Quality Assessment of JPEG-distorted Face Images: Influence of Affective Content

Silvia Corchs, Gianluigi Ciocca and Francesca Gasparini

Department of Informatics, Systems and Comunication, University of Milano Bicocca, Milano 20126, Italy

Abstract: In this work we investigate if the affective content of images influences the perception of image quality. Two database are generated and psychophysical experiments are conducted, where participants rate the stimuli in a five point Likert scale. We have fixed the semantic content, choosing only close-ups of face images, two emotion categories (happy and sad images) and JPEG-distortion. Also the influence of the background is considered. From the analysis of the subjective data we observe that the influence of affective content is more evident for images of very high or very low quality. The subjective scores are further used as ground-truth labels to train a five quality-class classifier. Two different feature spaces are used (visual features and quality metrics) to train a SVM classifier.

1 INTRODUCTION

Objective image quality assessment is mainly related to measuring the presence of distortions. Humans, while scoring the quality of images, are not always able to disregard all the factors related not only to the distortion presence but also to other aspects like complexity (Ciocca et al., 2017), image semantic (Siahaan et al., 2018; Triantaphillidou et al., 2007) and affective image content (Van Der Linde and Doe, 2012). When these different aspects concur to generate final subjective rates, the objective metrics, that measure only distortions, may not properly predict human judgments.

In particular, Image Quality Assessment (IQA) studies do not in general take into account how does the affective image content influence the quality perception process itself. Most available IQA databases are built starting from a set of pristine images which are subsequently corrupted by introducing graded simulated distortions (Sheikh et al., 2006; Larson and Chandler, 2010; Corchs et al., 2014). In general, such pristine images belong to different image content classes such as indoor, outdoor, landscape, close-up, etc., without considering the affective content dimension. On the other hand, recently a different type of IQA database has been presented that contains several authentic distortions on a very large number of images (Ghadiyaram and Bovik, 2016).

Nowadays Internet of Things and wearable computing is an active field of research and several smart devices are able to capture emotional reactions to audio visual stimuli. Within this field, knowing if the quality of the signal influences its affective perception can give insights in human computer interactions.

Up to our knowledge, only van der Linde and Doe (Van Der Linde and Doe, 2012) addressed the affective dimension issue, analyzing the influence of affective image content on subjective quality assessment. The authors set up a psycho-physical experiment and found that participants were unable to disentangle affective image content from objective image quality in a standard IQA procedure.

Since the semantic content of an image can influence in different ways users perception of quality (Siahaan et al., 2018), we here propose to fix the semantic content, focusing on images depicting close up of faces. Consequently, with respect to the affective content, we consider two emotion categories: happy versus sad images. Such choice is supported by the experimental results of van der Linde and Doe (Van Der Linde and Doe, 2012) who observed that the pleasantness of the viewed image is a factor that influences subjective rating of image quality. Moreover, users should be more affected by the emotions elicited by the face expressions. With respect to image distortions, as initial investigation, here we focus on JPEG compression artifacts.

Besides the analysis of the affective dimension on IQA, in this work we will also investigate the role of

the background when judging image quality. We set up a psychophysical IQA experiment as a single categorical stimulus with five anchors (excellent, very good, good, fair and poor quality). We also explore different classification methods to predict these five quality classes. Therefore, a classification approach to the IQA problem is proposed and investigated instead of the traditional regression analysis.

In Section 2 we present the two database generated in our Multi Media Signal Processing (MMSP) laboratory to study the influence of both background and affective content on IQA. In section 3 the psychophysical experiments are described and the collected data is analyzed (statistics, etc). In Section 4 the two different sets of objective features (visual features and No Reference quality metrics) are listed and the classification results are presented in Section 5.

2 THE MMSP-FACE DATABASE

The MMSP-Face database consists of 46 high quality pristine images depicting close-ups of happy (23) and sad (23) face images. In Figure 1 the corresponding thumbnails are shown (all images are under Creative Commons license). Starting from this database we have generated:

The **MMSP-FaceA** database consisting of 230 images. Each of the pristine images of the MMSP-Face database were JPEG distorted using the MAT-LAB imwrite function with the following q-factors: 10, 15, 20, 30.

