ASSESSING THE EFFECTIVENESS OF INPAINTING TECHNIQUES FOR ENHANCING FEATURE EXTRACTION QUALITY IN REMOTE SENSING IMAGERY

C. F. M. Fontoura Júnior [1,], G. P. Cardim [2], E. S. Nascimento [1], M. Colnago [3], W. C. de O. Casaca [4], E. A. da Silva [1]

[1] Faculty of Science and Technology, São Paulo State University - Unesp, Presidente Prudente Campus, Street Roberto Símonsen, 305, Pres. Prudente - SP, 19060-900, Brazil - caio.fontoura@unesp.br; erivaldo.silva@unesp.br; e.nascimento@unesp.br.
[2] Department of Engineering, Physics and Mathematics, São Paulo State University – Unesp, Araraquara Campus, Av. Prof. Francisco Degni, 55, Jardim Quitandinha, Araraquara - SP, 14800-900, Brazil - marilaine.colnago@unesp.br.
[3] Faculty of Engineering and Sciences - Department of Engineering, São Paulo State University – Unesp, Experimental Rosana Campus, Av. dos Barrageiros, 1881, Primavera - SP, 19274-000, Brazil - guilherme.cardim@unesp.br.
[4] Institute of Biosciences Letters and Exact Sciences, São Paulo State University – Unesp, São José do Rio Preto Campus, Street Cristóvão Colombo, 2265, São José dos Campos, 15045-000, Brazil - wallace.casaca@unesp.br.

ABSTRACT:

Remote Sensing (RS) images have been used in several applications of interest for society. Despite the precision and robustness derived from RS images, several aerial scenes exhibit imperfections and fall short of attaining ideal quality standards, as some of them present distortions such as noise, blur, and stripes. An alternative approach to deal with such distortions is by applying Inpainting techniques, however, under certain circumstances, this type of approach requires to be evaluated by quantitative metrics to assess the final quality of the reconstruction. Therefore, this paper focus on the issue of quantitatively evaluating inpainting results in the context of RS by analysing and comparing new evaluation metrics in contrast to the classical ones from the general literature of RS. More precisely, two inpainting techniques are applied for object removal and reconstruction of partially detected curvilinear cartographic features in RS images. Next, the obtained results are evaluated by taking six evaluation metrics to assess the agreement level between the metrics, as well as between qualitative evaluations conducted by human agents. Based on the evaluation of these metrics when applied to RS images, it can be concluded that the DISTS and VSI metrics are the most promising candidates for adaptation and application within the specific context of RS.

1. INTRODUCTION

Remote sensing (RS) images have been widely used in various areas that are critically important to society, such as ecological monitoring, agriculture, and urban planning (Rubel et al., 2022). However, RS scenes may exhibit imperfections, especially due to the process of capturing complex patterns from the scanned area, thus generating distortions such as noise, blur, scratches, stripes and other corrupted pixels. One way to address these distortions is by applying image inpainting techniques, as it can be successfully used to reconstruct these types of distortions, as well as removing objects such as clouds, shadows, and texts (Ieremeiev et al., 2020; Qureshi et al., 2017).

In the field of Cartography, the process of feature extraction has been crucial in various real-world applications. This task aims at identifying targets present on the earth's surface, such as roads, hydrography, airport runways, as well as for updating cartographic products. Due to anthropic and natural alterations of the earth's surface, it is of utmost importance to identify these modifications for tracking and vigilance purposes. In this context, among the existing classical digital image processing (DIP) techniques, there are several feature extraction methods that can detect partial or corrupted features so that the quality and of the results can be properly measured (Figueira et al., 2018).

Inpainting techniques aim to increase the visual quality of the image, by creating more detailed structures that, due to the flaws or absence of features information, were not captured before the inpainting reconstruction process. Despite the good results generated by the inpainting techniques, this type of task requires to be assessed by means of quantitative metrics to ascertain the final quality of the restoration (Azevedo, 2019; Basso et al., 2021). Qureshi et al. (2017) state that the issue of assessing the quality of images obtained via inpainting techniques remains a complex and challenging problem.

Despite the investigations carried out in this field, only a limited number of studies have proposed quantitative metrics to assess the quality of inpainting, particularly in the domain of Cartography and Remote Sensing. In fact, most of the published studies on validation metrics are strictly related to the development or use of inpainting techniques in conventional photographic digital images. In this sense, it is important to emphasize that the application of metrics to quantitatively evaluate the quality of partially detected curvilinear features via inpainting has been a field that is still little explored in the RS context.

The use of inpainting techniques has become increasingly frequent, and researchers have developed various inpainting methods according to specific needs. However, only by directly applying these techniques, it is not possible to quantitatively identify the changes of feature pixels in the images. Thus, specific metrics have been created or adapted to assess the quality of digital images processed after the elimination of the objects that interfere with the detection of features of interest.
The numerical evaluation of high and medium spatial resolution RS images allows for the quantitative validation of image reconstruction quality, and consequently, the establishment of a quantitative criterion for the quality of cartographic feature restorations. Thus, the use of inpainting quality metrics in the context of SR can lead to more robust and reliable validation criteria, eliminating the issue of human subjectivity while also promoting automation via computers, providing more well-

established quantitative evaluation metrics that would take into account the specific characteristics of SR images. The main goal of this paper is to evaluate the use of different evaluation metrics for quantifying and assessing the results from the inpainting techniques when they are applied to the process of extracting cartographic features, focusing on partially detected curvilinear features in SR images. By doing so, one can determine which metrics are best suited to evaluate the results obtained from the application of inpainting techniques.

