Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11872-2

1197: ADVANCES IN SOFT COMPUTING TECHNIQUES FOR VISUAL INFORMATION-BASED SYSTEMS

Blind image inpainting quality assessment using local features continuity

Amine Mohamed Rezki[1] · **Amina Serir**[1] · **Azeddine Beghdadi**[2]

Received: 31 October 2020 / Revised: 29 September 2021 / Accepted: 23 December 2021

Abstract
This paper deals with Blind Inpainted Image Quality Assessment BIIQA. Herein, we propose a new method that exploits the continuity of features around the boundaries of the retouched area. Indeed, we believe that the quality of inpainted images depends on how the edges and textures have been reproduced inside the hole. Besides, this concept has been formalized by the fact that features should be reproduced inside and outside the hole with respect to structures continuity. Furthermore, one could compare these features in terms of continuity and estimate the global quality of the inpainted image. And since the local structures are represented by patches, we proposed as a secondary contribution, an improvement of a patch classification algorithm. The strength of this metric unlike most existing IIQA metrics, is that it is completely blind and does not require any reference, making it well suited to the inpainting assessment, where reference is usually unavailable. The proposed BIIQA has been tested on TUM-IID database, where the results of four commonly used inpainting algorithms are provided and compared against IIQA state-of-the-art. The obtained results show clearly that our method outperforms the existing ones.

Keywords Inpainting · Image quality assessment · Inpainting evaluation · Image modification

1 Introduction

Inpainting which is the process that allows the reconstruction of images with lost parts or deleted zones has been increasingly used over the two last decades. Inpainting has been used in multiple applications: object removal, old images restoration, occlusion removal...

✉ Amine Mohamed Rezki
mrezki@usthb.dz

[2] Laboratory of Image Processing and Radiation(LTIR), Faculty of Electronics and Computer Science, University of Science and Technology Houari Boumediene (USTHB), Algiers, Algeria

[1] Laboratory of Information Processing and Transmission (L2TI), Institut Galilée, Université Sorbonne Paris Nord, 93430 Villetaneuse, France

and so on. Many methods in this topic have been proposed, with various results in terms of the performance which is often evaluated by subjective visual appreciation.

The purpose of quality assessment in case of inpainting is to expose how well the algorithm has filled the hole while preserving the continuity and the coherence of edges, textures, and smooth areas. Besides, the image quality assessment could quantify the details preservation. Thereby, in the inpainting process, an image quality assessment IQA could be used to tune the parameters of the used algorithm. In other words, in the case of inpainting based on setting some parameters, IQA could be used in an iterative way to tune the parameters in the aim to reach the best quality assessment. The iterative process is stopped when the quality measure reaches its maximum. More reliable methods consist of determining the image quality of inpainted images subjectively, when some observers could be invited to appreciate visually the processed image quality. This process is time-consuming and difficult to perform. Moreover, applying the subjective IQA to compare a new inpainting method against existing ones, is a very difficult task, since it requires to evaluate every time a great number of images. By this fact, using an objective IQA for inpainting should perform more accurate results in both comparative way and controlling the inpainting process.

Also, as the reference images in real inpainting applications is often nonexistent, all full-reference-based IQA i.e. PNSR, MSE, SSIM cannot be used for inpainting quality assessment. Herein we propose to conceive a new Blind Inpainted Image Quality Assessment B-IIQA, which consists of exploiting the image features contents and the coherence between inpainted areas and their close neighbors. Furthermore, the edge and texture continuity could be an efficient indicator of inpainting performance. The main contributions in the present work are:

- We propose a new completely blind metric dedicated to the evaluation of the quality of image inpainting. It is based on the classification of patches and the continuity of local structures.
- Since our metric is based on local structures, we have proposed an improved version of the patch classification algorithm [5].

The rest of this paper is organized as follows: the further section mentions the state-of-the-arts of IIQA. The proposed method is detailed in Section 3. The conducted experiments and the discussion of the obtained results are presented in Section 4. We conclude this work with some new prospects in the last section.

2 Related work

Inpainting is one of the challenging topics in image processing because it consists to fill conveniently unknown areas from their known neighbors. This is done automatically without human interaction, and blindly without any knowledge on the content of the area to be filled or available ground truth. As the reference image does not exist, then all commonly used metrics PSNR, MSE, SSIM, based on similarity with a reference, could not be used for the IIQA. However, in the literature, few IIQA metric were proposed and could be divided into two categories, full reference FR- IIQA and no reference NR-IIQA. To the best of our knowledge, unlike the IQA metrics, there is no IIQA based on the reduced reference(RR). The First IIQA metric proposed by Wang et al [17] is based on similarity by exploiting the three parameters: luminance, definition, and gradient to provide PWIIQ (Parameter Weight Image Inpainting Quality).

Ardis and Singhal [1] proposed two metrics. Average Squared Visual Salience (ASVS) and Degree of Noticeability (DN) using visual saliency that is inspired from the human visual attention model. They noticed that saliency takes high values on inside and around the artifacts areas. ASVS corresponds to the normalized sum of squared salience of inpainted pixels, and DN measures the attention change compared to the original image on inpainted regions, their results remain to be confirmed since they use only five observers. In [11] Mahalingam et al observe that there is a correlation between inpainting quality and visual attention, they rate inpainting quality by comparing gaze densities with the original within and outside the hole region of inpainted images, using eye-tracking experiments. In the same way, Oncu et al proposed BorSal [12] by applying GD equation only in borders, extends three pixels inside and outside the gap, and StructBorSal that combines the BorSal metric with SSIMIPT the color version of SSIM, to take into account the structural information on local neighbors.

The IIQA metrics based on saliency can be effective only on the presence of some special artifacts that grab human attention due to sudden change but they are insufficient to evaluate inpainting quality. By the way, Isogawa et al [10] prove that saliency map-based IQA methods are limited because the computational saliency maps do not reflect actual human gaze patterns, especially for inpainted images. Besides, the major drawback of all the measures presented above is the need for the reference image. Thus, they are useless for inpainting evaluation. The first NR-IIQA was proposed by Dang et al in [3], the authors have proposed to use the local coherence, weighted by the saliency map to give importance to the attractive areas, edges, corners, and textures. The approach is innovative but the coherence was computed basically, just the maximum of structure and hue similarity between each inpainted patch and the neighbor's patches. However, this method cannot detect some kind of inpainting defects like edge discontinuity. Recently, Wenjin et al [8] found a simple approach to evaluate inpainted images, but it can be applied just to a special kind of images named Thangka, related to the Asian culture, featured by asymmetrical patterns repetition.

More lately, some methods based on machine learning have been proposed. In [15] and [16], the authors have used Support Vector Regression (SVR) learned to predict the score of inpainting, by learning from local features as LBP pattern or DCT coefficients. In [10] and [9], the authors have treated the problem differently, starting from the idea that selecting the best is rather easier than providing the inpainting quality scores. They proposed to train learn-to-rank model, which gives as output the order of the inpainted images, from best to worst, without providing the absolute scores.

All those machine learning-based methods have been trained and tested on locally created datasets which makes the comparison with those metrics becomes a bit difficult. Indeed, the learning phase requires a great amount of labeled data, and actually, large inpainting database is not available. This is why, we propose a new method which is based on structures and features continuity or consistency to translate the perceptual inpainted image quality.

3 The proposed method

To date, the inpainting proposed algorithms are unable yet to reconstruct images with such quality that artists can do it manually. And almost all inpainting algorithms generate artifacts, sometimes invisible but often perceptible to the naked eye. Before starting in the heart of the subject and the description of the proposed method, it is necessary to review the different types of artifacts caused by inpainting algorithms .