The **MMSP-FaceB** database. Starting from the MMSP-FaceA database we have cut out the background and left only the face in each of the images. The images are then distorted as in MMSP-FaceA. This database also contains 230 images.

In Figure 2 an example image in its most distorted version is shown for MMSP-FaceA and MMSP-FaceB database. The datasets will be available online at www.mmsp.unimib.it

3 EXPERIMENTS

Two different experiments were conducted. The first one on the MMSP-FaceA database, and the second one on the MMSP-FaceB database. Observers were asked to judge the quality of each image of the database in a single stimulus presentation through a web-interface. They were explicitly instructed not to judge the emotional nor the affective content of the image. Before the start of the experiment, a grayscale chart was shown to allow the observers to
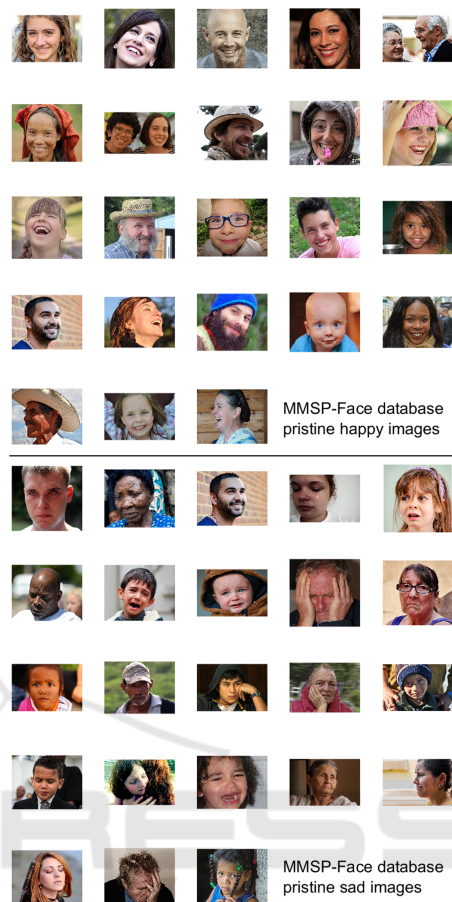

Figure 1: Thumbnails of pristine images belonging to MMSP-Face database.

Figure 2: An example image from MMSP-FaceA and MMSP-FaceB.

calibrate the brightness and the contrast of the monitor. After the calibration, six Ishihara tables were presented to the observers to estimate color vision deficiency. If the participants did not report correctly any of the six they were subsequently discarded from the subjects' pool. The stimuli were shown in random order, different for each subject. Participants rated the image quality in five discrete scales, ranging from one star to five stars, corresponding to: terrible (one star), not good (two stars), average (three stars), very good (four stars) and excellent quality (five stars). A web-interface was used in the experiment, where stimuli were presented for an unlimited time, up to response submission. In order to get the

observers accustomed to the experiment, seven practice trials were presented at the beginning of each experiment. The corresponding data were discarded and not considered for any further analysis. To avoid fatigue effects, the stimuli were divided in two subsets of 115 images each. Therefore, two sessions of experiments were performed for each of the considered database. Participants were recruited from the Informatics Department of the University of Milano Bicocca and were either students or researchers. None of the participants failed the Ishihara test. All the experiments reported in this article were conducted in accordance with the Declaration of Helsinki and the local guidelines of the University of Milano Bicocca (Italy). In the first experiment (MMSP-FaceA database), 23 observers participated in the first session, and 17 in the second session for a total number of ratings collected of $115 \times 23 + 115 \times 17 = 4,600$. In the second experiment (MMSP-FaceB database), 21 observers participated in the first session, and 19 in the second session for a total number of ratings collected of $115 \times 21 + 115 \times 19 = 4,600$.

