Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 171 (2020) 1877–1886

Procedia Computer Science

www.elsevier.com/locate/procedia

Third International Conference on Computing and Network Communications (CoCoNet'19)

Denoising and Segmentation of Epigraphical Estampages by Multi Scale Template Matching and Connected Component Analysis

Preeti P[a], Anish Kasi[a,1], Manish Shetty M[a,1], Mamatha H.R[a]

[a] PES University,Bangalore560085,India

Abstract

Epigraphy refers to the study of Epigraphs, or inscriptions. Estampage or stamping, is a term that is commonly used in epigraphy in order to get an exact copy of an inscription. An estampage is typically derived by applying wet paper onto the rock face, over which any ink material is wiped. This process also results in the appearance of unwanted noise which arises due to the texture of the rock, human errors or a variety of different factors. In this paper, we propose CCD: Connected Component Denoising , a novel approach for denoising and character segmentation via connected component analysis and multi scale template matching. A dynamic threshold is obtained via histogram density projections to effectively remove noisy areas. A comparative study is then performed between the different approaches (MSTM , CC , Hybrid). Inspection of the results of the hybrid approach by the HVS(Human Visual System) shows that it outperforms all the previous techniques.

Keywords: Connected Components, Projection , Denoising, PSNR, SSIM, MSTM

1. Introduction

Epigraphy is the detailed analysis of inscriptions on stones, pillars, coins and other surfaces. These inscriptions are of great value to a country as it indicates and explains the cultures and the advent of literature. The processing and preservation of these scriptures in India is done by the Archeological Survey of India. The engravings of interest can be spotted even today on a multitude of temple walls such as the famous Ajanta and Ellora caves.

[1]Equal Contribution

1877-0509
10.1016/j.procs.2020.04.201

The information extracted from these scriptures can be extended to endorse historic events like existence of Indian dynasties. Thus, it is important to explore this area or field. However the tasks that make up this process are tedious and require a lot of precision. The mechanism used to preserve these artifacts is to create a rubber copy of the inscriptions on paper by applying ink and the physical paper copy is then deciphered by experts. This task is extremely difficult as the textures , troughs and crests of the stones create unexpected non uniform noise. Nowadays, a digital copy of the scriptures are created which aids in preservation and security of the physical document.

The inputs to the system are images of these inscriptions with heavy noise and the task at hand is to denoise , segment and recognize characters which in turn helps decipher the inscription. Ideally any OCR model expects high clarity in the image containing the text to achieve a high precision in the recognition process.

Thus clarifying noise and finding segmentable components of text in the image is the major portion of processing the input , ie: digitized estampage images , would have to go through.

2. Literature Survey

Connected components labeling scans an image and groups its pixels into components based on pixel connectivity, i.e. all pixels within a connected component have pixel intensity values similar to each other and are in some way connected. Once the scan is complete, each pixel is given a label or a value corresponding to the component it belongs to. Extracting and labeling of various disjoint and connected components in an image is central to many automated image analysis applications.

[1] Makes use of connected component profiles for detection of language. A process called meaningful component selection, which removes "relatively small and large elements" determined through threshold values determined experimentally, is employed for denoising . The authors of [2] make use of connected components to extract text lines in handwritten documents. Neighborhood identification is done using tuning parameters and computing average width and height of the characters. Any component lying within a bounding box is assigned the same label .[3] Proposes a block based connected component labeling algorithm based on three optimization strategies. A simplified scan mask is obtained by selecting only a subset of the pixels in a scan mask to reduce redundant memory accesses. The second strategy is employed to reduce the judgement required in the next scan. The final strategy creates two binary decision trees to improve performance by eliminating redundant neighboring operations .The authors of [4] discuss a multi scan labelling algorithm for improved performance. Makes use of repeated forward and backward scan in order to resolve neighborhood relationships. However, the entire process takes more than four scans of the image and thus, is not time efficient. [5] Improved the multi scan labeling algorithm by implementing a lookup table. The look up table eliminates superfluous memory accesses by processing label equivalences during each neighborhood operation. The results of this improved version of the labeling algorithm has been experimentally shown to reduce up to 46% of execution time.

The authors of [6] classify character images into the following classes- adhesive, broken, normal, nosing and disputed depending on the length and width of the image. Horizontal projections are used for images falling under the classes of broken and adhesive in order to determine the approximate edge of a single license plate character. In the case of [8], the authors employ a vertical projection to achieve the same. The authors of [9] project the CAPTCHA i.e, the image onto the X-axis in order to analyze the density distribution. The projection value along with the variance value of a component help in differentiating between clutt
er items and characters.

