have been successfully used in a number of computer vision applications. They have, so far as we know, not been considered for no-reference image and video quality assessment.

The goal of each no-reference approach is to create an estimator based on the proposed features that would predict the Mean Opinion Score (MOS)[8] of human observers, without using the original (not-degraded) image or sequence data. In the study presented here, we evaluated the applicability of LBPs as features on which to base no-reference MOS estimation and the Multilayer Perceptron (MLP) [6] neural network as an estimator. To evaluate both the proposed approach images from a public and commonly LIVE Image Quality Assessment Database [13] were used. The experiments conducted show that the proposed approach is able to achieve no-reference image quality assessment beyond that of a widely used state-of-the art approach.

## 2 Related Work

When objective image and video quality is concerned, most studies focus on full-reference approaches, which assume the availability of pristine, uncompressed original content at the time of assessment. This is a fairly well researched and solved problem. Evaluation of a number of such approaches on the LIVE database is presented in [14].

In a realistic multimedia content delivery over telecommunication network scenario, these algorithms can only be used at server side, or in specific cases where the pristine content has been previously delivered to client side. The problem of no-reference quality assessment is more complex and the need for better solutions still exists.

Wang *et al.* [16] proposed an early no-reference approach to quality assessment in JPEG coded images. Their final measure is derived as a non-linear combination of a blockiness, local activity and a so-called zero-crossing measure. The combination is supposed to provide information regarding both blockiness and blurring (via the two latter measures) in JPEG coded images. Their approach remains to this date the one usually compared against, when no-reference quality assessment is concerned. Therefore, we evaluated the results of the approach proposed here against that of the approach proposed in [16].

More recently Culibrk *et al.* proposed a VQA approach that used different previously proposed artifact and quality metrics as basic features and machine learning algorithms to create a no-reference video quality (MOS) estimator [3]. The study demonstrated the viability of a machine approach and the ability to achieve superior quality estimation this way. In fact, one of the algorithms evaluated was MLP. In addition, the authors provided evidence that the basic features used by Wang *et al.* are among best predictors selected from the set of 35 classic measures and saliency-related features evaluated.

LBP's were originally proposed as a texture descriptor [11]. They have since then been employed as dynamic texture descriptors [18], as well as found a number of successful applications such as face recognition [1] and object detection [15]. To the best of our knowledge it has never been applied to the problem of image and video quality assessment.

## 3 LBP-VQA approach

The LBP operator is primarily used as an unifying approach to structural and statistical texture analysis. A detailed explanation of LBP derivation and its extensions is provided in [10]. Equation (1) describes the process of obtaining LBP code for a local area in a image, defined by the location of the central pixel  $(x_c, y_c)$ , radius from it  $R$  and number of pixels that surround it  $P$  at the distance of the radius  $R$ .

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^P s(g_p - g_c) 2^p \quad (1)$$



**Fig. 1.** Referent image (left) and images compressed with bitrates 0.45313, 0.3263, and 0.15771, respectively (right).

If the location of surrounding pixels  $(x_p, y_p)$ , where  $p$  denotes the index of neighboring pixel does not match the pixel grid exactly, the intensity of that pixel is obtained through bilinear interpolation. The differences in the values between the central pixel and those surrounding it, as defined by a local area (mask) with parameters  $R$  and  $P$ , are mapped to zero or one using the function  $s(g_p - g_c)$ . Where  $g_c$  and  $g_p$  denote gray value of central pixel and surrounding pixel respectively:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (2)$$

The LBP value is obtained as a sum of ones weighted by  $2^p$ , where  $p$  corresponds to the index of the pixel in the mask. Thus, the LBP value is a decimal representation of the binary number extracted from the predefined mask. The start of the pixel counting sequence within the mask is chosen arbitrarily the first time, and then must remain unchanged.

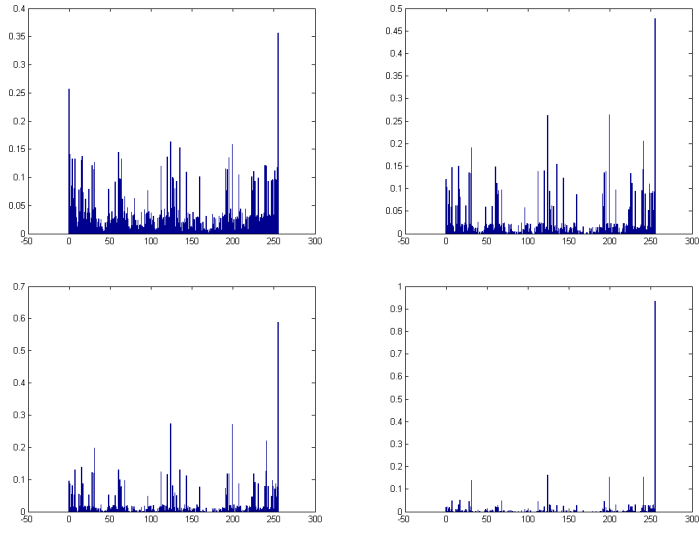
Finally, the feature vector is a histogram of values of LBP codes calculated for the whole image. Figure 1 shows referent and images compressed with different bitrates. Figure 2 shows the LBP codes for the same images. As artifacts introduced by JPEG compression become more visible, the distribution of LBP codes changes in such a way that only a few codes remain, representing blockiness and uniform regions. Therefore, histograms of LBP codes seem to be a good representation of image quality, as one can easily distinguish between those pertaining to compressed and non-compressed images (Figure 3). It should be noted that all histograms were normalized using  $\sqrt{\|h\|_1}$  norm, where  $h$  represents the histogram.

