Inpainted image quality assessment based on machine learning

V. Voronin[1] V. Marchuk[1] E. Semenishchev[1] S. Maslennikov[1] I. Svirin[2]

[1]Don State Technical University
Shevchenko 147
346500, Shakhty, Russian Federation
voronin_sl@mail.ru

[2]CJSC Nordavind
Varshavskoe 125
Moscow, Russian Federation
head@nordavind.ru

ABSTRACT

In many cases inpainting methods introduce a blur in sharp transitions in image and image contours in the recovery of large areas with missing pixels and often fail to recover curvy boundary edges. Quantitative metrics of inpainting results currently do not exist and researchers use human comparisons to evaluate their methodologies and techniques. Most objective quality assessment methods rely on a reference image, which is often not available in inpainting applications. This paper focuses on a machine learning approach for no-reference visual quality assessment for image inpainting. Our method is based on observation that Local Binary Patterns well describe local structural information of the image. We use a support vector regression learned on human observer images to predict the perceived quality of inpainted images. We demonstrate how our predicted quality value correlates with qualitative opinion in a human observer study.

Keywords

Inpainting, quality assessment, metric, visual salience, machine learning.

1. INTRODUCTION

Objective image quality metrics are designed to predict perception by humans based on an image processing without a human observer being involved. Such metric allow assessing an image quality quickly, but existing metrics behave differently in comparison with a quality perceived by human observers. Most of existing methods implement full-reference metrics where complete reference image is assumed to be known. In the case of image inpainting reference image just does not exist. This situation requires a no-reference or "blind" quality assessment approach.

Objective methods for assessing perceptual image quality have traditionally attempted to quantify the visibility of errors between a distorted image and a reference image using a variety of known properties of the human visual system. The most fundamental problem with the traditional approach is the definition of image quality. In particular, it is not clear that artifact visibility should be associated with loss of quality. Some artifacts may be clearly visible but very hard to model numerically.

Several works on objective image inpainting quality assessment have been published in recent years. For instance, an analysis of gaze patterns was involved to quality assessment in work [Ven10]. Authors postulate that perceived by human image quality is related to so-called “saliency”. To quantitatively assess saliency, they compute the gaze density for a given image inside and outside the inpainted region. Resulting quality estimates are achieved as a relation of the gaze densities of an image inside and outside the hole region. Authors have used an eye tracker to estimate a gaze density. This method has the same disadvantages as subjective evaluation.

Most of proposed approaches use saliency maps to estimate visibility of different artifacts in inpainted region. The key idea is based on the change of the saliency map before and after inpainting. In paper [Pau09] this problem addressed by two proposed metrics: average squared visual salience (ASVS) and degree of noticeability (DN). Drawbacks of these metrics are related to the fact that they do not take into consideration the global visual appearance of the image. In [Pau09] proposed another visual saliency based metric. He defined a normalized gaze density measure that uses the original image as a reference, and shows that if there is any change in the saliency map corresponding to the inpainted image, then this change is related to the perceptual quality of the inpainted image. Authors use the visual coherence of the recovered regions and the visual saliency describing the visual importance of an area. This approach shows promising results but addresses only few possible inpainting artifacts.

There is a work that generalize some previous methods like Structural Similarity Index (SSIM)

[Pau10] for image inpainting. This approach is able to achieve good results, but it's completely lacking a high level modeling of a human visual system.

At this point, we may conclude that abovementioned approaches are quite efficient for particular tasks. Nevertheless, existing approaches are weakly correlated with a human perception and, thus, additional investigation on this topic is needed.

2. IMAGE INPAINTING QUALITY

At first the inpainting problem was approached as "error concealment" in the field of telecommunications. The goal of this technique was to fill-in image blocks that have been lost during data transmission. More recently, more elaborated techniques for digital image inpainting such as one presented by Bertalmio et al. [Ber01] have been developed. During the last decade, many methods addressing inpainting problem have been proposed. It leads to the natural need of robust inpainting performance metric. Typically, subjective expert-based approaches are involved which is expensive and time consuming procedure. So, alternative approach has to be developed to address the problem of objective image quality assessment.

The problem of inpainted image quality assessment is highly related to the human visual system modeling problem. In order to design a good quality metric one should take into account its different properties. One way to do this is to model it with machine learning techniques.

Let's introduce basic notations used in our work. The whole image domain I is composed of two disjoint regions: the inpainting region Ω, and the known region Φ $(\Phi = I - \Omega)$ as shown at the figure 1.