For each of the 230 images in each database, we collected the set of discrete values (ratings) assigned by the observers. In this section we analyze the normalized frequencies (with respect to the total number of observers) of the subjective quality class (from one to five stars) for each objective quality setting, obtained for MMSP-FaceA and MMSP-FaceB database respectively. Let us first consider the influence of the background on the subjective perception of the quality classes. We compare the experimental results from MMSP-FaceA and MMSP-FaceB dataset in Figure 3. The normalized frequencies are shown for each of the quality settings. The objective quality settings are indicated as follows: 5 (original image), 4 (image JPEG-compressed with q-factor=30), 3 (q-factor=20), 2 (q-factor=15) and 1 (q-factor=10). We observe that the most distorted images are labeled with one star more frequently when no background is present (75% versus 70%). A similar behavior of the data can be observed for quality settings 2, 3 and 4. In fact, even if the differences decrease, for these quality settings , the corresponding images are more frequently rated as two, three and four stars in absence of background: 53% vs 46%, 38% vs 35% and 36% vs 32% respectively. However, with respect to the original pristine images, these were more frequently labeled with five stars when the background was present (63% versus 56%). We now investigate how do the two emotion categories (happy-sad) influence the quality assessment at different levels of JPEG compression. In Figure 4 we plot again the normalized frequency as a function of the objective quality settings but now

considering separately the subset of happy and sad images respectively. In the presence of background (first column of Figure 4), the most distorted images from the sad subset show a higher probability to be labeled as one star (72%) with respect to the happy images (67%). This difference becomes smaller when the background is not present (second column of the Figure), where the corresponding probabilities are 76 % (sad) and 74 % (happy) respectively. On the other side, we observe that the probability of pristine images to be classified with five stars is greater in the case of happy images compared to sad ones, in both databases. In the case of MMSP-FaceA data (first column of Figure 4) we have: 66 % versus 59%, while for the MMSP-FaceB data (second column of the figure), we have: 60% against 51%. For quality settings 2 and 4, the probability of happy images to be labeled as two and four stars respectively are slightly higher than the corresponding sad ones, for both databases. In what follows, the final subjective quality class assigned to a given image is the class with highest frequency among all the participants. Using the statistical mode we discard the influence of outliers. Besides aggregating the subjective data using the statistical mode, we also evaluate the Mean Opinion scores (MOS) as traditionally done in the field of IQA. In Figure 5 we compare the MOS obtained using the two different databases. A high correlation is found between the subjective quality perception in both experiments; the linear Pearson correlation coefficient is equal to 0.97.

4 OBJECTIVE FEATURES

One of our goal is to predict the five quality classes (from one to five stars) using features computed from the images themselves (i.e. objective features). In this section we list and briefly describe the set of considered features. These features will be used in the next section to train and test a quality classifier. We have considered two groups of features: features obtained from No Reference (NR) quality metrics (subsection 4.1), and visual features (subsection 4.2).

4.1 No Reference Quality Metrics

We consider fourteen NR qualtiy metrics frequently used in the literature: seven are specific to measure JPEG distortions (metrics M1 to M7) and the rest are general purpose ones (metrics M8 to M14). A brief description of them follows:

M1 (Wu and Yuen, 1997): It is the most well known metric in the spatial domain. It measures the
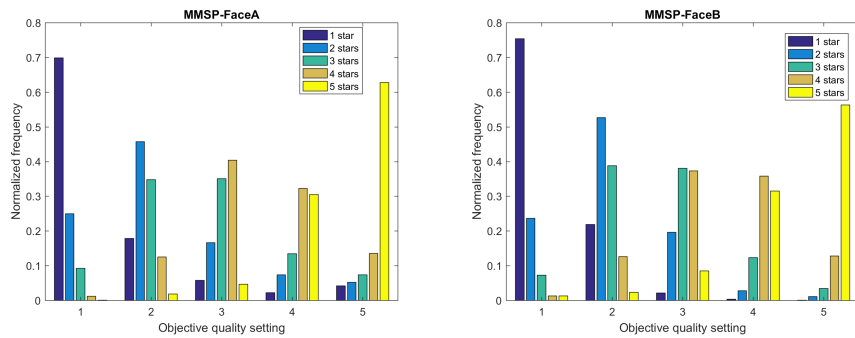
Figure 3: Normalized frequencies (with respect to the total number of observers) are plot for each of the objective quality settings on MMSP-FaceA and MMSP-FaceB dataset respectively. The objective quality settings are: 5 (original image), 4 (image JPEG-compressed with q-factor=30), 3 (q-factor=20), 2 (q-factor=15) and 1 (q-factor=10).
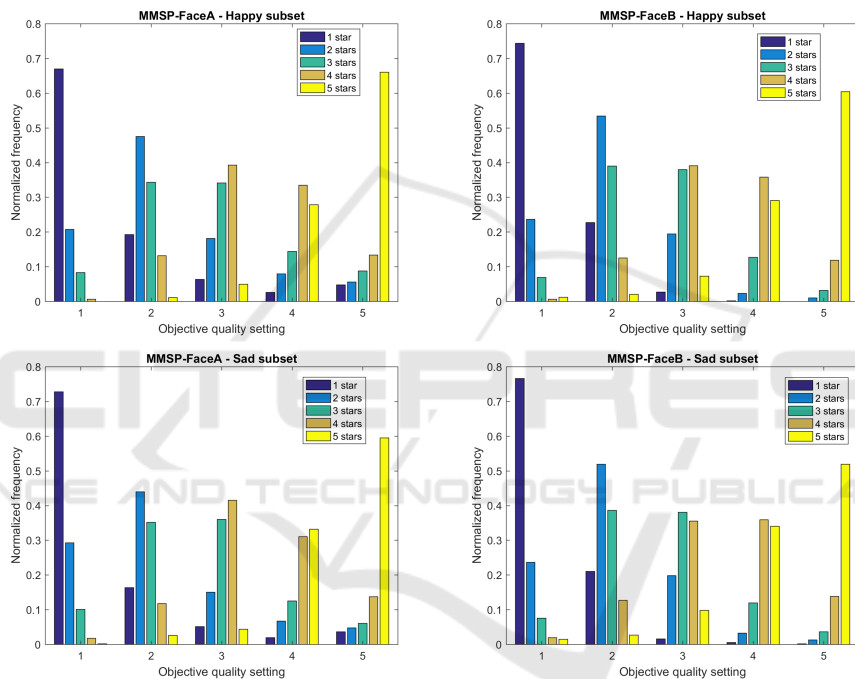

Figure 4: Normalized frequencies (with respect to the total number of observers) for each of the objective quality settings in the subset of happy and sad images for MMSP-FaceA and MMSP-FaceB dataset respectively. The objective quality settings are: 5 (original image), 4 (image JPEG-compressed with q-factor=30), 3 (q-factor=20), 2 (q-factor=15) and 1 (q-factor=10).
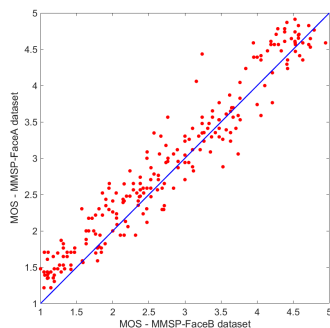

Figure 5: Correlation between the MOS obtained for MMSP-FaceA and MMSP-faceB database.

blockiness separately in horizontal and vertical direction, after which the two directions are combined into a single quality value.

M2 (Babu et al., 2004): it integrates edge amplitude, edge length, background activity and background luminance to evaluate block edge impairment.

M3 (Wang et al., 2000): It is formulated in the frequency domain. It models the blocky image as a non-blocky image interfered with a pure blocky signal. It detects and evaluates the peaks in the power spectrum of the blocky signal.

M4 (Wang et al., 2002): The method works in the frequency domain and is based on gradient features. It considers blurring and blocking as the most signifi-

cant artifacts generated during the JPEG compression process.

M5 (Pan et al., 2004): It is based on gradient features and it examines the blocks individually, measuring the severity of blocking artifacts locally. The local metric is averaged over all possible blocks to yield a unique score.

M6 (Chen and Bloom, 2010): The absolute difference between horizontally adjacent pixels is computed, normalized, and averaged along each column. A one-dimensional discrete Fourier transform is thereafter employed and vertical and horizontal blockiness measures are derived.