2. LITERATURE REVIEW

2.1 Inpainting Techniques

According to Li and Wen (2012), the task of digital inpainting has been inspired on ancient art in which people repaired cracks in works of arts during the Renaissance period in Europe. With the evolution of computers, the first digital image inpainting technique was introduced by Bertalmio et al. (2000) in the computational context, which consisted of filling in the missing or lost information in digital photographs based on the available information present in the images.

The technology of RS imagery is a widely discussed topic in the field of DIP, as RS images are used to drive several research areas such as agriculture, urban and territorial planning, earth sciences, meteorology, monitoring of environmental disasters, weather conditions, defense areas, and others. The information contained in aerial images provides more reliable attributes to explore large-scale problems such as change detection, land use and cover classification, spectral indices, and identification of different types of rocks and minerals (Lakshmanan and Gomathi, 2017). However, some elements in the image often present damaged or missing pixels (dead pixels) caused by the unwanted presence of clouds and shadows. In addition, some of these elements may be occluded, noisy, blurred, or stained. These types of distortions limit the execution of post-processing methods such as target classification and recognition (INCE, 2019). Therefore, the goal of the inpainting task is to perform the full reconstruction or restoration of the target image, recovering relevant features or removing unwanted objects from the image. Previously, the inpainting technique was widely used in conventional photographs or paintings to remove scratches, folds, objects, noise, or texts. Nowadays, it has been widely used in various digital products in order to obtain images with quality equivalent to the original images without any damages (Lakshmanan and Gomathi, 2017).

2.2 Evaluation Metrics

According to Sara, Akter, and Uddin (2019), there are objective and subjective ways to evaluate image quality. Subjective evaluation is usually more challenging and time-consuming compared to the objective one. In terms of objective assessment, these are image quality metrics that rely on different aspects of their development. In the past decade, various evaluation metrics have been developed to image quality assessment. According to Søgaard et al. (2016), those metrics can be classified into two ways: Full-Reference (FR) and No-Reference (NR). FR aims to evaluate image or video quality by comparing the distorted image with a reference image, which is usually taken as the original image without distortion and in optimal quality, while the NR holds the opposite case.

In recent decades, there has been significant growth in visual quality metrics, each with their particularities when applied on the general context of conventional digital images (Egiazarian et

al., 2018). In the RS context, the main validation metrics are the following: Mean Squared Error (MSE), the Peak Signal to Noise Ratio (PSNR), and the Structural Similarity (SSIM) (Basso et al., 2021). Although PSNR and MSE gauge the images in terms of their common content and the type of distortion, these metrics do not correlate well with subjective ratings (Huang and Jing, 2020; Zhang, Shen and Li, 2014). As a result, MSE, PSNR, and SSIM may not properly reflect the results in relation to the visual reality of RS images, thus requiring qualitative inspections together with quantitative examinations to propose metrics that better reflect the agreement between visual quality and numeric evaluation.

It is worth mentioning that there are several evaluation metrics that have been explored not in the context of RS but rather in the general application of digital image editing in an effort to assess the quality of results obtained by inpainting techniques (Qureshi et al., 2017). Table 1 lists three evaluation metrics that achieve satisfactory results in the general context of digital image: the Deep Image Structure and Texture Similarity (DISTS) index (Ding et al., 2020), the Visual Saliency-Induced (VSI) Index (Zhang, Shen and Li, 2014), and the Feature Similarity (FSIM) Index (Zhang et al., 2011). Classic evaluation metrics such as MSE, PSNR, and SSIM were also included.

Full-Reference Metric	Description
MSE	Classic pixel-based mean square error.
PSNR	Peak signal-to-noise ratio based on MSE.
SSIM (Wang, 2004)	Gauges the similarity structures between two images. It considers low-level features such as luminance, contrast, and structural information.
FSIM (Zhang et al., 2011)	Gauges the similarity structures between two images. It takes both low-level and high-level image features.
VSI (Zhang, Shen and Li, 2014)	Compares the saliency-induced regions of the reference and processed images, and then computes the similarity between them.
DISTS (Ding et al., 2020)	Takes a pre-trained convolutional neural network (CNN) devoted to predicting the perceptual similarity between two images.

Table 1. Description of main image quality evaluation metrics.

In recent studies, the PSNR, SSIM, and FSIM metrics have been increasingly employed in the RS context. Jiang et al. (2021) proposed an image dehazing method for RS imagery based on encoder-decoder architecture, which associates wavelet transformation and deep learning technology. Meanwhile, Huang and Jing (2020) developed an algorithm capable of reconstructing high-resolution images by combining wavelet transformation and generative adversarial network technology to improve high-frequency details in low-resolution images. Both studies used the above-listed metrics to evaluate the best results for the inpainting task.

Based on theoretical references, the VSI and DISTS metrics have not yet been applied to RS images, particularly, the literature has covered applications of these metrics in

conventional photographic images (Zhu et al., 2022) and biomedical images (Du et al., 2022). In (Zhu et al., 2022), the image quality evaluation was used to assess a new perspective of grouping in conventional photographic images. Among these metrics, the VSI was able to be located among the best results, which reflected the reality of the study. In (Du et al., 2022), a fusion method was developed to preserve high-intensity texture and color information. Thus, it was necessary to use the VSI metric for the quantitative evaluation of the images. The VSI metric also appears in works related to the creation of other metrics for validation, such as in (Huang et al, 2022). Finally, in (Lu et al., 2022), the VSI metric was employed to validate the creation of a low-light enhancement network with prior gradient assistance (GPANet). This network is responsible for extracting edge features and removing unwanted noise, by introducing Sobel Filter and Laplacian Filter features.