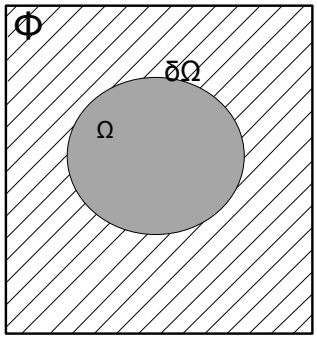

Figure 1. Image model.

Given an image I and a region Ω inside it, the inpainting problem consists in modifying image values of the pixels in Ω so that this region does not stand out with respect to its surroundings. The purpose of inpainting might be to restore damaged portions of an image (e.g. an old photograph where folds and scratches have left image gaps) or to remove unwanted elements present in the image (e.g. a microphone appearing in a film frame). The region Ω is always given by the user, thus the localization of Ω is not a part of the inpainting problem.

It is very difficult to compare the "original" image and an inpainted one, because inpainted region can be large and very different from the corresponding region of the original image. In some cases, a visual image quality may be nearly perfect, but objective quality in terms of pixel-oriented metrics like PSNR will be poor. On way to model human attention and to estimate the visibility of different image areas is to use so-called saliency maps. We exploit this approach together with machine learning to model relations between local geometric patterns and perceived by a human observer image quality.

3. THE PROPOSED METHOD

In [Fra14], we have proposed the inpainting quality assessment technique based on a machine learning approach. Our method allows to receive both low and high level inpainted image descriptions. Next, we have used a support vector regression learned on human observer images to predict the perceived quality of inpainted images. One of the major problems there was a computational complexity.

The main workflow of proposed method is presented on the Figure 2.

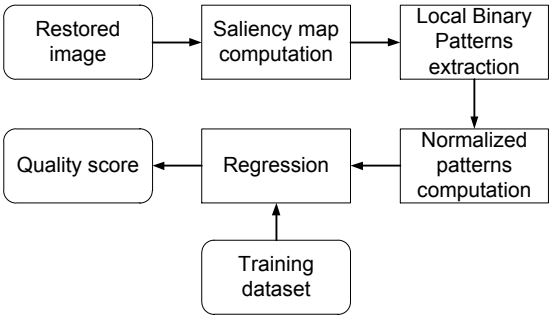

Figure 2. Overall algorithm block scheme.

The first step is to compute an importance of each sub-region of inpainted area. To approach this problem, we use a visual saliency which plays an important role in human visual perception. Human eye at each time clearly sees only a small portion of the space, while a much larger portion of the space is perceived very 'blurry'. The latest information is sufficient to assess the importance of different areas and to draw attention to important areas of a visual field. Most of methods give so-called saliency map: a two-dimensional image in which each pixel value is related to an importance of this region.

It is believed that two stages of visual processing are involved: first, the parallel, fast, but simple pre-attentive process; and then, the serial, slow, but complex attention process. Innovation denotes the novelty part, and prior knowledge is the redundant

information that should be suppressed. In the field of image statistics, such redundancies correspond to statistical invariant properties of our environment. It is widely accepted that natural images are not random, they obey highly predictable distributions. In the following sections, we will demonstrate a method to approximate the "innovation" part of an image by removing the statistical redundant components. This part, we believe, is inherently responsible to the popping up of regions of interest in the pre-attentive stage.

In our work, we use spectral residual approach [Xia07]. It defines an entropy of the image as:

$$H(Image) = H(Innovation) + H(Prior\ Knowledge).$$

This model is independent of features, categories, or other forms of prior knowledge of the objects. By analyzing the log-spectrum of an input image, authors extract the spectral residual of an image in spectral domain. They have proposed a fast method to construct the corresponding saliency map in spatial domain.

Given an input image I with a Fourier decomposition F, the log spectrum $L(F)$ is computed from the down-sampled image with height equal to 64 pixels.

If the information contained in the $L(F)$ is obtained previously, the information required to be processed is:

$$H(R(F)) = H \cdot L(F) | A(F),$$

where $A(F)$ denotes the general shape of log spectra, which is given as a prior information. $R(F)$ denotes the statistical singularities that is particular to the input image. To compute area importance metric we have used the following expression:

$$Q = \frac{1}{\|\Phi\|} \cdot \sum_{p \in \Phi} S(p),$$

where S is the saliency map corresponding to the inpainted image, which gives $S(p)$ as the saliency map value corresponding to pixel p. We have used Q value as a threshold level at the next step and calculated the assessment only for those recovered areas for which $S(p) > Q$.