## 4 Experiments and Results

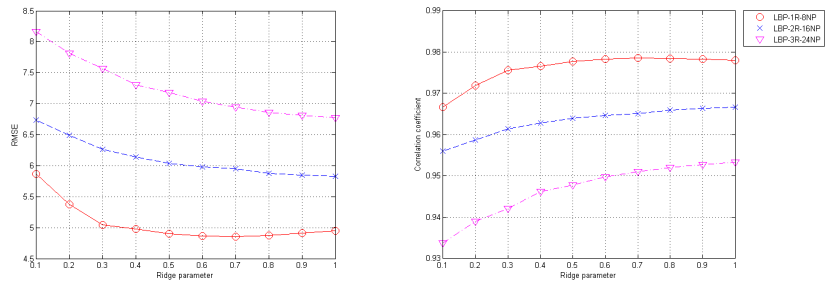
The evaluation was performed using images from the LIVE Image Quality Assessment Database [13]. The images were compressed using JPEG and form



**Fig. 2.** LBP codes for images in Figure 1.



**Fig. 3.** Histograms of LBP codes for images in Figure 1.



**Fig. 4.** Correlation coefficients and RMSE values for different MLP ridge parameter values and LBP codes generated for diverse neighborhoods:  $R = 1$ ,  $N = 8$  - circles,  $R = 2$ ,  $N = 16$  - crosses, and  $R = 3$ ,  $N = 24$  - triangles.

the release 1 of the database. The database was created from twenty-nine high-resolution 24-bits/pixel RGB color images (typically 768 x 512) using different compression ratios. This yielded 204 images. For all 233 images, subjective MOS is available.

Feature extraction was implemented in C++ using OpenCV library [2]. We evaluated the MLP neural network as a MOS predictor. The network was trained and its performance tested using the machine learning package Weka [17]. The neural network had one hidden layer and was trained by minimizing the squared error plus a quadratic penalty with the BroydenFletcherGoldfarbShanno (BFGS) method [4].

To provide for better evaluation we considered different values for the key parameters. The estimator proved to be most sensitive to the value of the ridge parameter, which controls the overfitting. Figure 4 shows how the performance of model changes for different ridge parameter values. We considered three sizes for the neighborhoods used to extract the LBP codes. The results shown were obtained using 10-fold cross-validation and the network that had just 2 neurons in the hidden layer. Due to the complexity of the BFGS model, 2 neurons in the hidden layer are the default in Weka.

As Figure 4 shows, good correlation and low RMSE can be achieved using the proposed approach. For the ridge parameter value of 0.7, correlation of 0.9767 and a RMSE of 0.4890 is achieved between the predicted and subjective MOS. When tested on the train set, slightly better result are obtained, 0.9867 for the correlation coefficient and 0.3740 for RMSE, indicating that the model generalizes well and is stable.

Figure 5 shows the scatter plots of predictions of the proposed approach and the well-known model proposed Wang and Bovik [16]. The scatters show the comparison of model predictions, using both 10-fold cross-validation and when tested on the train set. The evaluation on the training set was included as the final model of Wang and Bovik was fitted using all the images in the dataset, making the evaluation of their model, as conducted here, equivalent to testing on the training set.