(a) (b) (c)

Fig. 1 Examples of inpainting weakness and defects

3.1 Inpainting weakness and resulting artifacts

The encountered artifacts in the inpainting process can be summarized into three types [13]: the blur, edge discontinuity, texture inconsistency. Figure 1a depicts, the blur effect introduced by inpainting without taking into account the existing edges. Figure 1b shows an example of edge discontinuity[6] where we see clearly that the edges do not converge properly during the completion. Figure 1c illustrates the problem of texture inconsistency.

3.2 Method description

To discuss our method, we must start by giving briefly the concept of image inpainting and some notations used in this work. Let us consider the hole area Ω the missing part of the image or the region that should be inpainted, and Φ the outside area that contains intact pixels. And usually, these two regions are controlled using a binary image called the binary mask image. Figure 2 illustrates the general inpainting concept and notations used in this work.

The inpainting process consists on completing the hole Ω, in a way to respect the local continuity of structures and features from the Φ area into Ω.

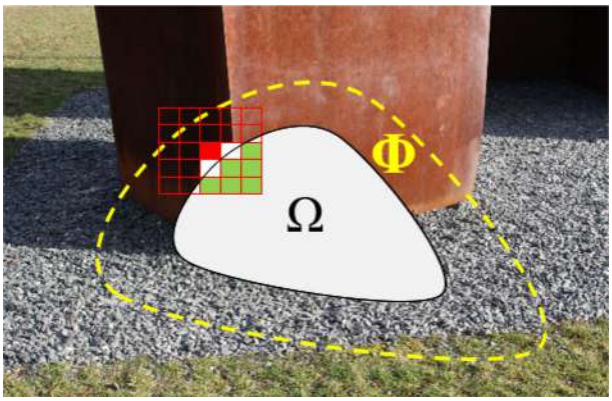


Fig. 2 Inpainting concept and used notations

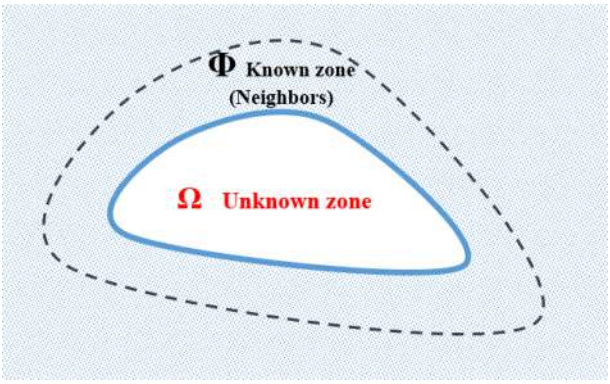


Fig. 3 Block subdivision. p_i in red, $q_{i,j}$ in green

Thus, let us consider the patches inside and outside the mask. These patches could be classified as:

- Patches p_i, which are completely outside the mask and are close to the frontiers, where i=1,2 ....M. With M is the number of patches p_i, (the red patch in Fig. 3).
- Patches $q_{i,j}$, which belong completely to the mask and are close to the frontiers, where j=1,2 ...J_i. And J_i is the number of patches $q_{i,j}$ in the neighborhood of p_i, (the green patches in Fig. 3).
- Patches that have some pixels inside the mask and others outside (white patches in Fig. 3), are ignored and are not considered during the inpainting quality evaluation. For each patch p_i there exist a number J_i of patches $q_{i,j}$ which belong to the neighbor of p_i of size WxW. Figure 3 depicts an example of patch p_i in red color and its 5x5 neighboring patches (two direct neighbors in all directions), where only five $q_{i,j}$ patches with green color could be considered to evaluate the quality of inpainting on the patch p_i. Indeed, we consider only the patches which are completely inside the hole.

The aim of this paper is to propose a new metric that evaluates the quality of image inpainting. This quality assessment method is blind and does not require any reference. The proposed method is based on estimating the contribution of each patch p_i to evaluate the overall image quality score. We believe that the inpainting quality assessment depends essentially on the structure's continuity between the Φ and Ω areas. Roughly speaking, edges must be correctly continued in the masked area. Moreover, texture and smooth areas should be similarly continued. This is why we propose to treat each type of local structure separately.

Therefore, in the proposed assessment process, we first classify all patches pi and $q_{i,j}$ into edges, textures or smooth patches. Then, we evaluate the continuity of local structures by exploiting the local features extracted from edge, texture and smooth patches.

For each patch p_i, we compute the local contribution score Qe_i or Qt_i or Qs_i depending on its type. This contribution score translates how the reconstruction is performed over this patch. Finally, we obtain the global quality score of the inpainted image using the weighted sum of all contribution scores. The proposed IIQA is summarized in the flowchart in Fig. 4.

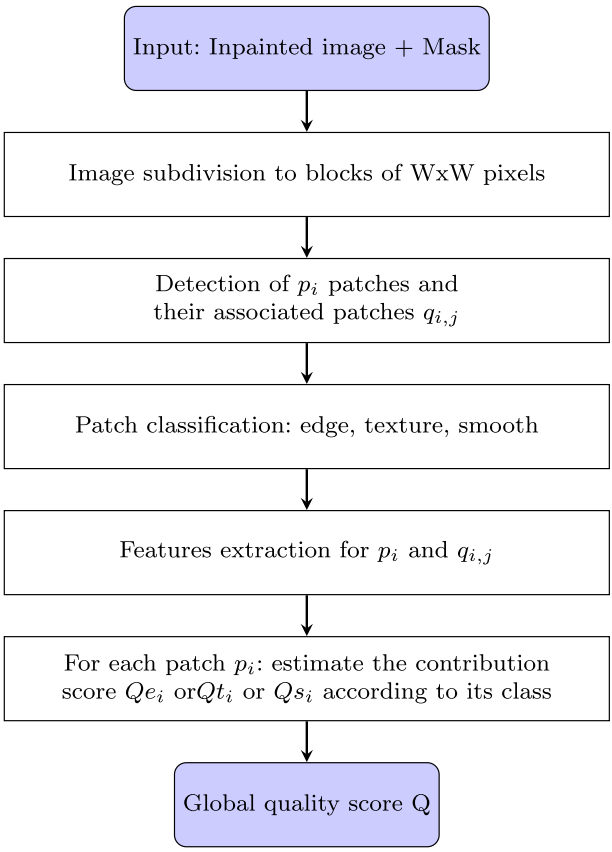


Fig. 4 Flowchart of the proposed blind IIQA method

3.3 Selection of patches of interest

By observing and analyzing several inpainted images, we noticed that inpainted patches situated in the borders of Ω attract more the human eye, and the consistency of these patches with their neighborhood is usually correlated with the overall quality of the image. For this, after subdividing the input image into W x W blocks we select only the patches that are close to the mask borders, and we assume that those patches could contribute to our quality evaluation process. Let S denotes the set of all selected patches p_i and their associated ones $q_{i,j}$ that meet the criteria mentioned previously. The p_i patches are selected by using the binary dilation tool over the mask, then the associated $q_{i,j}$ patches are detected from the neighborhood of p_i. After that, we perform the patch's classification and features extraction for each patch S_k belonging to the set S as it will be explained in the next paragraphs.

3.4 Patch classification

To correctly model the continuity of local structures, we propose to classify all selected blocks into three classes smooth, texture, or edge. So we attribute to each block of WxW size, one label from the three according to its content. To determine the most appropriate

W-value that could well discriminate edges from textures and smooth regions, we tried 4x4, 8x8, and 16x16 patch sizes.

If one considers patches of size 4x4, the analysis of the texture which can be considered as a set of micro-structures, can only be a partial analysis. In this case, the analysis can either fall on a smooth part of the texton or a part presenting marked structures which can be confused with contours. Thus, the texture should be observed in its globality as a set of textons, in order to avoid that the micro-structures could be considered as micro-contours which should distort the evaluation of the continuity of the structures.