The saliency map is an explicit representation of proto-objects [Tan11]. We use a simple threshold segmentation to detect proto-objects in a saliency. Given $S(x)$ of an image, the object map $O(x)$ is obtained:

$$O(x) = \begin{cases} 1 & if\ S(x) > threshold \\ 0 & otherwise \end{cases}.$$

Empirically, we set $threshold = E(S(x)) \times 3$, where $E(S(x))$ is the average intensity of the saliency map. While the object map $O(x)$ is generated, proto-objects can be easily extracted from their corresponding positions in input image.

After that we perform feature extraction for found proto-objects. Features are characteristic properties of the artifacts whose value should be similar for artifacts in a particular class, and different from the values for artifacts in another. The choice of appropriate features depends on the particular application.

At present work we use local binary pattern operator (LBP), introduced by Ojala et al. [Tim02]. It is based on the assumption that texture has locally two complementary aspects, a pattern and its strength. The LBP was proposed as a two-level version of the texture unit to describe the local textural patterns. As the neighborhood consists of 8 pixels, a total of 256 different labels can be obtained depending on the relative gray values of the center and the pixels in the neighborhood. An example of an LBP computation is shown on Figure 3.

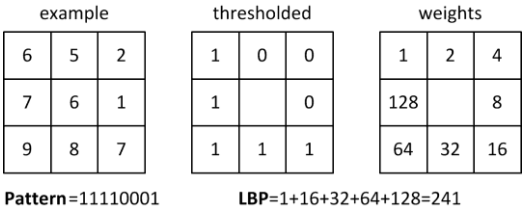

Figure 3. Example of local binary pattern computation.

Learning stage involves word frequency histogram of local binary patterns in salient regions as a feature vector and Support Vector Regression (SVR) as a learning method. Illustration of feature vector is presented in Figure 4.

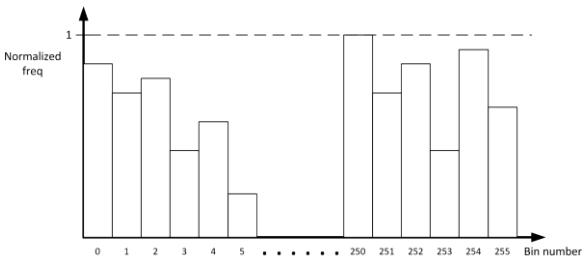

Figure 4. Example of normalized pattern histogram.

Support Vector Machine (SVM) and other kernel methods have achieved a lot of attention recent years, and it has been reported to outperform nearest

neighbor classifier in texture classification. Also, boosting based approaches such as AdaBoost, and bagging classifiers like the Random Forest classifier have been successfully applied to texture classification. The problem of texture retrieval in some extent is related to texture classification.

To compare histograms we use Earth Movers Distance (EMD) [Lev01]. This is a common way to compare two probability distributions (in our case presented by histograms). To incorporate EMD distance into the SVM framework, we use extended Gaussian kernels:

$$K(S_i, S_j) = \exp(-\frac{1}{A} D(S_i, S_j)),$$

where $D(S_i, S_j)$ is EMD if S_i and S_j are image signatures. The resulting kernel is the EMD kernel, A is a scaling parameter that can be determined through cross-validation. We have found, however, that setting its value to the mean value of the EMD distances between all training images gives comparable results and reduces the computational cost.

To learn a regression function, we use a support vector machine regression. As a result, the classification algorithm can be written as:

$$a(x) = sign\left(\sum_{i=1}^{n} \lambda_i c_i x_i \cdot x - b\right).$$

We will use $a(x)$ as a predicted value of image quality.

4. EXPERIMENTAL RESULTS

The experimental method for the subjective quality assessment was chosen the Mean Opinion Score (MOS) [Rib11]. The MOS values are based on subjective data obtained from the experiment. Participants were presented with one inpainted image at a time in a random order and different to each observer. Given an image, the participants were asked to judge the overall image quality of the inpainted image using the quality scale: Excellent, Good, Fair, Poor, Bad. In order to be able to analyse the obtained subjective data, each of the five adjectives in the descriptive quality scale had an equivalent numerical value, or score (not shown to the observers). Accordingly, Excellent corresponded to a 5 score and Poor to a 1 score. The MOS was obtained for each reproduction by computing the arithmetic mean of the individual scores given by participants:

$$MOS = \frac{1}{n} \sum_{i=1}^{n} \text{Score}_i,$$

where n denotes the number of observers, and Score_i the score given by the observer to the inpainted image under consideration. The criterions used to estimate quality presented in the Table 1.