M7 (Muijs and Kirenko, 2005): The key algorithm is based on the principle that block discontinuities can be characterized as edges that stand out from the spatial activity in their vicinity. The visibility of a block edge is determined by the contrast between the local gradient and the average gradient of the adjacent pixels.

M8 (Moorthy and Bovik, 2010): the Blind Image Quality Index is a general purpose metric based on natural scene statistics that does not require any knowledge of the distorting process. BIQI has been evaluated on the LIVE database and in particular on its subset of JPEG-distorted images.

M9 (Mittal et al., 2012): the Blind/Referenceless Image Spatial QUality Evaluator is a general purpose metric based on natural scene statistic. It uses scene statistics of locally normalized luminance coefficients to quantify possible losses of naturalness in the image due to the presence of distortions.

M10 (Saad et al., 2010): the BLind Image Integrity Notator using DCT-Statistics (BLIINDS) is a general-purpose metric that uses natural scene statistics models of discrete cosine transform coefficients to perform distortion-agnostic NR IQA.

M11 (Mittal et al., 2013): the Natural Image Quality Evaluator is based on the construction of a quality aware collection of statistical features based on a space domain natural scene statistic model without training on human-rated distorted images.

M12 (Ghadiyaram and Bovik, 2015): it is a deep belief network that takes model-based statistical image features derived from a very large database of authentically distorted images.

M13 (Liu et al., 2016): Blind IQA by relative gradient statistics and adaboosting neural network.

M14 (Gu et al., 2015): it uses the free energy principle for blind IQA together with classical human visual system-inspired features such as structural information and gradient magnitude.

4.2 Visual Features

A total of 21 visual features are considered. We evaluate the following group of features usually adopted for texture analysis and image classification: Coarseness, Contrast, Directionality, Linelikeness, Roughness, (Tamura et al., 1978), Edge density (Mack and Oliva, 2004), Local Binary Pattern (LBP) (Ojala et al., 1996), and Histogram of Oriented Gradients (HOG), developed by Ludwig et al. (Junior et al., 2009). All these features are 1-Dimensional (1-D), except LBP (2891-D) and HOG (1296-D). We also consider a group of features related to simple chromatic properties: Chroma Variance (1-D) (Ciocca et al., 2016), Number of Regions (1-D) (Comaniciu and Meer, 2002), Colorfullness (1-D) (Hasler and Suesstrunk, 2003), Color Histogram in the HSV color space (32-D), color simple statistics (mean and standard deviation) in the RGB color space, (6-D), Autocorrelogram obtained quantizing the RGB color space in 64 colors, (64-D). Finally, we take into account features related to photographic properties, and visual perception: Feature Congestion and Subband Entropy (Rosenholtz et al., 2007), image complexity (Corchs et al., 2016), entropy, edge contrast and measure of enhancement (Schettini et al., 2010), a measure of the degree of focus (Minhas et al., 2009), all of them 1-D and a 113-D aesthetic feature vector (Bhattacharya et al., 2013).

5 CLASSIFICATION RESULTS

As pointed out in the introduction, we here propose to classify images within five quality classes instead of using regression models as traditionally done in IQA literature. A similar framework was previously proposed in (Corchs et al., 2014). The features within each group, either quality metrics or visual, are concatenated together forming a single feature vector of size 14 and 4412 respectively (many of the visual features are multidimensional as indicated in section 4.2). The ground truth labels are the subjective quality classes (from one to five stars) obtained from the experimental session conducted on MMSP-FaceA and MMSP-FaceB database. We recall that for each of the 230 images a subjective quality class was assigned using the mode statistic. We report in Table 1 the classification results obtained using Support Vector Machine (SVM) classifiers for MMSP-FaceA database. The best performance (in terms of accuracy) was achieved using a linear kernel in case of visual features and a quadratic kernel in case of quality metrics. A five-cross validation scheme was ap-