Unlike the 4x4 case, considering 16x16 size patches, most edge patches are classified as texture patches. Indeed, a window of such size may contain several close edges and they will be considered as a single texture set. In this case, the assessment of edge continuity is ignored. For this reason, it is important to choose an intermediate window size, such as 8X8, which gives more significant scores.

Figure 7(b-d) depicts an example of patch classification, using window sizes of 4x4, 8x8, and 16x16. We can see that the classification result using 8x8 patches is the closest to reality. In the present work, we will divide the input image into patches of 8x8 pixels.

The classification idea is inspired from [5]. Firstly, the Discrete Cosine Transform (DCT) is applied to the 8x8 image block. Then, the DCT coefficients $C_{i,j}$ are used to make a preliminary decision, by computing for each patch S_k, four parameters A_{tot}, $R1$, $R2$ and $R3$ as follows:

$$A_{tot}(S_k) = \sum_i \sum_j |C_{ij}| - |C_{00}|, \quad R1(S_k) = \frac{med(V, H, D)}{\max(V, H, D)},$$

$$R2(S_k) = \frac{\min(V, H, D)}{med(V, H, D)}, \quad R3(S_k) = \frac{\min(V, H, D)}{\max(V, H, D)},$$

where the sets V,H and D are given by:

$$V = \{C_{i,j}/i = 1, 2 \quad and \quad j = 3, 4, 5\}, \quad H = \{C_{i,j}/i = 3, 4, 5 \quad and \quad j = 1, 2\}$$

$$and \quad D = \{C_{i,j-1}, C_{i,j}/i = 3, 4\} \cup \{C_{i,j+1}/i = 2, 3\}.$$

According to [5], we can distinguish the smooth patches from the others just by comparing the A_{tot} value to a fixed threshold T_0. And to differentiate edge patches from texture ones, we can use the following condition:

For each considered block S_k
if ($R1(S_k) > T_1$) and ($R2(S_k) > T_2$) and ($R3(S_k) > T_3$) **then**
  $S_k$is a texture.
else
  $S_k$is an edge.
end if

where T_1, T_2, T_3 are thresholds.

The authors of [5] do not specify the values of these four thresholds. Thus, to establish the relationship between the four parameters and the patch class, we have selected 1250 patches and we have manually labeled them into 230 smooths, 510 textures, and 510 edges. Then, for each patch, A_{tot}, $R1$, $R2$ and $R3$ parameters have been computed.

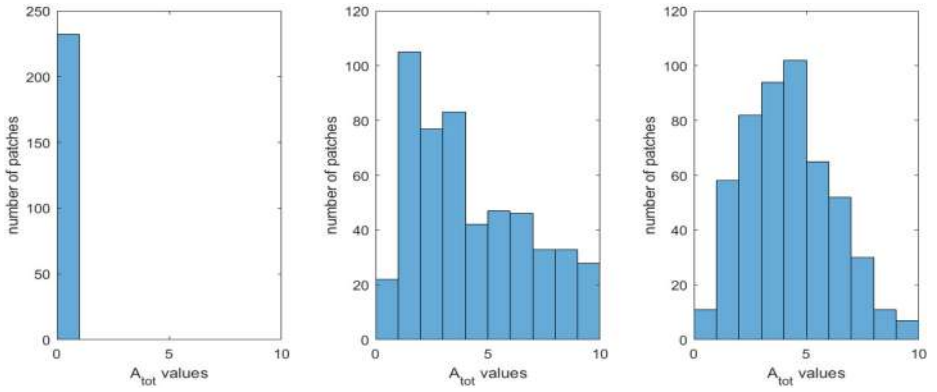


Fig. 5 Distribution of A_{tot} values for 1250 patches, (a) smooth, (b)edge and (c) texture patches

Figure 5a shows the distribution of A_{tot} values according to the three classes. We can see clearly that all smooth patches have a small value of A_{tot}. And for $A_{tot} \leq 1$ only 0.3% of texture and edge patches could be confused as smooth ones, as shown in Fig. 5b and c.

Besides, to set threshold values and to verify the classification result using the compound condition, we have studied the effect of each sub-condition separately. Then, we have applied each sub-condition to the selected 510 edge and 510 texture patches. For each one of them, we have varied the associated threshold, and the Correct Classification Rate CCR has been evaluated.

Figure 6 represents for each sub-condition, the variation of the CCRs for both the edges and texture patches separately and that the intersection of the two curves corresponds to the optimal point. Indeed, if the threshold is set to greater or less than the value of optimal point, one of the two classes (texture or edge) will be favored over the other.

By combining the three sub-conditions and using the optimal thresholds values T1=0.6, T2=0.66, T3=0.38, the CCR attained 71% for edge and texture patches.

Figure 7 shows an image subdivided into blocks of 8x8 pixels, and classified into three categories: edge, texture, or smooth. Figure 7-d depicts the patch classification results when we apply the method inspired from [5]. It is worth noticing that edge patches could be confused with texture ones, because it is not always true that DCT coefficients of texture patches are distributed uniformly, especially for structured and oriented textures.

For the purpose to improve this classification results and avoid false edges, we propose to use the Canny edge detector. Indeed, this operator does not only detect the variations of the gray levels that are also found in the textured areas but favors the edges continuity by introducing the thresholding by hysteresis and the information on edge direction. Therefore, the number of canny-edge points detected in each selected patch could discriminate edges from texture areas. Indeed, if the number of edge points Ne in S_k is greater or equal to four points, then the patch contains an edge. On the contrariwise, it contains a texture. Figure 7-e shows the result of patches classification by adding the condition of canny-edge points number to the first classification. It is worth noticing that some patches have been classified as an edge in the first time, are correctly classified as texture after adding the canny-edge condition. Using this new compound condition improves the CCR to 84% for edge and texture patches. Thus, all selected patches are classified using the Algorithm 1.

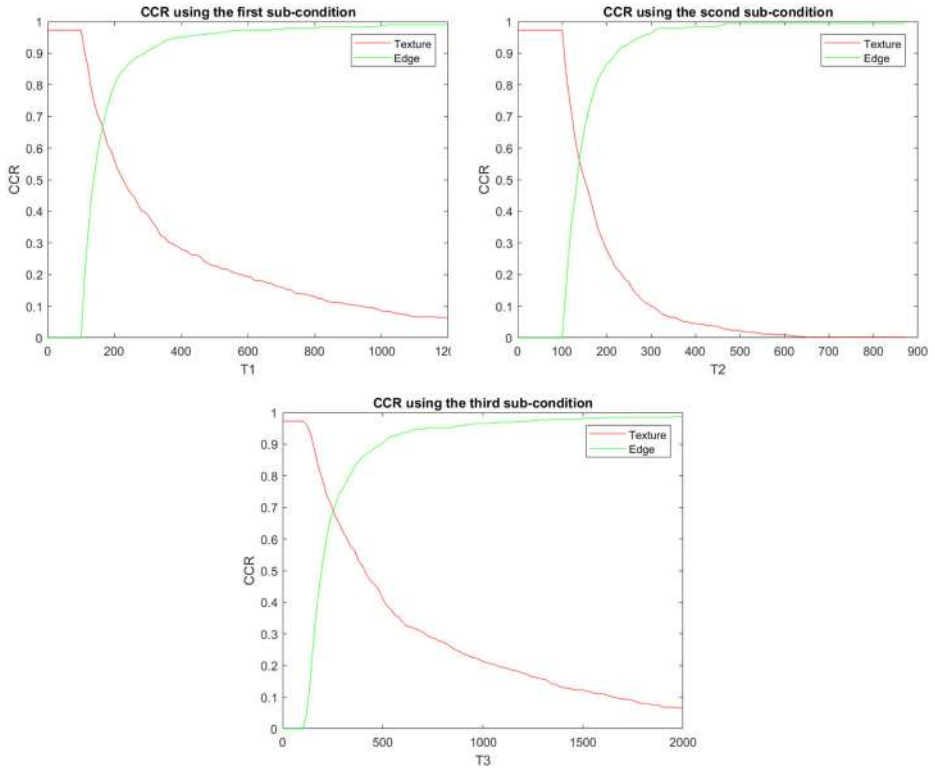


Fig. 6 Correct classification rates of texture patches in red, and edge patches in green, as function of T_1, T_2, T_3, respectively from the top to the bottom

Algorithm 1 Block classification

Inputs : Inpainted image I, Canny-edge image I_e, list of selected patches S.