Histogram equalization is a technique which is frequently used for improving image quality. Smoothening and Continuous intensity allocation provide stronger averaging effects and reduce the effect of outliers [10]. The slope differences of a smoothed histogram density projections are used to calculate optimal thresholds [11][12].

There are various metrics for evaluation of image quality. Image quality evaluation methods are divided into objective and subjective methods [13]. Subjective methods are related to Human Visual perception and do not have any explicit criteria for judgement [14]. Objective methods are based on comparisons and make use of explicit numeric criteria for evaluation of image quality [15]. The most widely used and accepted metrics are PSNR and SSIM. SSIM was created by Wang et. Al [16], and is considered to have some correlation with the perception of quality by the Human Visual System (HVS).

3. Background Concepts

Understanding the various individual pieces of the complete algorithm is important to its working. The pipeline can be broken down into 3 major pieces :

- White Background Removal
- Average Character Area Finder
- Connected Component Analysis

In the first step there is a need for some form of a bounding box solution. Initially few simple morphological operations are performed to remove certain noise.After which contours have come of use in the process of finding the smallest box that encloses the image of interest. Contours are curves joining continuous points of same intensity or color. To reduce the use of memory in memorizing the various contours found the approximation of the contour is stored. Then the contour with the largest area is found and represented as 4 points of a rectangle.

The next piece is a supplementary to the last and final piece of the pipeline. To remove noise from the different connected components identified there must be a threshold area given to the algorithm, which defines on an average the area of a character. This threshold provides an estimate as to how big an expected component should be.To find this area horizontal and vertical projections of the image were used to approximate the spacing between lines and characters. An example of projections would be as below. Each pixel value in every row is added to give the horizontal projection and each column is added to give the vertical projection.

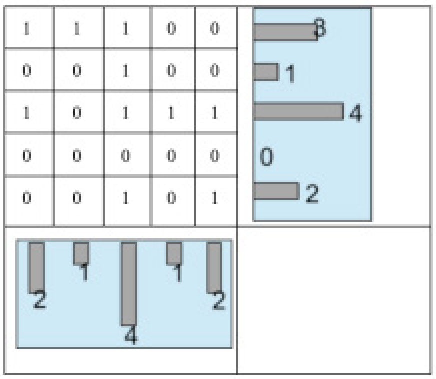

Fig 1 : Horizontal and Vertical projections of an image

$$\sum_{In\ row} pixelval\ () = Horizontal\ Projection \quad (1)$$
$$\sum_{In\ col} pixelval\ () = Horizontal\ Projection \quad (2)$$

The crests and troughs in the projections created gives us an idea of the spacing between characters. In the above example assuming text is represented by 1 and other parts of the image is 0 , we can see that every alternate row of pixels is a sentence of characters. Thus line spacing being 1. Similarly using vertical projection we can make an assumption that character spacing is also 1. This gives us on an average an area of character being 1x1. *8In the above step the projections tend to be erratic and random due to inherent noise. To find clear line separation or character separation the projections are smoothened , as shown below.

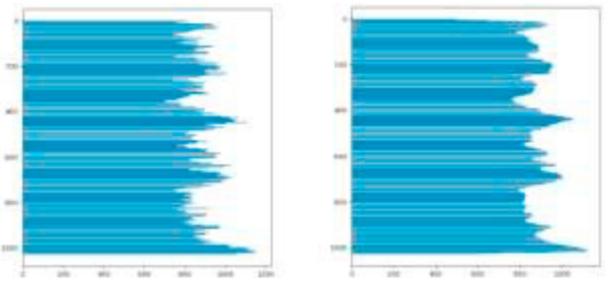

Fig 2 : Original Projection and Smoothened Projection

In the final part of our pipeline we find connected components of characters in the image. The approach to denoising documents followed here is not the usual one that removes noise from the subject of the image , but the other way round. That is the algorithm separates the subject from the noisy background. The idea of neighbourhood of a pixel is well known and has various meanings. But intuitively it means the pixel touching the pixel of interest , which can be a maximum of 8. Neighbourhoods can be defined more specifically like