plied. We observe that using the quality NR metrics as feature space of the SVM classifier outperforms the corresponding classification using visual features. The confusion matrices are shown in Tables 2 and 3, where the number of images of each quality class are reported. Analyzing in detail the confusion matrix from Table 2 we can note that quality class five is the best predicted (only 3 images were misclassified). Quality class four is not predicted at all, and we observe that is the class less populated (only 19 images were labeled with 4 stars). Images of quality class three are mainly misclassified in quality two. Also for quality two, high rates of misclassifications are observed. With respect to quality class one, 70% of correct predictions are found. Similar results are observed from the confusion matrix in Table 3 but the number of misclassifications is decreased in particular for qualities two and three. We repeat the classification task but considering separately the subset of happy and sad images. The corresponding accuracies are shown in Table 1. As in the case of the whole database, the same conclusion is achieved: the performace of classification is higher when considering the quality metrics' space. For each feature space we have that: for visual features, both subsets show a decrease of classification performance with respect to the whole database; this decrease is more evident for the sad subset of images. With respect to the quality metrics' space, the classification performance is slightly increased for both happy and sad subsets with respect to the whole dataset.

Table 1: Performance in terms of accuracy (%) of SVM classifiers considering MMSP-FaceA.

Features	MMSP-FaceA	HAPPY	SAD
Visual	59	57	47
Quality	72	73	73

Table 2: Confusion matrix obtained using SVM classifier with visual features on MMSP-FaceA.

Class	Predicted				
Real	1⋆	2⋆	3⋆	4⋆	5⋆
1⋆	42	16	1	0	0
2⋆	14	36	12	0	0
3⋆	3	26	21	0	1
4⋆	0	3	9	0	7
5⋆	0	0	1	2	36

We performed the same experiments on the MMSP-FaceB database. Classification accuracies are shown in Table 4 and the corresponding confusion matrices for the whole dataset are reported in Tables 5 and 6. Even if in general similar classification performances are obtained, the distribution of true posi-

Table 3: Confusion matrix obtained using SVM classifier with quality metrics features on MMSP-FaceA.

Class	Predicted				
Real	1⋆	2⋆	3⋆	4⋆	5⋆
1⋆	46	13	0	0	0
2⋆	3	51	8	0	0
3⋆	0	19	31	0	1
4⋆	0	1	11	1	6
5⋆	0	0	1	0	38

tives and misclassifications has changed. In particular, from Table 5, comparing the present results with those of MMSP-FaceA data (Table 2), class 2 seems now to be better predicted while class 5 confirms to be the one best predicted. Neither in this case class 4 is predicted. From the comparison of Tables 3 and 6, the distribution of true and false positivies does not show important differences. Finally, with respect to the comparison of the whole dataset MMSP-FaceB and the happy and sad subsets (Table 4) we note that, considering the quality metrics' feature space, the sad (happy) images present a higher (lower) classification performance with respect to the whole dataset.

Table 4: Performance in terms of accuracy (%) of SVM classifiers considering MMSP-FaceB.

Features	MMSP-FaceB	HAPPY	SAD
Visual	57	54	50
Quality	72	69	75

Table 5: Confusion matrix obtained using SVM classifier with visual features on MMSP-FaceB.

Class	Predicted				
Real	1⋆	2⋆	3⋆	4⋆	5⋆
1⋆	29	15	2	0	0
2⋆	7	50	13	0	0
3⋆	1	20	20	5	0
4⋆	0	7	16	1	8
5⋆	0	0	2	3	31

Table 6: Confusion matrix obtained using SVM classifier with quality metrics features on MMSP-FaceB.

Class	Predicted				
Real	1⋆	2⋆	3⋆	4⋆	5⋆
1⋆	39	7	0	0	0
2⋆	3	56	11	0	0
3⋆	0	10	32	3	1
4⋆	0	2	17	3	10
5⋆	0	0	1	0	35

As a final experiment, let us apply our quality classification proposal on the database from (Van Der Linde and Doe, 2012), composed of JPEG-compressed images that also include different affective contents. The authors considered 100 of the 730

Table 7: Performance in terms of accuracy (%) of SVM classifiers, considering the dataset of Van Der Linde.