For each patch S_k:

-Calculate the DCT coefficients.
-Compute$A_{tot}(S_k)$

if $(A_{tot}(S_k) < T_0)$ **then**

$\Rightarrow S_k$ is smooth.

else
Compute: $R1(S_k), R2(S_k), R3(S_k)$ and N_e, .

if (($R1(S_k) > T_1$) and ($R2(S_k) > T_2$) and ($R3(S_k) > T_3$)) or (N_e <4) **then**

$\Rightarrow S_k$is a texture.
else
$\Rightarrow S_k$ is an edge.
end if
end if

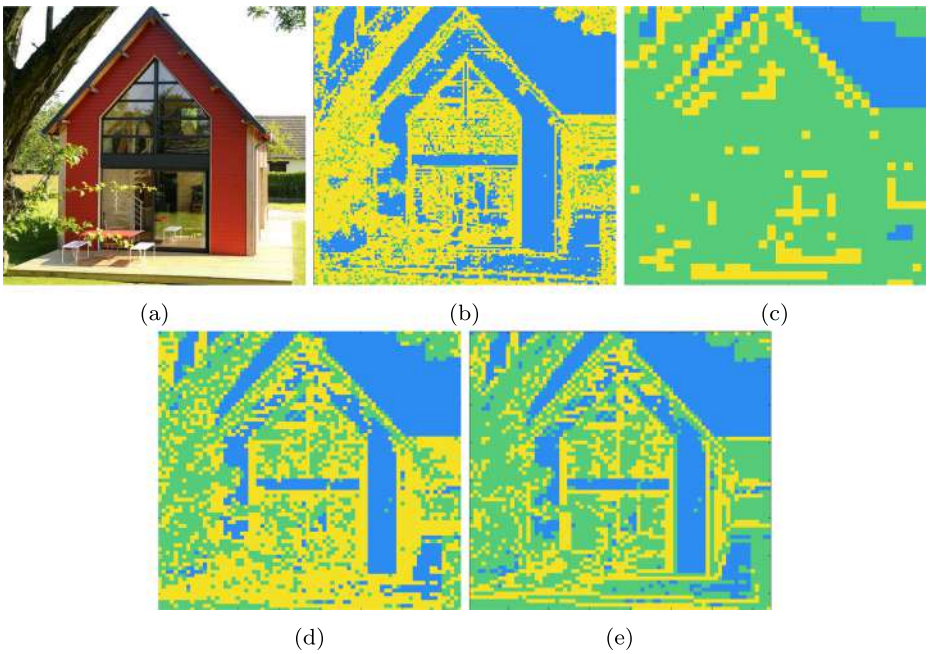


Fig. 7 Example of patches classification. (a) the input image. (b-c) patch classification using 4x4 and 16x16 block sizes respectively. (d) 8x8 patch classification result without canny-edge condition. (e) 8x8 patch classification with canny-edge condition

3.5 Features extraction

To evaluate the continuity of local structures between neighboring blocks, we must extract some features according to the class of each block as follows:

- For edge patches we assume that the patch p_i classified as edge, respects the edge continuity if there exists at least one patch among the $q_{i,j}$ patches, classified as edge too. Moreover, the edge orientation of the patch p_i is close to the selected $q_{i,j}$ edge direction. Thus, we calculate first the principal orientation that we have quantified into 36 integer values (from 1 to 36) to represent all directions from 0° to 360°. Figure 8 shows an example of edge direction for an original and inpainted image, where we can see for the original image, the edges are well continued and most of the adjacent patches have the same direction. Contrariwise, for the badly inpainted image, many edges are broken and the orientations of adjacent blocks are different.

- For textures patches, we assume that the texture continuity is respected if the extracted features for both p_i and a selected texture patches $q_{i,j}$ are very close. To extract texture features, we have considered only the top-left 4x4 block from the 8x8 DCT block as a feature vector as shown in Fig. 9, and we have taken off the first coefficient known as DC coefficient. The remaining fifteen DCT coefficients are considered as a feature's descriptor $Coef_{dct}$ during the evaluation of the texture uniformity.

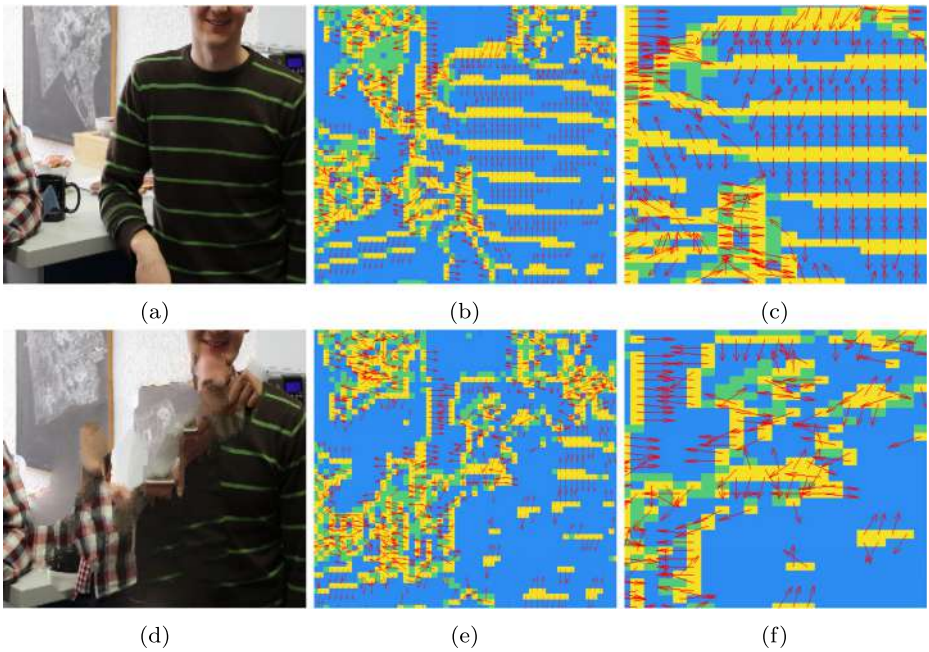


Fig. 8 The orientation of edge patches of an original image on the top, and of a badly inpainted one on the bottom. (c) and (f) are the zoom versions of (b) and (e)

- For smooth patches, which is the simplest case, we have characterized the smooth patch by its mean gray-level value, which should be very close to the mean gray value of a selected smooth patch $q_{i,j}$. It is enough feature to verify the similarity of smooth areas.

3.6 IIQA evaluation and contribution score Q_i

After having classified all the patches of the S set (all p_i patches and their associated ones $q_{i,j}$) as edges, textures, and smooths. And having extract the representative features for each one of them according to its class. Where, for edges we used orientations, for textures we used DCT coefficients and for the smooth area we used the mean intensity value. We start to assess the continuity of structures from the outside to the inside of the hole.