MOS	Quality	Criteria
5	Excellent	Artifacts are Imperceptible
4	Good	Artifacts are perceptible buy not annoying
3	Fair	Artifacts are slightly annoying
2	Poor	Artifacts are annoying
1	Bad	Artifacts are very annoying

Table 1. Quality criteria's

For evaluation purposes we use database of 300 images. Note, that the test images have been chosen to have different geometrical features: texture, structure and real images. After applying the missing mask, all images have been inpainted by four different methods [Oli01, Ber00, Tel04, Cri04]. For each inpainted image, its quality was assessed by 10 human observers. The results were divided into two disjoint subsets. The first was used for training, the second - to verify the results. Some of images from test database are presented at Figures 5-7 (a - images with missing pixels, b - images reconstructed by the Smoothing, c - images reconstructed by the Navier-Stokes, d - images reconstructed by the Telea, e - images reconstructed by the EBM).

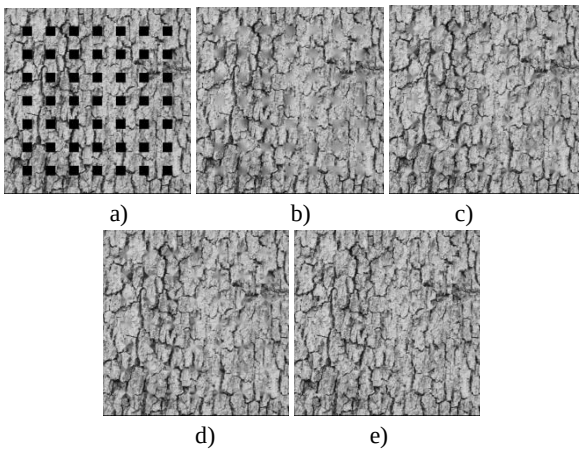

Figure 5. Examples of texture images from test database.

Thus, in an attempt to establish a ranking of the considered algorithms in terms of perceived quality of the inpainted images, and considered the database described above, a psychophysical experiment will be carried out, according to the specifications. The obtained raw perceptual data will be statistically analyzed in order to determine the ranking of the inpainting algorithms. To evaluate the objective quality assessment methods, we use the MOS. Furthermore, the prediction accuracy of proposed

metric was evaluated using Spearman rank order correlation coefficient (SRCC) for proposed metric results and subjective MOS estimation. Results of numeric comparison are presented in the Table 2.

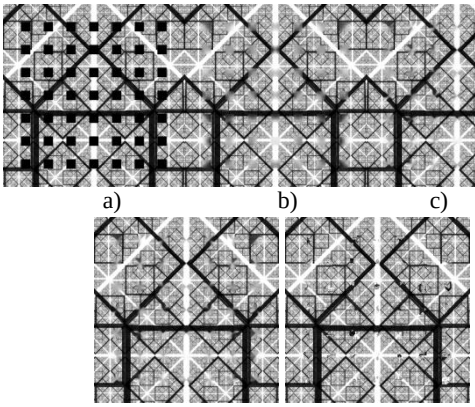

Figure 6. Examples of structure images from test database.

Proposed objective quality metric shows strong correlation with perceived by human quality. Thus, our approach is quite efficient to estimate quality of the inpainted images.

Methods		*MOS*	$\overline{MOS}$	*SRCC*
Smoothing [Oli01]	texture	2.21	2.08	0.84
	structure	1.58		
	image	**2.45**		
Navier-Stokes [Ber00]	texture	**3.15**	2.69	0.95
	structure	1.73		
	image	3.21		
Telea [Tel04]	texture	3.08	2.78	**0.96**
	structure	2.02		
	image	**3.23**		
EBM [Cri04]	texture	**4.41**	**3.69**	0.93
	structure	3.13		
	image	3.54		

Table 2. Spearman rank correlation of proposed metric results and subjective MOS estimation

Results of comparison Spearman rank order correlation coefficient, which finds the linear relationship between two variables using the formula for our method with several popular methods are presented in Tables 3.

DN	**SSIM**	**ASVS**	**PROPOSED METRIC**
0.61	0.71	0.68	**0.78**

Table 3. CC comparison

These tables show that our approach outperforms known and widely used algorithms on a selected image dataset in term of correlation coefficient.