Although the DISTS is a recent metric, it has been successfully used in several applications. However, in the RS field, no related studies were found. Liu and Yeoh (2021) created an automated approach to recognizing concrete crack patterns in images, and for validation purposes, DISTIS was used to recognize crack patterns through similarity comparisons. The DISTS metric is also appeared as a validation tool in scientific works related to the generation of new metrics, such as Underwater Image Enhancement (UIF) (Zheng et al., 2022a) and Mean and Deviation of Deep and Local Similarity (MaD-DLS) (Sim et al., 2020). Finally, in (Zheng et al., 2022b), the DISTS metric was applied together with other evaluation metrics, such as FSIM, to evaluate the quality of image resolution enhancement artifacts.

2.2.1 MSE (Mean Square Error)

The MSE metric can be understood as the Mean Squared Deviation (MSD) of a statistical estimator. The MSE always presents positive values and when closer to zero, indicates better results. It is also considered a traditional metric that requires a reference image to be computed (Figueira et al., 2020; Sara, Akter and Uddin, 2019; Wafy and Ebaid, 2016). The MSE between two images is defined as follows:

$$MSE = \frac{1}{NxM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(z'(i,j) - z''(i,j)\right)^2 \quad , \tag{1}$$

where $z'(i, j)$: reference image
$z''(i, j)$: distorted image
M: number of rows
N: number of columns

Considering the process of synthesizing information in digital image, the goal of any inpainting or noise removal technique is to improve the fidelity and visual quality of a distorted image. Although the MSE quantifies the distortion between the reference image and the processed image, it does not consider certain relevant features of the target image such as texture and inherent patterns (Ndajah et al., 2010).

2.2.2 PSNR (Peak Signal to Noise Ratio)

The PSNR is a well-established metric that calculates the ratio between the maximum signal and the distortion noise values that affect the quality of the signal representation. In other words, it takes the total of gray levels in the image and its corresponding pixels from the reference image. The higher the PSNR value, the better the result, indicating that the target and reference images are similar. In order to compute this metric for practical purposes, one may calculate the MSE metric (Jagalingam and Hegde, 2015; Tiefenbacher et al., 2015; Sara; Akter and Uddin, 2019; Figueira et al., 2020). PSNR is calculated by using a logarithmic function, as signals have a very wide dynamic range. Additionally, PSNR measures the difference between pixel values individually. In general, the relationship between two images is calculated in decibels (db) (Tiefenbacher et al., 2015; Sara, Akter and Uddin, 2019). SNR can be expressed as (Rabbani; Jones; 2010):

$$PSNR = 10log_{10}\left(\frac{(peak\ value)^2}{MSE}\right), \tag{2}$$

where $peak\ value$ = The maximum among the pixels of the two images

The metrics PSNR and MSE have widely been used due to their simplicity of use and easy mathematical implementation. However, these metrics are not normalized in representation (Sara; Akter and Uddin, 2019).

2.2.3 SSIM (Structured Similarity Index Method)

The SSIM metric belongs to the FR group of methods and provides the normalized mean value of structural similarity between the distorted and the reference images (Sara, Akter, and Uddin, 2019). SSIM measures distortions by combining three factors: loss of correlation, luminance distortion, and contrast distortion (Ndajah et al., 2010).

Mathematically, SSIM can be expressed as (Wang et al., 2004):

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_1)}, \tag{3}$$

where μ: mean
σ: standard deviation
x: original imag
y: target image
σ_{xy}: covariance of x and y
C_1 and C_2: constants that prevent numerical instability

SSIM is an evaluation metric that predicts the quality of images when there is a need to measure the structural similarity between the input and reference images. SSIM is more accurate compared to PSNR and MSE (Figueira et al., 2020).

2.2.4 FSIM (Features Similarity Index Matrix)

The FSIM metric consists of two stages. The first one involves calculating the similarity map between images, while in the second stage, the similarity map is grouped, and then the similarity score is computed. FSIM is a very robust metric for validation tasks (Zhang et al., 2011). FSIM provides a normalized average value of the similarity of features between the original and distorted image, and their values range from 0 to 1, where a result closer to 1 indicates that the target and reference images are similar (Sara, Akter and Uddin, 2019). FSIM was originally designed for grayscale images, but since chrominance information also affects the Human Visual System (HVS) in image perception, the metric has been improved to incorporate chrominance information for color images, i.e., RGB images (Zhang et al., 2011). FSIMC/FSIM equation can be expressed as follows (Zhang et al., 2011):

$$FSIM_C = \frac{\sum_{X\epsilon\Omega}S_L(X)\cdot[S_C(X)]^\lambda\cdot PC_m(X)}{\sum_{X\epsilon\Omega}PC_m(X)} \quad , \tag{4}$$

where   λ: adjust the chromatic components (λ > 0)
Ω: represents the whole image spatial domain
PC_m (X): gauges the importance of $S_L(X)$ in the overall similarity between two images
S_L (X): similarity computed at the location x
S_C (X): chrominance similarity measure

2.2.5 VSI (Visual Saliency-Induced Index)

The VSI quality metric gauges the preservation of visual saliency after image processing, comparing the similarity between the distorted and the original images. VSI assumes that HSV is sensitive to prominent features in the image such as edges and contrasts, and that preserving these features is essential to maintaining the visual quality of the image.

Visual saliency (VS) has been used in various applications over the past decades, such as neurobiology and computer science. VSI takes VS as a resource to calculate the quality map of the processed image. Subsequently, the quality score is gathered, and VS is applied as a weighting function to reflect the importance of a particular region. Furthermore, VSI is a low-complexity metric and presents satisfactory results compared to other metrics used in the validation process for conventional images (Zhang, Shen and Li, 2014).