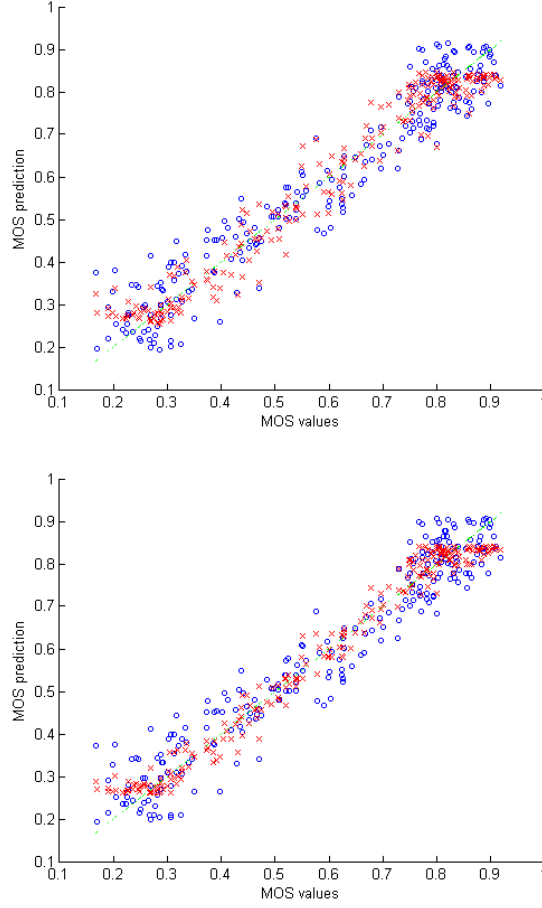
As the scatters show, the performance of the proposed approach is significantly better than that reported by Wang *et al.* in [16], as their model achieved an average RMSE of 0.7256. In addition, proposed model correlates with MOS very well.

## 5 Conclusion

The paper proposes an approach to image and video quality assessment based on LBPs as features and an MLP neural network estimator.

The proposed approach was evaluated on a standard database of JPEG images and achieved superior performance when compared to a classic no-reference image quality assessment methodology.

The study demonstrated the usefulness of LBPs and neural networks for quality assessment. Something that has not been explored before. In the future the



**Fig. 5.** MOS predictions for the proposed approach (crosses) and Wang and Bovik model (circles), when cross-validated (top) and evaluated on the training set (bottom)

study will be extended to evaluate the proposed approach on H.264 videos and to create a blockiness artifact detector which will be able to detect isolated artifacts that appear due to network-introduced errors in multimedia transmission.

## Acknowledgment

This research is supported in part by the FP7 project QoSSTREAM.

## References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 28(12), 2037–2041 (2006)
2. Bradski, G.: The OpenCV Library. *Dr. Dobbs's Journal of Software Tools* (2000)
3. Culibrk, D., Mirkovic, M., Zlokolica, V., Pokric, M., Crnojevic, V., Kukolj, D.: Salient Motion Features for Video Quality Assessment. *IEEE Trans. on Image Processing* 20, 948–958 (2010)
4. Dennis, J.E., Schnabel, R.B.: Numerical methods for unconstrained optimization and nonlinear equations, vol. 16. Society for Industrial and Applied Mathematics (1987)
5. Ferzli, R., Karam, L.: A no-reference objective image sharpness metric based on just-noticeable blur and probability summation. *Proceedings of IEEE 2007 International Conference on Image Processing* 3, III –445–III –448 (16 2007-Oct 19 2007)
6. Haykin, S.: *Neural Networks: A Comprehensive Foundation*. Macmillan, New York (1994)
7. Index, C.: Forecast and methodology, 2011-2016. White paper, CISCO, December (2012)
8. ITU-R BT.500: Methodology for the Subjective Assessment of the Quality of Television Pictures. Video Quality Experts Group (2002)
9. Kirenko, I.: Reduction of coding artifacts using chrominance and luminance spatial analysis. *Consumer Electronics, 2006. ICCE '06. 2006 Digest of Technical Papers. International Conference on* pp. 209–210 (Jan 2006)
10. Mäenpää, T.: *The Local Binary Pattern Approach to Texture Analysis: Extensions and Applications*. Oulun yliopisto (2003)
11. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24(7), 971–987 (2002)
12. Seshadrinathan, K., Bovik, A.: An information theoretic video quality metric based on motion models. In: *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics*. pp. 25–26. Citeseer (2007)
13. Seshadrinathan, K., Soundararajan, R., Bovik, A., Cormack, L.: Study of subjective and objective quality assessment of video. *Image Processing, IEEE Transactions on* 19(6), 1427–1441 (2010)
14. Seshadrinathan, K., Soundararajan, R., Bovik, A., Cormack, L.: A subjective study to evaluate video quality assessment algorithms. In: *SPIE Proceedings Human Vision and Electronic Imaging*. vol. 7527. Citeseer (2010)

15. Wang, X., Han, T., Yan, S.: An hog-lbp human detector with partial occlusion handling. In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 32–39. IEEE (2009)
16. Wang, Z., Sheikh, H.R., Bovik, A.C.: No-reference perceptual quality assessment of jpeg compressed images. In: Proceedings of IEEE 2002 International Conferencing on Image Processing. pp. 477–480 (2002)
17. Witten, I.H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann, San Francisco (2005)
18. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. Pattern Analysis and Machine Intelligence, IEEE Transactions on 29(6), 915–928 (2007)