To estimate the contribution quality score of the considered patch p_i, we seek among its associated interior patches $q_{i,j}$, the patch $q_{i,k}$ that is most similar to it. The selected patch $q_{i,k}$ must belong to the same class as p_i and have the closest features to the latter. After that, we propose to estimate whether the structures and textures continuities between p_i and $q_{i,k}$ are respected using the formulas which will be discussed in details below.

3.6.1 Edge continuity score

We consider that an edge could be well continued over two blocks when they have almost the same or very close orientation. So for each edge block p_i, we consider the associated edge patches $q_{i,j}$. The edge continuity is respected if there exists at least one edge patch $q_{i,k}$ which has the same or close quantified edge orientation as $Orient(p_i)$. The edge continuity is not respected if there is no edge patch $q_{i,k}$. And in the case where there exist edge patches

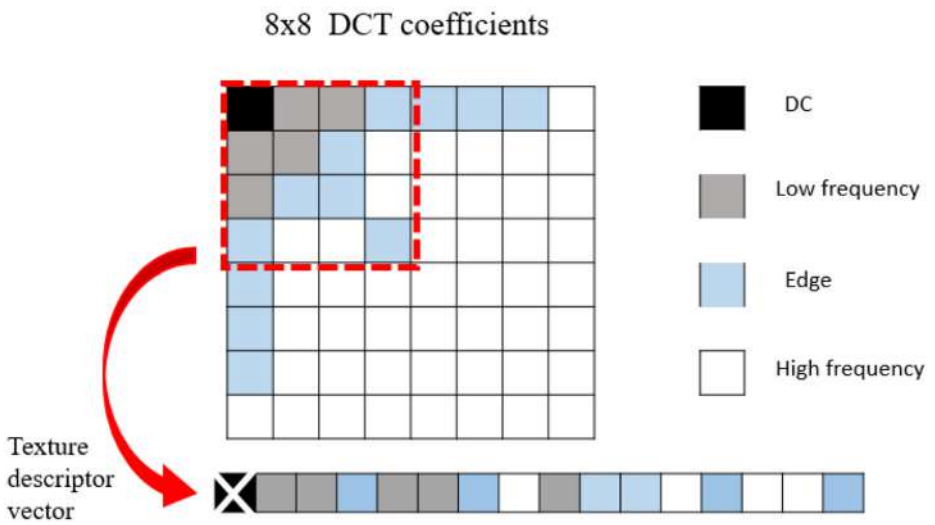


Fig. 9 Texture feature descriptor extracted from DCT coefficients

associated to p_i, but they have not the same orientation as p_i, we propose to consider $a_i = \min_{j=1,2,\ldots J_i}(|Orient(p_i) - Orient(q_{i,k})|)$ which is the minimum distance between the quantified edge orientation of p_i and all the associated edge patches $q_{i,k}$. As all orientations are quantified with integer values 1,2,...36, the difference between two orientations a_i also takes integer values (0, 1, 2, 3... 35).

Herein, the edge continuity depends on a_i value. The greater the value a_i, the less continuity is respected. Therefore, for each patch p_i, we propose an edge score Qe_i which is equal to 1 when the edge continuity is respected and 0 in the opposite case. Indeed, when the difference of quantified orientations a_i is equal to 0 or 1, this means that the orientations of edges are shifted with less than 10°, it is considered that the continuity of the edge in the patch p_i is well ensured on q_i and the edge score for this patch is equal to 1. In opposite, when the difference a_i is greater than 4, the edges suddenly change orientation with a minimum offset of 40°, we have considered that the continuity of the edge orientation is broken, so the edge quality takes 0 as score.

In the case of a_i belongs to [2; 4] the edge continuity is expressed by $1 - \alpha \times a_i$, where α is a constant. This latter is set to 0,2 to weaken the local edge score and by consequence penalize the global score. This strong penalization is based on the knowledge that the distortion caused by edge discontinuity is more attractive and annoying to human perception.This is done in the aim to quantify the degree of edge continuity when a_i belongs to [2; 4] and spread the score over the interval [0, 2; 0, 6]. Then, we propose to evaluate the edge continuity Qe_i for each patch p_i as follows:

$$Qe_i = \begin{cases} 1, & \text{if} \quad a_i = 0, 1 \\ 1 - 0.2 \times a_i, & \text{if} \quad a_i = 2, 3, 4 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

3.6.2 Texture consistency score

We consider that texture is correctly continued over two blocks when they have almost the same or very close texture features. So for each texture block p_i, we consider the associated texture patches $q_{i,k}$. The texture consistency is correctly continued, if there exists at least one texture patch $q_{i,k}$ which has the same or close texture features as p_i. Herein, we propose to characterize the patch texture by the DCT coefficients. To evaluate the texture consistency score Qt_i and similarity between each texture patch p_i with the associated texture patches $q_{i,k}$, we use the Euclidean distance between their DCT feature vectors $Coef_{dct}$, as follows:

$$Qt_i = 1 - \min_{j=1,2,\ ...J_i} (dist(Coef_{dct}(p_i), Coef_{dct}(q_{i,k}))) \tag{2}$$

If no texture patch $q_{i,k}$ exists, Qt_i should be set to 0.

3.6.3 Smooth continuity score

We believe that when the patch p_i is smooth, it exists a patch $q_{i,j}$ which is smooth too, and has a close mean value. Perceptually, when there is a large discrepancy with neighboring smooth patches values, the inpainting score should have a value that represents this inconvenience. Then for smooth patches, we compare the average value of the smooth patch $\mu(p_i)$ with those of their associated smooth patches $\mu(q_{i,k})$, and the smooth similarity score Qs_i is evaluated as follows:

$$Qs_i = 1 - \frac{\min_{j \in J_i}(|\mu(p_i) - \mu(q_{i,k})|)}{max(\mu(p_i), \mu(q_{i,k_{min}}))} \tag{3}$$

If there is no smooth patch among the $q_{i,k}$ patches, Qs_i should be set to 0.

3.6.4 Global quality evaluation

When evaluating the quality of the painted image, we consider all the patches p_i and their associated quality assessment Qe_i, Qt_i, Qs_i. Let us consider m1, m2, m3, the number of edge, texture, and smooth patches, respectively, and M the sum of $m_1 + m_2 + m_3$. then we propose to evaluate the global B-IIQA as follows:

$$Q = \frac{1}{M} \times \left(\sum_{i=1}^{m_1} Qe_i + \sum_{i=1}^{m_2} Qt_i + \sum_{i=1}^{m_3} Qs_i \right) \tag{4}$$

Let us consider the percentage of edge patches $\alpha = \frac{m_1}{M}$, the percentage of texture patches $\beta = \frac{m_2}{M}$, and $\gamma = \frac{m_3}{M}$ for smooth patches. Let us consider $\bar{Q}e$, $\bar{Q}t$ and $\bar{Q}s$ the mean values of edge, texture and smooth scores, respectively. The equation (4) could be rewritten as follows:

$$Q = \alpha\ \bar{Q}e\ + \beta\ \bar{Q}t\ + \gamma\ \bar{Q}s \tag{5}$$

4 Experimental results

Most of the inpainting algorithms and inpainting quality measures are evaluated on local datasets, created or collected by the authors. The first image inpainting dataset is Technische Universitt München Image Inpainting Database TUM-IID [14]. All experiments in this