According to Kumar, Bhandari, and Kumar (2022), the closer the VSI value is to 1, the better the output of the VS between the reference and distorted images. In other words, the higher the VSI value, the better the result, and the lower the distortion of visual saliency. VSI equation can be expressed as follows (Zhang, Shen and Li, 2014):

$$VSI = \frac{\sum_{X \epsilon \Omega} S\,(X) \cdot VS_m\,(X)}{\sum_{X \epsilon \Omega} VS_m\,(X)} \quad , \qquad (5)$$

where   Ω: represents the whole spatial domain
S (X): similarity computed at the location x
VS_m (X): gauges the importance of S(x) in the overall similarity

2.2.6 DISTS (Deep Image Structure and Texture Similarity index)

The DISTS metric allows the computation of structural distortions (artifacts due to noise or blur) with a tolerance for texture resampling, where a texture region is transformed into a new sample wherein the pixels are different but, in terms of visual perspective, the texture is identical (Liu and Yeoh, 2021).

DISTS takes a convolutional neural network (CNN) to transform the reference and distorted images into new representations. Then, a set of measures is created to capture the visual appearance of the image textures. Finally, the texture parameters and global structural measures are combined to form an image quality evaluation (Ding et al., 2020). In more technical terms, DISTS relies on a deep neural network that is trained to learn image quality. Thus, a pre-trained convolutional neural network (CNN) is trained to predict the perceptual similarity between the reference and distorted images.

According to Ding et al. (2020), DISTS is a metric that is robust to mild geometric distortions, and it performs satisfactorily in texture classification and retrieval. It varies from 0 to 1, as 0 indicates that the distorted image resembles the reference image. DISTS equation can be expressed as (Ding et al., 2020):

$$D(x, y; \alpha, \beta) = 1 - \sum_{i=0}^{m}\sum_{j=0}^{ni}\left(\alpha_{ij}\, l(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)}) + \beta_{ij}\, s(\tilde{x}_j^{(i)}, \tilde{y}_j^{(i)})\right) \quad , (6)$$

where   {α_{ij}, β_{ij}}: positive learnable weights
l($\cdot$): texture similarity function
s($\cdot$): structure similarity function

3. METHODOLOGY

3.1 Database

To drive our analysis, a cartographic database was used, which is composed of images from the IGC's aerophotogrammetric and cartographic collection with a spatial resolution of 45 cm, as shown in Figure 1. The original images have a frame of 10,000 x 10,000 pixels. To optimize processing time and take regions of interest in the images, the images were previously cropped. As a result, thousands of images were created with frames of approximately 300 x 300 pixels.

Figure 1. Images from the IGC aerial photogrammetric and cartographic collection.
Source: IGC (2010).

3.2 Application of Inpainting Techniques

Among the inpainting techniques, two of them where taken in this study: the algorithms proposed by Baixo (2022), and the one described by Azevedo (2019). The choice of both algorithms is justified because they are recent proposals formulated to address RS images. The automatic algorithm presented by Azevedo (2019) consists of three steps: in the first two steps, the goal is to obtain the shadowed regions of image. Pre-processed images are then used as input data, based on two properties that characterize shadows.

To enhance the shadows in images, a combination of top-hat transformation and closure is applied, along with an area parameter calculated using the shadow index NSDVI, which detects low spectral responses in these regions. This transformation is a morphological image processing tool that aims to recover relevant image structures through mathematical operations. In the third stage, masks of the detected shadows are used to guide the process of reconstructing shadow regions by using a variant of the inpainting method described in Casaca et al. (2014), which employs a useful mechanism of block-based pixel replication, whose the goal is to copy "blocks of pixels" that have the same characteristics as the pixels that need to be reconstructed. This is done by calculating the distance that embeds information about the structure (homogeneous regions and edges) of the image. Figure 2(a) shows an illustration of how this approach fills the image. Starting from the so-called region of dynamic sampling (H_L(p)), the pixel priority is determined by computing the most suitable block (H_m(qˆ)) to be

filled in the target region of interest (Ω) within a specific neighborhood (*p*). For this purpose, a similarity measure based on NRMSD (Normalized Root Mean-Square Deviation), (Equation 7) (Azevedo, 2019) was applied.

$$d(p,q) = \frac{\|p-q\|_{\Delta U}}{\sqrt{\|p\|_{\Delta U}^2 + \|q\|_{\Delta U}^2}} \quad , \tag{7}$$

on what,

$$\|p-q\|_{\Delta U} = \sqrt{(p-q)^2 \Delta U(p-q)}, \ \|p\|_{\Delta U} = \sqrt{p^T \Delta U p}, \ \|q\|_{\Delta U} = q^T \Delta U q, \tag{8}$$

where   ΔU = diagonal matrix determined by the Laplacian
            p = specific neighborhood

The term ΔU is the diagonal matrix determined by the Laplacian of u: $\Delta U_{ii} = \Delta U_{pi}, p_i \in H_n(p) \cap \Lambda \Omega_p$. Within the sampling region (ΛΩ_p) (Figure 2(b)), the NRMSD measure compares the fixed block ($H_n(p)$) with all possible candidate blocks ($H_n(q\hat{})$). The block that minimizes the NRMSD distance between ($H_n(p)$) and ($H_n (q\hat{})$) will be the best block ($H_m(q\hat{})$), for all $H_n(q) \cap \Lambda \Omega_p$. From the selected block, a small region ($H_m(q\hat{})$) is taken to reconstruct the missing region in the neighborhood H(p) of *p* (Figure 2(c)) (Azevedo, 2019).
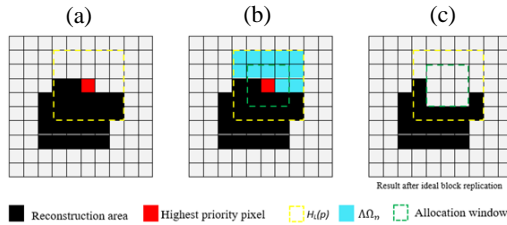

Figure 2. Block filling mechanism through dynamic sampling. Source: Adapted from Azevedo (2019).