Features	(Van Der Linde and Doe, 2012)
Visual	33
Quality metrics	35

affective color images provided in the Geneva Affective PicturE Database (GAPED) (Dan-Glauser and Scherer, 2011). Images "negative" (spiders, snakes, human concerns), "positive" and "neutral" were chosen. For each of the original images, four JPEG compressed versions were generated (q-factors 10, 15, 20 and 30). The final database is composed of 500 images. Each of the 500 stimulus images was rated by 5 participants. Since the authors set up the experiment using a 10 points discrete scale, we had to collapse their 10 points scale to our five points one. Therefore, we have reordered their results as follows: (1-2) was assigned to label 1, (3-4) to label 2, (5-6) to label 3, (7-8) to label 4 and (9-10) to label 5. The final label of each image is obtained applying the statistical mode. In Table 7 we show the classification results. The low accuracies observed in this case can be due to several factors. While our database contains only images belonging to a unique semantic content (face), the database by (Van Der Linde and Doe, 2012) contains different semantic contents (animals, humans, etc.). Another factor that could explain the differences is that our database contains images representing only happy and sad emotions, while the images used by (Van Der Linde and Doe, 2012) span many different values in the valence-arousal space of emotions. Finally, with respect to the authors' main conclusion, i.e. that the pleasantness of an image (its valence) is a significant factor that influences subjective rating, we can conlcude from our experimental data that the main influences (differences) are found when assessing the quality of original pristine images, while these differences decrease as the level of distortion increases.

6 CONCLUSIONS

From our experimental results we can conclude that the influence of affective content on the subjective perception of image quality is mainly observed when images are of very high or very low quality. In particular, happy images in their pristine version were more frequently rated as best quality compared to sad pristine images. On the contrary, sad images in their most distorted version were more frequently rated as worst quality compared to the most distorted version of the happy ones. The quality perception was also influenced by the background of the images. When such background was eliminated and only the faces were used as stimuli, the frequency of ratings of pristine images as best quality was lower than in the experiment where background was present, while images most distorted without the background were assigned to the worst quality more frequently compared to the stimuli with background. The classification approach using SVM and a feature space composed of NR quality metrics was able to predict the five quality classes with an accuracy of 72%. Different issues have to be considered in the near future. Increasing the cardinality of the database is important both from the subjective perception side and also to better apply machine learning classification strategies. Moreover other quality metrics for JPEG distortions proposed in the last years can be considered and a proper feature selection strategy will be also considered to reduce redundancies in the chosen metrics. Deep learning based classifiers will also be taken into account in the next future.

REFERENCES

Babu, R. V., Bopardikar, A. S., and Perkis, A. (2004). A perceptual no-reference blockiness metric for jpeg images. In *ICVGIP*, pages 455–460.

Bhattacharya, S., Nojavanasghari, B., Chen, T., Liu, D., Chang, S.-F., and Shah, M. (2013). Towards a comprehensive computational model foraesthetic assessment of videos. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 361–364. ACM.

Chen, C. and Bloom, A. (2010). A blind reference-free blockiness measure. In *Lecture Notes in Computer Science*, volume 6297, pages 112–123. Springer-Verlag Berlin Heidelberg.

Ciocca, G., Corchs, S., and Gasparini, F. (2016). Genetic programming approach to evaluate complexity of texture images. *Journal of Electronic Imaging*, 25(6):061408–061408.

Ciocca, G., Corchs, S., and Gasparini, F. (2017). A complexity-based image analysis to investigate interference between distortions and image contents in image quality assessment. In *International Workshop*

on Computational Color Imaging, pages 105–121. Springer.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619.

Corchs, S., Gasparini, F., and Schettini, R. (2014). No reference image quality classification for jpeg-distorted images. *Digital Signal Processing*, 30:86–100.

Corchs, S. E., Ciocca, G., Bricolo, E., and Gasparini, F. (2016). Predicting complexity perception of real world images. *PloS one*, 11(6):e0157986.

Dan-Glauser, E. S. and Scherer, K. R. (2011). The geneva affective picture database (gaped): a new 730-picture database focusing on valence and normative significance. *Behavior research methods*, 43(2):468.