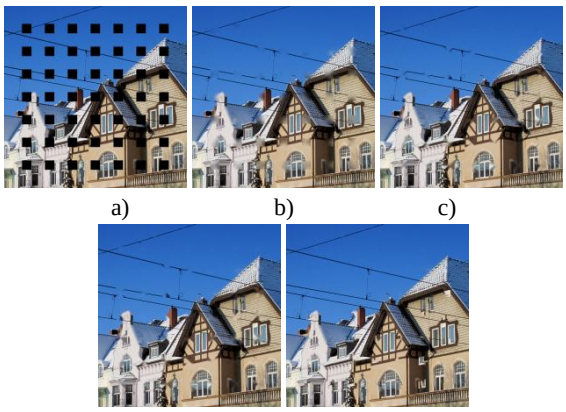

Figure 7. Examples of real images from test database.

5. CONCLUSION

In this work we have presented a novel no-reference inpainting quality assessment technique which is based on a machine learning approach. Our method use inpainted image description by local binary patterns weighted by visual importance. Next, we have used a support vector regression learned on human observer images to predict the perceived quality of inpainted images. We have demonstrated that predicted quality value highly correlates with a qualitative opinion in a human observer study.

6. ACKNOWLEDGMENTS

The reported study was supported by the Russian Foundation for Basic research (RFBR), research project №15-37-21124\15 and grant of the President of the Russian Federation for the young scientists №МК-6986.2015.8.

7. REFERENCES

[Ber01] Bertalmio, M., Bertozzi, A., Sapiro, G. Navier-Stokes, fluid dynamics, and image and video inpainting, Hawaii: Proc. IEEE Computer Vision and Pattern Recognition (CVPR), pp. 213-226, 2001.

[Ber00] Bertalmio, M., Sapiro, G., Caselles, V. and Balleste, C. Image inpainting, New Orleans: Proceedings of SIGGRAPH, pp. 102-133, 2000.

[Cri04] Criminisi, A., Perez, P., and Toyama K. Region filling and object removal by exemplar-based image inpainting, IEEE Transactions on Image Processing 13, pp. 1200–1212, 2004.

[Fra14] Frantc, V.A., Voronin, V.V., Marchuk, V.I., Sherstobitov, A.I., Agaian, S., Egiazarian, K. Machine learning approach for objective inpainting quality assessment, Proc. SPIE 9120, Mobile Multimedia/Image Processing, Security, and Applications 2014, 91200S, 2014.

[Lev01] Levina, E., Bickel, P. The Earth Mover's Distance is the Mallows Distance: Some Insights from Statistics, Proceedings of ICCV 2001 (Vancouver, Canada), pp. 251–256, 2001.

[Oli01] Oliveira, M., Bowen, B., Kenna, R. Mc, and Chang, Y.-S. Fast Digital Image Inpainting, In Proc. VIIP, pp. 261-266, 2001.

[Pau09] Paul A., and Singhal, A. Visual salience metrics for image inpainting, Proc. IS&T/SPIE Electronic Imaging, 2009.

[Pau10] Paul, A., Singhal, A., and. Brown, C. Inpainting quality assessment, Journal of Electronic Imaging, vol. 19, pp. 011002-011002, 2010.

[Rib11] Ribeiro, F., Florencio, D., Cha, Zhang, Seltzer, M. CROWDMOS: An approach for crowdsourcing mean opinion score studies, IEEE International Conference on, Acoustics, Speech and Signal Processing (ICASSP), pp. 2416 – 2419, 2011.

[Tan11] Tang, Huixuan, Neel, Joshi, and Ashish, Kapoor. Learning a blind measure of perceptual image quality, IEEE Conference on, Computer Vision and Pattern Recognition (CVPR), pp. 305 – 312, 2011.

[Tel04] Telea, A. An image inpainting technique based on the fast marching method, Journal of Graphics Tools, vol. 9, no. 1, ACM Press, pp. 25-36, 2004.

[Tim02] Timo, Ojala, Pietikainen, Matti, and Maenpaa, Topi. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on, Pattern Analysis and Machine Intelligence, pp. 971-987, 2002.

[Ven10] Venkatesh, Vijay, M., and Cheung, S.S. Eye tracking based perceptual image inpainting quality analysis, Image Processing (ICIP), 17th IEEE International Conference on IEEE, pp. 1109 – 1112, 2010.

[Xia07] Xiaodi, Hou, and Zhang, Liqing. Saliency detection: A spectral residual approach, IEEE Conference on, Computer Vision and Pattern Recognition, 2007.