The approach proposed by Baixo (2022) was taken for the inpainting application due to the fact that it is an improved method regarding the popular Total Variation (TV) model (Shen and Chan, 2002), which was enhanced by Schonlieb (2015). Baixo (2020) adapted the TV approach formulated by Schonlieb (2015) for recovering missing parts in RS images, including different RS bands. In short mathematical terms, Equation (9) – the TV-RS inpainting model –, summarizes the inpainting process as applied on (Baixo, 2022):

$$u_t = |\nabla u| \ div \left(\frac{\nabla u}{|\nabla u|} \right),$$
$$u \ |\delta \Omega = u^0| \ \delta \Omega \tag{9}$$

where   *div(f)*: represents the divergence operator
            δΩ: the restoration region

In Equation (9), *div(f)* represents the divergence operator, and the boundary condition requires that the output image, *u=u(x,y,t)*, should be equal to the initial image with respect to the information contained in the boundary of the restoration region in δΩ (Baixo, 2022; Schonlieb, 2015).

From the above-described Partial Differential Equation, the TV-RS model is derived, by numerically estimating the solution from Equation (9). In order to do so, it is necessary to discretize the Ω region (inpainting domain). Next, the partial derivatives are discretized via the Finite Differences Method (FDM)

(Baixo, 2022). As a result, the term *u_t* is numerically approximated via FDM, by using Equation (10):

$$u_t = \frac{\partial u}{\partial t} = \frac{u(t+\Delta t) - u(t)}{\Delta t}, \tag{10}$$

where   Δt: temporal step

After simplification of Equation (10), the following expression is reached, which introduces a recursive process that converts the damaged image (u^n) into an improved image ($u^{(n+1)}$):

$$u^{n+1} \cong u^n + \Delta t u_t^n, \tag{11}$$

Finally, for the numerical implementation of TV-RS restoration model, one may apply the following discretized expressions:

$$u_t^n = |\nabla u^n| \ div \left(\frac{\nabla u^n}{|\nabla u^n|} \right), \tag{12}$$

$$u_t^n = \frac{u_{xx}^n (u_y^n)^2 + u_{yy}^n (u_x^n)^2 - 2(u_x^n u_y^n u_{xy}^n)}{(u_x^n)^2 + (u_y^n)^2}, \tag{13}$$

From the discrete model in Equation (12), the inpainting can be computed efficiently (Schonlieb, 2015; Baixo, 2022). The input data are: the image to be restored, the restoration domain (the mask), the temporal step, and the maximum total number of iterations.

3.3 Validation Steps and Metrics Assessment

The restauration via inpainting techniques were assessed by taking six evaluation metrics applied to high spatial resolution RS images, creating a solid benchmark for metrics validation.

Two correlation coefficients were also used as evaluation criteria: the Spearman rank correlation coefficient (SRCC) and the Kendall rank correlation coefficient (KRCC). The SRCC was computed to verify the correlation between qualitative and quantitative analysis, i.e., the visual/subjective analysis of the metrics. Qualitative comparisons were performed for each evaluation metric: MSE, PSNR, SSIM, FSIM, VSI, and DISTS. To gauge the correlation between human agents from a qualitative point of view, KRCC was applied as part of our analysis. For the KRCC inspection, two users were invited to rank the images according to the best restoration from a subjective/visual point of view after inpainting was applied. Both users are doctoral students in cartographic sciences, but one is from the geodesy field and the another is from the remote sensing field. Our goal is to investigate whether existing quantitative metrics are compatible or not with qualitative visual results for basic features present in RS images.

4. RESULTS

The inpainting techniques from Azevedo (2019), Figures 3 and 4 (a)-(c), and Baixo (2022), Figures 3 and 4(d), were applied on a set of RS images with curvilinear features and different textures. In order to eliminate the issue of human subjectivity, all the six validation metrics were applied (see Table 2). From the the generated results, it was found that the classic metrics (MSE, PSNR, and SSIM) were generally unable to translate the obtained results in numerical terms for some of the sampled images, because they do not consider certain specific characteristics in digital images, such as texture. Table 2 presents the scores of the six evaluation metrics for each