Ghadiyaram, D. and Bovik, A. C. (2015). Feature maps driven no-reference image quality prediction of authentically distorted images. In *Human Vision and Electronic Imaging XX*, volume 9394, page 93940J. International Society for Optics and Photonics.

Ghadiyaram, D. and Bovik, A. C. (2016). Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387.

Gu, K., Zhai, G., Yang, X., and Zhang, W. (2015). Using free energy principle for blind image quality assessment. *IEEE Transactions on Multimedia*, 17(1):50–63.

Hasler, D. and Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. In *Electronic Imaging 2003*, pages 87–95. International Society for Optics and Photonics.

Junior, O. L., Delgado, D., Gonçalves, V., and Nunes, U. (2009). Trainable classifier-fusion schemes: An application to pedestrian detection. In *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, pages 1–6. IEEE.

Larson, E. C. and Chandler, D. M. (2010). Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006–011006.

Liu, L., Hua, Y., Zhao, Q., Huang, H., and Bovik, A. C. (2016). Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, 40:1–15.

Mack, M. and Oliva, A. (2004). Computational estimation of visual complexity. In *the 12th Annual Object, Perception, Attention, and Memory Conference*.

Minhas, R., Mohammed, A. A., Wu, Q. J., and Sid-Ahmed, M. A. (2009). 3d shape from focus and depth map computation using steerable filters. In *International Conference Image Analysis and Recognition*, pages 573–583. Springer.

Mittal, A., Moorthy, A. K., and Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708.

Mittal, A., Soundararajan, R., and Bovik, A. C. (2013).

Making a completely blind image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212.

Moorthy, A. K. and Bovik, A. C. (2010). A two-step framework for constructing blind image quality indices. *IEEE Signal processing letters*, 17(5):513–516.

Muijs, R. and Kirenko, I. (2005). A no-reference blocking artifact measure for adaptive video processing. In *Proceedings of the 13th European Signal Processing Conference 2005*.

Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59.

Pan, F., Lin, X.and Rahardja, S., Lin, W., Ong, E., Yao, S., Lu, Z., and Yang, X. (2004). Locally-adaptive algorithm for measuring blocking artifacts in images and videos. In *Proceedings of the International Symposium on Circuits and Systems*, volume 3, pages 925–928. IEEE.

Rosenholtz, R., Li, Y., and Nakano, L. (2007). Measuring visual clutter. *Journal of vision*, 7(2):17–17.

Saad, M. A., Bovik, A. C., and Charrier, C. (2010). A dct statistics-based blind image quality index. *IEEE Signal Processing Letters*, 17(6):583–586.

Schettini, R., Gasparini, F., Corchs, S., Marini, F., Capra, A., and Castorina, A. (2010). Contrast image correction method. *Journal of Electronic Imaging*, 19(2):023005–023005.

Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11):3440–3451.

Siahaan, E., Hanjalic, A., and Redi, J. A. (2018). Semantic-aware blind image quality assessment. *Signal Processing: Image Communication*, 60:237–252.

Tamura, H., Mori, S., and Yamawaki, T. (1978). Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, 8(6):460–473.

Triantaphillidou, S., Allen, E., and Jacobson, R. (2007). Image quality comparison between jpeg and jpeg2000. ii. scene dependency, scene analysis, and classification. *Journal of Imaging Science and Technology*, 51(3):259–270.

Van Der Linde, I. and Doe, R. M. (2012). Influence of affective image content on subjective quality assessment. *JOSA A*, 29(9):1948–1955.

Wang, Z., Bovik, A. C., and Evans, B. L. (2000). Blind measurement of blocking artifacts in images. In *Proc. International Conference on Image Processing*, volume 3, pages 981–984. IEEE.

Wang, Z., Sheikh, H., and Bovik, A. C. (2002). No-reference perceptual quality assessment of jpeg compressed images. In *Proc. International Conference on Image Processing*, volume 1, pages 477–480. IEEE.

Wu, H. and Yuen, M. (1997). A generalized block-edge impairment metric for video coding. *IEEE Signal Processing Letters*, 4:317–320.