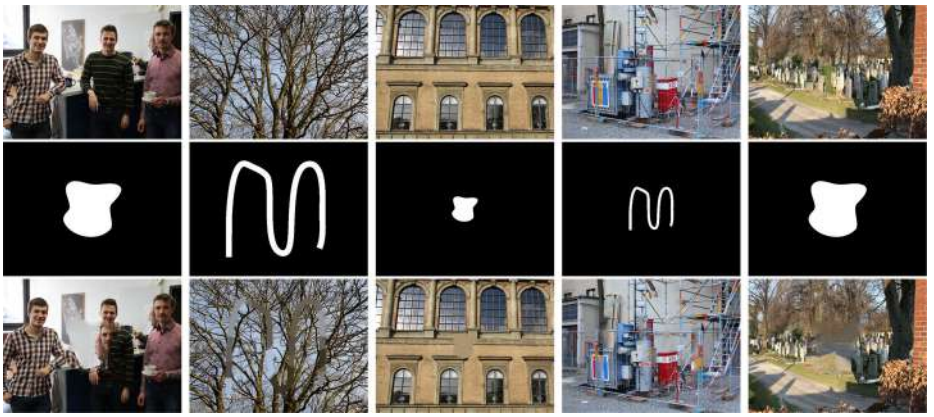


Fig. 10 Some images from the TUM-IID [14] database. Top: original images. Middle: the used masks for inpainting(two compacted and two spread-out). Bottom: inpainted images using different algorithms

paper are conducted in this dataset. In this section, we first give a brief description of the TUM-IID dataset. Then, we discuss the efficiency of the proposed metric in two parts:

- Performance evaluation: we present quantitative results, comparing the scores of the proposed objective metric with the subjective scores provided by the dataset, and qualitative results by showing some examples inpainted images assessment.
- Comparative study: we compare the proposed metric with most of the related works.

4.1 IIQA dataset description

TUM-IID database contains 17 original intact images, variate in terms of texture and structure. Each of these images is inpainted with four inpainting algorithms, and by using four masks(two compacted and two spread-out masks).

The inpainting algorithms used in this database are those of Bugeau [2], of Herling [7], that based on principle of Total Variation [4] and of Xu [18]. In total 272 inpainted images are provided to do a comprehensive comparison. In addition, this database provides also the psycho-physical subjective experimental data by providing Difference Mean Opinion Scores DMOS. Figure 10 shows some images and masks taken from the TUM-IID dataset.

4.2 Results and discussion

4.2.1 Performance evaluation

To evaluate the performance of the proposed method, we have

evaluated its correlation with human subjective scores DMOS, using Spearman Rank Order Correlation Coefficient SROCC measure. Figure 11 shows clearly that the B-IIQA correlates well with the DMOS quality. Moreover, the correlation of the proposed IIQA with DMOS depends on the mask of the inpainting process and is greater than 0.8.

Furthermore, the proposed metric allows us to objectively compare the inpainting algorithms results. Figures 12, 13, 14 and Table 1 present examples of comparison between the results of four inpainting methods. The considered images, are characterized by different

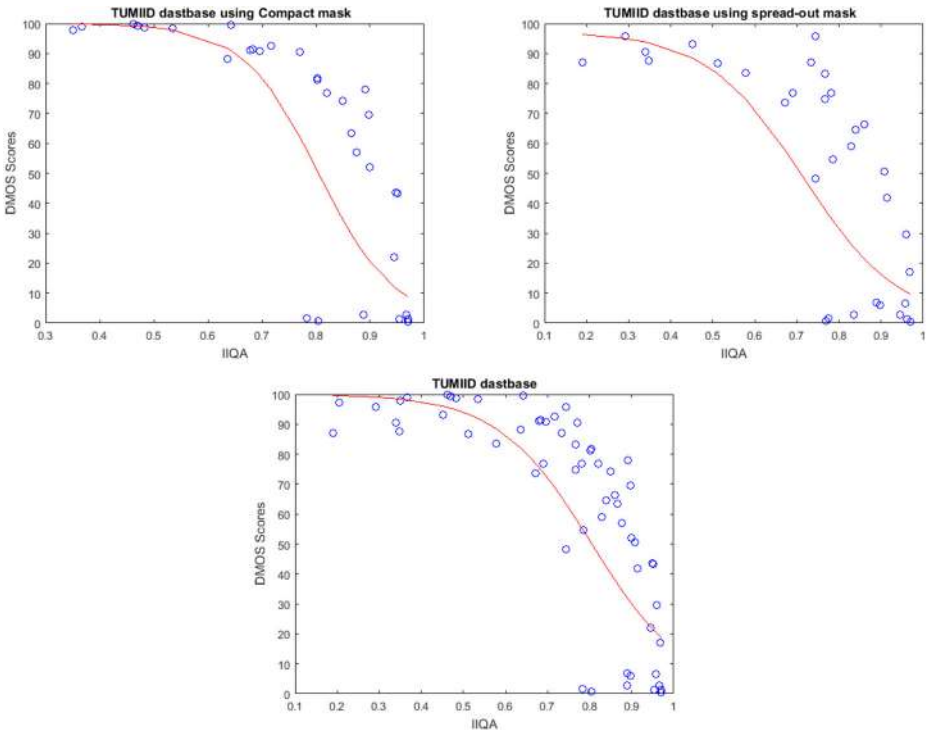


Fig. 11 Regression plots between DMOS and predicted score B-IIQA for compacted masks, spread-out masks and for all TUM-IID masks

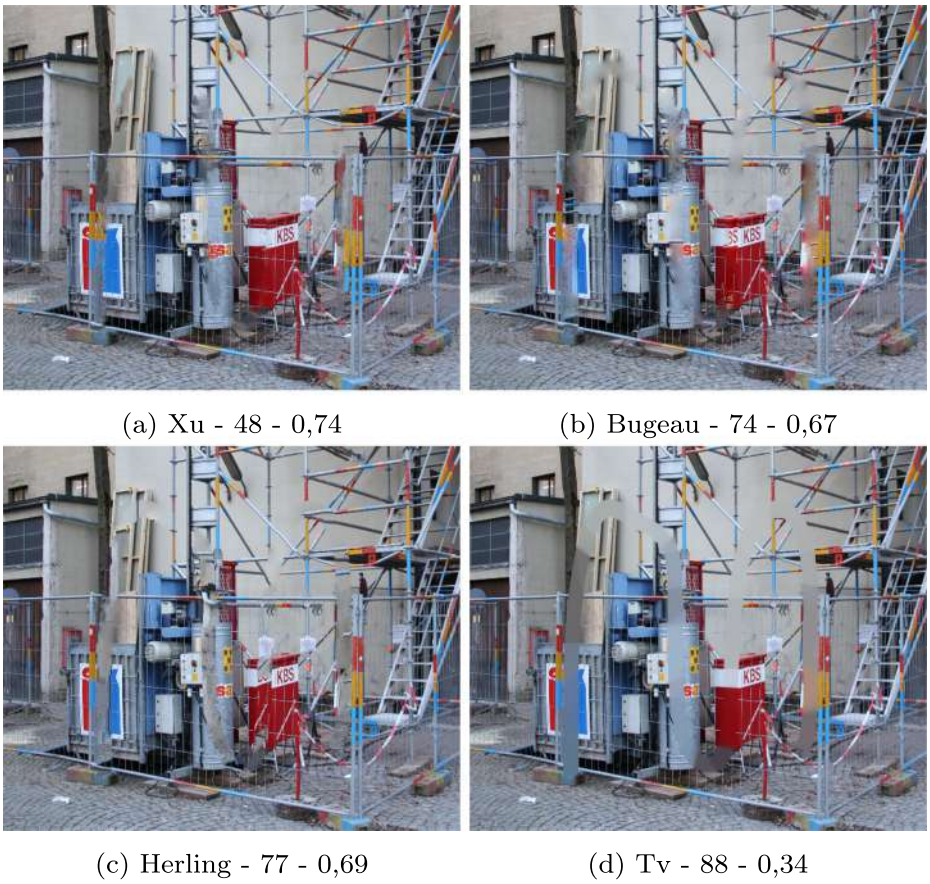
types of textures: structural (image 1 of Fig. 12) , stochastic (image 2 of Fig. 13), presence of a well-structured object (image 3 of Fig. 14).