reconstructed images, as well as the qualitative examination performed by a human agent, who ranked the results according to his subjective criteria. From Table 2 and Figure 3(a), best quantitative scores for MSE, PSNR, and SSIM were obtained, but visually, the inpainting presented a less satisfactory result. Therefore, the following question arises naturally: "What is the rationale behind the metric indicating that this particular outcome is the optimal one?" This can be justified by the fact that TV-RS inpainting technique smoothed the image significantly, discarding texture patterns. By only inspecting the evaluation metrics, these impose that the best results occur when the targets are excessively smoothed while for a human agent, the result may appear artificial.
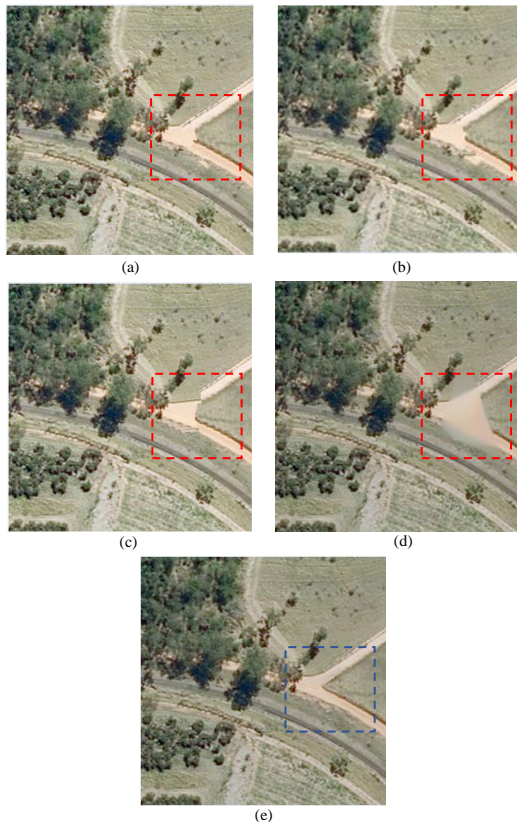
In fact, one can verify that in visual terms, the image was not reconstructed properly, as it reproduced a distortion/blur as part of the image. However, the FSIM, VSI, and DISTS metrics indicated coherent results compared to the subjective analysis.

Table 3 presents the numerical and qualitative results of a cartographic land use feature image. The DISTS and FSIM metrics coincide with the visual examination. However, the remaining metrics diverge between the first and second Figures 4(a) and 4(b) regarding the subjective analysis. From these results, the following question arise: "What makes a quantitative metric capable of indicating a more accurate result for one image than another?" Notice that the metrics proposed in the general inpainting literature were not designed to evaluate patterns and certain peculiarities of remotely sensed images.
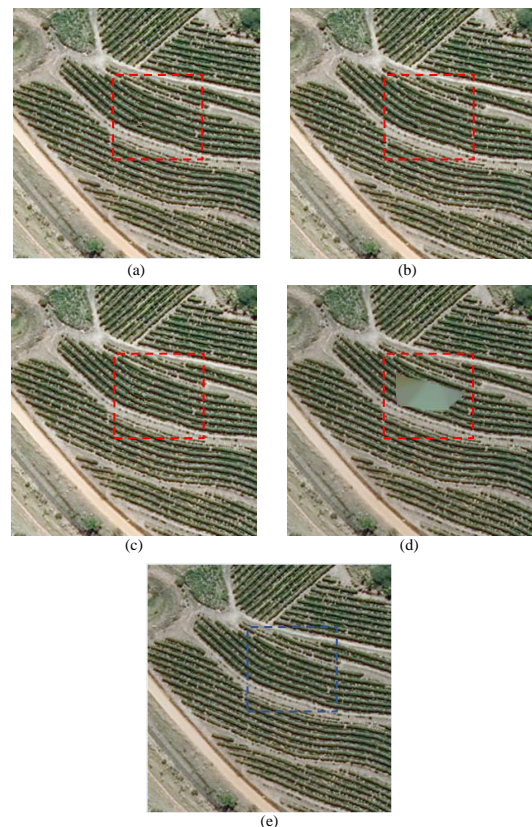

Figure 3. (a)-(c) Inpainting from Azevedo (2019), (d) from Baixo (2022), and (e) the reference image.

Figure 4. (a)-(c) Inpainting from Azevedo (2019), (d) from Baixo (2022), and (e) the reference image.

	Fig. 3(a)	Fig. 3(b)	Fig. 3(c)	Fig. 3(d)
MSE ↓	3	4	2	1
PSNR ↑	2	1	3	4
SSIM ↑	2	1	3	4
FSIM ↑	1	2	3	4
VSI ↑	1	2	3	4
DISTS ↓	1	2	3	4
Subjective score	**1**	**2**	**3**	**4**
MSE ↓	1056,7	1226,2	1007,8	995,4
PSNR ↓	17,9	17,2	18,1	18,2
SSIM ↑	0,9690	0,9684	0,9677	0,9766
FSIM ↑	0,9826	0,9807	0,9806	0,9785
VSI ↑	0,9943	0,9937	0,9935	0,9893
DISTS ↓	0,0365	0,0387	0,0457	0,0481

Table 2. Ranking of classified images according to their quality computed by each metric.

	Fig. 4(a)	Fig. 4(b)	Fig. 4(c)	Fig. 4(d)
MSE ↓	1	3	2	4
PSNR ↑	4	3	2	1
SSIM ↑	1	2	3	4
FSIM ↑	2	1	3	4
VSI ↑	1	2	3	4
DISTS ↓	2	1	3	4
Subjective score	**2**	**1**	**3**	**4**
MSE ↓	931,0	1132,7	1321,2	4923,3
PSNR ↑	18,4	17,6	16,9	11,2
SSIM ↑	0,9828	0,9820	0,9812	0,9754
FSIM ↑	0,9896	0,9900	0,9891	0,9810
VSI ↑	0,9952	0,9950	0,9948	0,9776
DISTS ↓	0,0245	0,0242	0,0247	0,3475

Table 3. Ranking of classified images according to their quality computed by each metric.

Concerning the visual and quantitative analysis, SRCC and KRCC were applied by considering the users' selections. User 1 had little experience in the area compared with User 2, causing their choices divergent from User 2, who has some knowledge about RS imagery (Table 4). Regarding Figure 3, when comparing the user's choices with the results of each quantitative metric, it was observed that almost none of the metrics were correlated with their choices, i.e., only SSIM gave a satisfactory correlation. On the other hand, almost all metrics demonstrated correlation with User 2's qualitative analysis, except for SSIM metric.

Metrics	Figure 3		Figure 4	
	User 1	User 2	User 1	User 2
MSE	0	-0.8	-0.4	-1
PSNR	0	0.8	0.4	1
SSIM	-1	-0.4	0.8	0
FSIM	-0.4	-1.0	-0.2	-0.8
VSI	-0.4	-1.0	-0.2	-0.8
DISTS	0.4	1.0	0.2	0.8

Table 4. Measures of performance (SRCC).

According to the KRCC, the value between users was 0.33, indicating weak correlation between their choices. It should be noted that a value closer to one indicates strong correlation. Thus, it was found that qualitative analysis is subjective among users themselves so that each user can evaluate the image and identify features that they believe to be more relevant in the reconstruction context, and some quantitative metrics cannot be properly correlated with visual analysis. Therefore, there is a necessity in adapting or creating new evaluation metrics that is capable of taking into account RS image features as evaluation criteria. Regarding Figure 4, the behavior was similar to the one observed in Figure 3, i.e., weak correlation for User 1 and strong correlation for User 2. According to the KRCC, the value between users was 0.67, indicating moderate correlation between their choices. It should be noted that in both cases, the DISTS metric achieved a strong correlation in both scenarios for User 2. DISTS reached attractive results, putting it as a strong candidate to be adapted to deal with RS images. On the other hand, the SSIM metric showed divergent results in relation to the other evaluation metrics in both analyses.

The most dependable metrics will hold significant importance, particularly in post-processing endeavors like image classification. This is due to the fact that while the reconstructed image may visually appear satisfactory, it may not align with the numerical reality. Additionally, since each user possesses varying levels of visual acuity and experience within this process, employing metrics that yield more consistent results between qualitative and quantitative analyses becomes necessary. In the course of reconstruction, the target may acquire information from adjacent pixels that do not belong to the same class, further emphasizing the need for images to exhibit concise and coherent visual as well as numerical information.

5. CONCLUSIONS

The main contribution of this work lies in the identification of more robust and appropriate metrics to be applied on the evaluation and assessment of inpainting quality for partially detected cartographic features. Based on the evaluation of the metrics applied to RS images, it was possible to measure and elect DISTS and VSI metrics as potential metrics to be applied

to the specific context of RS, including post-processing applications. Since our experimental analysis has been almost exclusively carried out in the context of photographic digital images, we believe that a more reliable treatment of RS data in Cartographic Sciences will have a broad impact on the use of inpainting metrics. In addition, we focus on maximizing the accuracy obtained during the evaluation task so as to obtain more precise results to drive the process of updating cartographic products. As future work, we plan to apply the analyses to a broader dataset and orbital satellite images.

REFERENCES

Azevedo, S. C., Oliveira, R. F. D., Casaca, W., Silva, E. A. 2019: Deshadowing of high spatial resolution imagery applied to urban area detection. *Boletim de Ciências Geodésicas*, 25. doi.10.1590/s1982-21702019000S00003

Baixo, F, P. 2022. Avaliação do modelo de Variação Total no retoque de feições cartográficas (Master in Cartographic Sciences), Presidente Prudente, SP.

Basso, D., Colnago, M., Azevedo, S., Silva, E., Pina, P., Casaca, W. 2021: Combining morphological filtering, anisotropic diffusion and block-based data replication for automatically detecting and recovering unscanned gaps in remote sensing images. *Earth Science Informatics*, 1-14. doi.org/10.1007/s12145-021-00613-6.

Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C. Image inpainting. 2000: In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 417-424. doi.org/10.1145/344779.344972.

Casaca, W., Boaventura, M., De almeida, M. P., Nonato, L. 2014: Combining anisotropic diffusion, transport equation and texture synthesis for inpainting textured images. *Pattern Recognition Letters*, 36, 36-45. doi.org/10.1016/j.patrec.2013.08.023.

Ding, K., Ma, K., Wang, S., Simoncelli, E. P. 2020: Image quality assessment: Unifying structure and texture similarity. IEEE transactions on pattern analysis and machine intelligence, 44(5), 2567-2581. doi.10.1109/TPAMI.2020.3045810.

Du, J., Li, W., Peng, Y., Zong, Q. 2022: Image Fusion by Multiple Features in the Propagated Filtering Domain. *Social Science Research Network.* dx.doi.org/10.2139/ssrn.4087606.

Egiazarian, K., Ponomarenko, M., Lukin, V., Ieremeiev, O. 2018: Statistical evaluation of visual quality metrics for image denoising. *In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, *IEEE.* 6752-6756. doi.10.1109/ICASSP.2018.8462294.

Figueira, A. L. C., Da silva, E. A., Araújo, M., Dias, B. S. M. D. C., Cardim, G. P. 2020: Use of two inpainting Techniques to

Restore Partially detected Cartographic Features. *International Journal of Advanced Engineering Research and Science*, 7(4), 299-304, 2020. dx.doi.org/10.22161/ijaers.74.38.

Figueira, A. L. C.1, Cruz, B. S. M., Silva. E. A., Cardim, G. P., Wallace, C. O. C. 2018: Metodologia para Reconstrução de Feições Curvilíneas Parcialmente Detectadas após Processo de Extração de Feições Cartográficas em Imagens Digitais. *XVIII Simpósio Internacional de La Sociedad Latinoamericana Depercepción Remota y Sistemas de Información Espacial*, La habana, Cuba

Huang, H., Zeng, H., Hou, J., Chen, J., Zhu, J., Ma, K. K. 2022: A spatial and geometry feature-based quality assessment model for the light field images. *IEEE Transactions on Image Processing*, 31, 3765-3779. doi.10.1109/TIP.2022.3175619.