The algorithm proposed by Xu and al [18] reproduces the best structures as well as, it is difficult to perceive or discriminate the inpainted areas. In the case of Bugeau [2] and Herling [7] algorithms, very close results have been obtained, where some structures have been correctly reproduced. The last algorithm TV [4] almost made no changes to the corrupted image.

These results have been correctly quantified by the proposed quality measure. Since the first method had the best score, and the two following methods have a very close score, and the lowest score was awarded to TV method. Thus, the proposed method indicates that the ability to reconstruct the structured areas is correctly taken into account.

Also for the stochastic texture areas presented by Fig. 13 we can see clearly that Herling's method can reproduce effectively the texture of the missing part. Visually, we can classify the remaining methods as follows: Bugeau, Xu, Tv, when the mask trace is partially visible to well visible. This state of affairs is clearly expressed by the quality proposed. Where the results are also classified in the same order. Therefore, the proposed BIIQA score indicates that the ability to construct textured areas is taken into account.

Finally, the third example could represent the most complicated case for inpainting , where the structures and textures are reconstructed with respect to the spatial continuity but do not respect the human perceptual coherence and logical form that should have. It is worth noticing that all images in Fig. 14 have a bad DMOS score. The proposed metric



(a) Xu - 48 - 0,74 (b) Bugeau - 74 - 0,67

(c) Herling - 77 - 0,69 (d) Tv - 88 - 0,34

Fig. 12 Example of image with structural texture, inpainted with four different methods. The inpainting method, the DMOS score, and BIIQA score are mentioned, respectively below each image

has classified them in terms of local spatial continuity. Besides, the coherence of the well-scored image Fig. 14a is more acceptable than the other inpainted images. The proposed method provides a bad score for images that remain unchanged or in the case of the hole is replaced by a smooth area.

For the third case, we can say that the proposed metric fails in the evaluation of inpainting when it is applied to images containing a high level of information. In other words, our metric cannot correctly assess the illogical reconstruction of shapes and forms, which require some intelligence. In order to improve the quality measurement results, it is necessary to introduce information on the consistency of forms or structures, which can be obtained through a learning-based system.

The obtained quantitative and qualitative results confirm that the proposed metric is effective in assessing the continuity of local structures in inpainted images. The distinction between edge, texture and smooth via patch classification, allowed us to treat and model each type of inpainting distortion differently, as well as to emphasize the contribution of edge discontinuity and texture inconsistency to the quality assessment, knowing that these two artifacts are more attractive to the human eye.

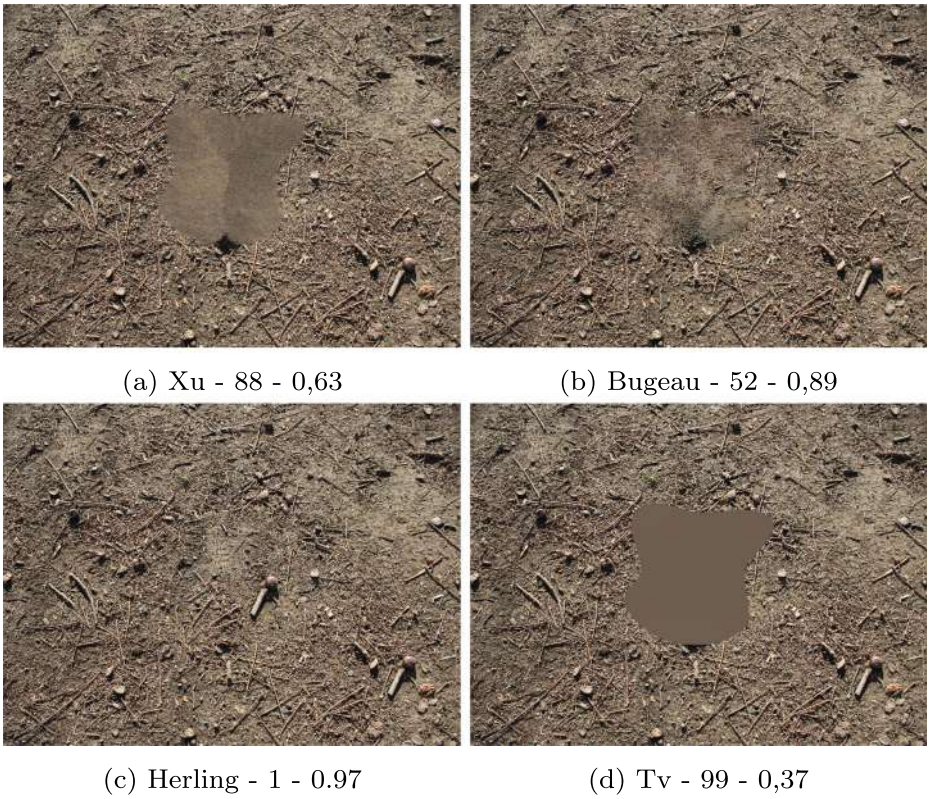
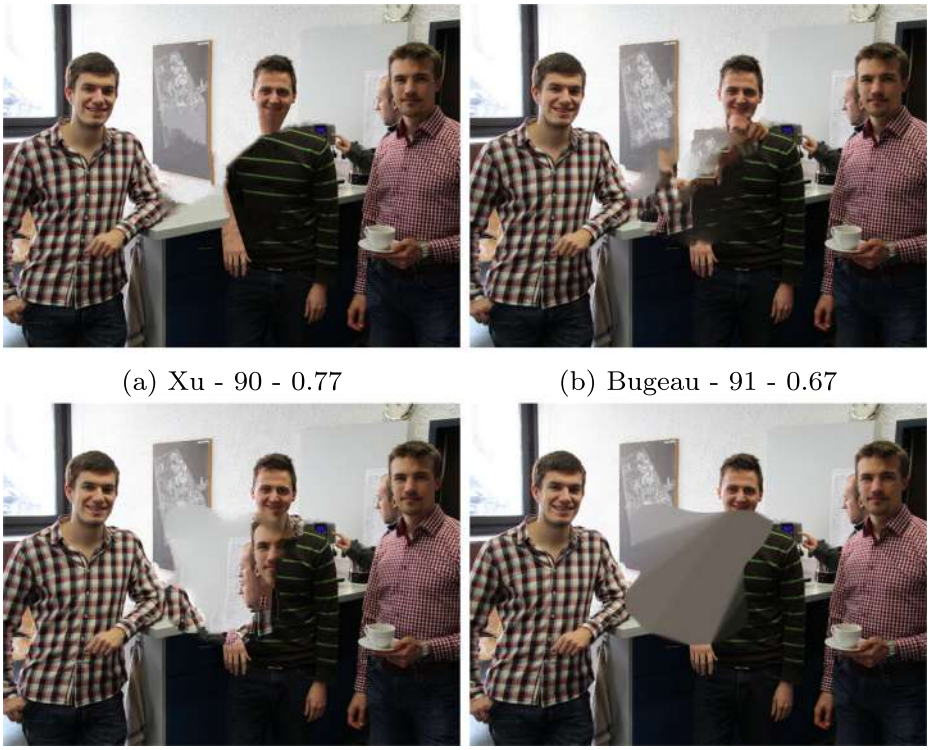


Fig. 13 Example of image with stochastic texture, inpainted with four different methods.The inpainting method, the DMOS score and BIIQA score are mentioned, respectively below each image

4.2.2 Comparative study

In order to compare our method to the related works, we have considered the existing methods in the literature except [10] and [16] which are machine learning-based methods. Indeed, the comparison against these methods is practically complex, since they use locally created datasets and their model is trained and tested using collected subjective scores, and neither the model nor the used data are available. Moreover, the method [10] ranks the inpainted images from the best to the worst, without giving any quality score, which makes the quantitative comparison meaningless, in this case.