Huang, Z. X., Jing, C. W. 2020: Super-resolution reconstruction method of remote sensing image based on multi-feature fusion. *IEEE Access*, 8, 18764-18771. doi.10.1109/ACCESS.2020.2967804.

Ince, T. 2019: Hyperspectral Image Denoising Using Group Low-Rank and Spatial-Spectral Total Variation. *IEEE Access*, 7, 52095–52109. doi:10.1109/ACCESS.2019.2911864.

Jagalingam, P.; Hegde, A. V. 2015: A review of quality metrics for fused image. *Aquatic Procedia*, 4, 133-142. doi.org/10.1016/j.aqpro.2015.02.019.

Jiang, B., Chen, G., Wang, J., Ma, H., Wang, L., Wang, Y., Chen, X. 2021: Deep dehazing network for remote sensing image with non-uniform haze. *Remote Sensing*, 13(21), 1-23. doi.org/10.3390/rs13214443.

Ieremeiev, O., Lukin, V., Okarma, K., Egiazarian, K. 2020: Full-reference quality metric based on neural network to assess the visual quality of remote sensing images. *Remote Sensing*, 12(15), 2349. doi.org/10.3390/rs12152349.

Kumar, R., Bhandari, A. K., Kumar, M, 2022: Haze Elimination Model-Based Color Saturation Adjustment With Contrast Correction. *IEEE Transactions on Instrumentation and Measurement*, 71, 1-10. doi:10.1109/TIM.2022.3177725.

Lakshmanan, V., Gomathi, R. 2017: A survey on image completion techniques in remote sensing images. In: *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN). IEEE*, 1-6. doi:10.1109/ICSCN.2017.8085669.

Li, M., Wen, Y. 2012: A new image inpainting method based on TV model. *Physics Procedia*, 33, 712-717. doi.org/10.1016/j.phpro.2012.05.125.

Liu, Y., Yeoh, J. K. 2021: Automated crack pattern recognition from images for condition assessment of concrete structures. *Automation in Construction*, 128, 1-14. doi.org/10.1016/j.autcon.2021.103765.

Ndajah, P., Kikuchi, H., Yukawa, M., Watanabe, H., Muramatsu, S. 2010: SSIM image quality metric for denoised images. In: *Proc. 3rd WSEAS Int. Conf. on Visualization, Imaging and Simulation*. 53-58.

Qureshi, M. A., Deriche, M. D., 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, 177-191. doi.org/10.1016/j.jvcir.2017.09.006.

Rubel, A., Ieremeiev, O., Lukin, V., Fastowicz, J., Okarma, K. 2022. Combined No-Reference Image Quality Metrics for Visual Quality Assessment Optimized for Remote Sensing Images. *Applied Sciences*, 12(4), 1-19. doi.org/10.3390/app12041986.

Sara, U., Akter, M., Uddin, M. S. 2019: Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *Journal of Computer and Communications*, 7(3), 8-18. doi:10.4236/jcc.2019.73002.

Sim, K., Yang, J., Lu, W., Gao, X. 2020: MaD-DLS: mean and deviation of deep and local similarity for image quality assessment. *IEEE Transactions on Multimedia*, 23, 4037-4048. doi.10.1109/TMM.2020.3037482.

Schonlieb, C. B. 2015: *Partial Differential Equation Methods for Image Inpainting*. Cambridge University Press.

Søgaard, J., Krasula, L., Shahid, M., Temel, D., Brunnström, K., Razaak, M. 2016: Applicability of existing objective metrics of perceptual quality for adaptive video streaming. *Electronic Imaging*.

Shen, J., Chan, T. F. 2002: Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*, 62(3), 1019-1043. doi.org/10.1137/S0036139900368844.

Tiefenbacher, P., Bogischef, V., Merget, D., Rigoll, G. 2015: Subjective and objective evaluation of image inpainting quality. In: *2015 IEEE International Conference on Image Processing (ICIP). IEEE*, 447-451. doi:10.1109/ICIP.2015.7350838.

Wafy, M, Ebaid, R. 2016: Quality Assessment of Visual Cryptography Images. *Australian Journal of Basic and Applied Sciences*, 10(15), 215-222.

Zhang, L., Zhang, L., Mou, X., Zhang, D. 2011: FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8), 2378-2386. doi.10.1109/TIP.2011.2109730.

Zhang, L., Shen, Y.,Li, H. 2014: VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image processing*, 23(10), 4270-4281. doi:10.1109/TIP.2014.2346028.

Zheng, Y., Chen, W., Lin, R., Zhao, T., Le Callet, P, 2022a: UIF: An objective quality assessment for underwater image enhancement. *IEEE Transactions on Image Processing*, 31, 5456-5468. doi.10.1109/TIP.2022.3196815

Zheng, R., Jiang, X., Ma, Y., Wang, L, 2022b: A Comparison of Quality Assessment Metrics on Image Resolution Enhancement Artifacts. *In 2022 International Conference on Culture-Oriented Science and Technology (CoST), IEEE*. 200-204.

Zhu, C., Xu, J., Feng, D., Xie, R., Song, L. 2022: Edge-based video compression texture synthesis using generative adversarial network. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10), 7061-7076. doi.10.1109/TCSVT.2022.3169951.