All considered methods were evaluated by applying them on the same public database TUM-IID. Then, the obtained SROCC values are compared with those obtained for our proposed method. Table 2 summarizes the obtained results of the all considered approaches applied to TUM-IID database. It is worth noticing that the proposed method outperforms the No-reference method VISCOM, and is close to the best FR method GD-in for a compacted mask. Moreover, the proposed B-IIQA approach performs well for spread-out masks compared to all FR and NR approaches.



(a) Xu - 90 - 0.77 (b) Bugeau - 91 - 0.67

(c) Herling - 99 - 0.64 (d) Tv - 99 - 0.46

Fig. 14 Example of image with structured objects, inpainted with four different methods.The inpainting method, the DMOS score and BIIQA score are mentioned, respectively below each image

Table 1 B-IIQA and DMOS scores of original and inpainted images

			Inpainted images			
		Original image	Xu [18]	Bugeau [2]	Herling [7]	TV [4]
image1	DMOS	0,66	48,31	73,69	76,79	87,63
	B-IIQA	0,8	0,74	0,67	0,69	0,34
image2	DMOS	1,58	90,5	91	99,49	99,87
	B-IIQA	0,78	0,77	0,68	0,64	0,46
image3	DMOS	0,39	88,12	52,11	1,23	98,93
	B-IIQA	0,97	0,63	0,89	0,97	0,37

Table 2 Image-wise correlation performance of IIQA metrics

reference	IIQA metrics	SROOC		
		C-Mask	S-mask	Mean
	ASVS	-0,914	-0,886	-0,90
	DN	-0,91	-0,857	-0,883
	GDout	0,14	0,8	0,47
FR	PWIIQ	0,48	0,854	0,669
	StrucBorSal	0,74	0,74	0,74
	BorSal	0,71	0,8	0,755
	GDin	0,885	0,8	0,842
	VisCOM	0,68	0,714	0,697
NR	**B-IIQA**	**0,872**	**0,809**	**0,840**

5 Conclusion

Image inpainting quality assessment is a complex problem, and is different from the ordinary image quality assessment, because of the specificity of the artifacts introduced by the inpainting process and the absence of reference image. Most of the proposed methods in the literature are full reference. In this work, we have proposed a new blind image quality assessment, based on block classification by content, and structure continuity in the frontier of the used mask in the inpainting process. The proposed method is based on the summation of local inpainted patch quality assessment. The local scores depend on how the structures of patches have been continued from outside to the inside the hole delimited by the mask. In this aim, we have then defined the structure continuity for three types of structures: smooth, edges and textured patches, and the local associated scores have been provided. As the proposed algorithm depends on the classification accuracy of patches, we have then modified the existing algorithm to improve its performance.

By testing the proposed method on the public inpainting database TUM-IID that provides the mean opinion score which allows us to prove the correlation with the human perception and to compare our results against the existing IIQA methods, we have shown that our method outperforms the state-of-the-art. That is why the proposed B-IIQA can be used to control the iterative inpainting algorithms. Indeed, in such algorithm, each inpainted block could be chosen depending to its contribution to the global IIQA score. In the same way, the NR-IIQA could be computed at each iteration to evaluate if the best quality of the image is reached or not.

References

1. Ardis PA, Singhal A (2009) Visual salience metrics for image inpainting. In: Rabbani M, Stevenson RL (eds) Proceedings of SPIE 7257, Visual Communications and Image Processing, San Jose, CA, p 72571W. https://doi.org/10.1117/12.808942
2. Bugeau A, Bertalmio M, Caselles V, Sapiro G (2010) A comprehensive framework for image inpainting. IEEE Trans Image Process 19(10):2634–2645

3. Dang TT, Beghdadi A, Larabi M-C (2013) Metric for visual coherence evaluation of color image restoration. In: 2013 colour and visual computing symposium (CVCS). IEEE, Norway, pp 1–6. https://doi.org/10.1109/CVCS.2013.6626268
4. Getreuer P (2012) Total variation inpainting using split bregman. Image Processing On Line 2:147–157
5. Gong H, Hang H (1994) Scene Analysis for DCT Image Coding, Signal processing of HDTV. Elsevier, Amsterdam, pp 425–434. https://doi.org/10.1016/B978-0-444-81844-7.50052-8
6. Guillemot C, Le Meur O (2014) Image Inpainting : Overview and Recent Advances. IEEE Signal Processing Magazine 31(1):127–144. https://doi.org/10.1109/MSP.2013.2273004
7. Herling J, Broll W (2012) Pixmix: A real-time approach to high-quality diminished reality in International Symposium on Mixed and Augmented Reality. IEEE, Piscataway, pp 141–150
8. Hu W, Ye Y, Zeng F, Meng J (2018) A new method of thangka image inpainting quality assessment, J. Vis. Commun image r. https://doi.org/10.1016/j.jvcir.2018.12.045
9. Isogawa M, Mikami D, Takahashi K, Kojima A (2016) Eye gaze analysis and learning-to-rank to obtain the most preferred result in image inpainting. In: 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, pp 3538–3542. https://doi.org/10.1109/ICIP.2016.7533018
10. Isogawa M, Mikami D, Takahashi K, et al. (2019) Image quality assessment for inpainted images via learning to rank. Multimed Tools Appl 78:1399–1418. https://doi.org/10.1007/s11042-018-6186-z
11. Mahalingam VV, Cheung S-CS (2010) Eye tracking based perceptual image inpainting quality analysis, IEEE, Hong Kong. https://doi.org/10.1109/ICIP.2010.5653640
12. Oncu AI, Deger F, Hardeberg JY (2012) Evaluation of digital inpainting quality in the context of artwork restoration. In: Proceeding of the 12th International Conference on Computer Vision, pp 561–570. https://doi.org/10.1007/978-3-642-33863-2_58
13. Qureshi M, Deriche M, Beghdadi A, Amin A (2017,) A critical survey of state-of-the-art image inpainting quality assessment metrics. Journal of Visual Communication and Image Representation. 49. https://doi.org/10.1016/j.jvcir.2017.09.006
14. Tiefenbacher P, Bogischef V, Merget D, Rigoll G (2015) Subjective and objective evaluation of image inpainting quality, IEEE, Canada
15. Viacheslav V, Vladimir F, Vladimir M, Nikolay G, Roman S, Valentin F (2014) Low-level features for inpainting quality assessment. In: 2014 12th International Conference on Signal Processing (ICSP) Hangzhou, pp 643–647. https://doi.org/10.1109/ICOSP.2014.7015082
16. Voronin VV, Vladimir A, Frantc VI, Marchuk AI, Sherstobitov K (2015) Egiazarian No-reference visual quality assessment for image inpainting SPIE, 9399 Image Processing: Algorithms and Systems XIII. https://doi.org/10.1117/12.2076507
17. Wang S, Li H, Zhu X, Li P (2008) An evaluation index based on parameter weight for image inpainting quality, IEEE, Hunan. https://doi.org/10.1109/ICYCS.2008.461
18. Xu Z, Sun J (2010) Image inpainting by patch propagation using patch sparsity. IEEE Trans Image Process 19(5):1153–1165