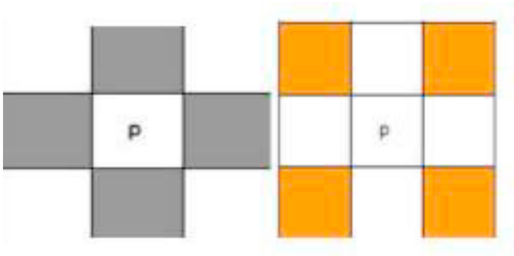

Fig 3 : 4 neighbourhood and d- neighbourhood of a pixel p

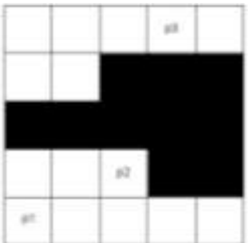

Fig 4 : Connectivity and connected components

Connected pixels are defined as a set of pixels that are belonging to each others neighbourhood forming a chain of connectivities. In Fig 3 , p1 is connected to p2 but neither p1 or p2 is connected to p3 due to the set of black pixels in between that remove connectivity.

This is expanded further to find areas of pixels that are connected to each other and define these as connected components of an image. An example of 2 components found in an image is given in Fig 5..

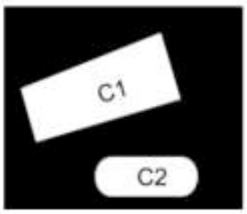

Fig 5 : 2 Connected Components in an image

4. Connected Components

4.1 *Union-Find Data Structure and Algorithm*

A disjoint-set , also known as Union-Find data structure is a data structure that keeps track of a set of elements partitioned into a collection of disjoint (non overlapping) subsets. The algorithm performs 2 operations :

1. Find : This operation determines which subset a particular element belongs to. Can be used to find some form of similarity between elements belonging to the same subset. It thus answers the question " Are two grid points connected ? ".
2. Union : Combines two individual subsets into a single subset. It thus acts as an operation that adds an edge between two points.

Now a connected component as defined earlier is set of pixels that are belonging to each others neighbourhood forming a chain of connectivities. For the sake of the algorithm we define a connected component simply as a set of mutually connected vertices.

We start off with each pixel or vertex being a unique component. An observation can now be made that applying one union operation reduces the no of components by 1. The algorithm in its simplest form, runs 2 passes on the image. The first pass in the classical algorithm makes use of dynamic programming for labeling. It goes through each pixel and looks at the labels that have already been assigned to 2 neighbouring pixels , one top and one left of the pixel of interest. This creates 2 cases :

1. In the case that the pixels are background pixels , create a new label for both of them.
2. In the case that either one or both are not background pixels , leave the labels as it is.

Once this is done the smaller label number is assigned to the pixel of interest and then a check is made if the current pixel connects the 2 neighbours. If so the union operation is called and the smaller component is made a child of the larger one.

The second pass of the algorithm is just to clean up any mistakes made during the first. For example assigning multiple labels to the same component. It does this pass by trying to reach root pixel of that component by following the tree structure and then assigns the final label to all the pixels in the path taken. Thus the algorithm in the end returns a list of labels assigned to each pixel which represents the component that the pixel belongs to.

5. Proposed Method

5.1 *Dataset Source*

The Archeological Survey of India (ASI) has provided a set of 144 epigraphs for testing. Each of these images are in binary black and white format or grayscale format. Most of these images contain significant amounts of noise and preprocessing must be done in order to serve as an input image to the next stage (Segmentation). A sample input image has been shown below.

Fig 6: Sample image

5.2 *Our Method*

- The source image used is the output images of the MSTM algorithm. These images serve as ideal inputs since our previous approach effectively removed all small noise and also helped splitting large chunks of character components into smaller more useful connected components. The results section further explains as to why this ensemble technique is more efficacious than just connected component denoising.
- The source image is then projected on to the X-axis and Y-axis. A histogram relating the axis to the pixel density is plotted and an estimation of the average character size is obtained.
- The projections observed tends to be erratic in nature and thus to find clear line separation or character spacing , the projections are smoothened.

- A percentage of this obtained character area is then used as a threshold value for the connected component analysis. The threshold is an indicator for meaningful component preservation.
- 8 neighbor connected component labelling is performed and each pixel is assigned a respective value based on the component as it is located in. Connected component analysis is then performed to remove all components that fall below the threshold value.
- The output image is then obtained and image quality is assessed using PSNR and SSIM values.
- The output image was segmented using the connected components method by using gradients to colorize the different components found.
- Then the gradient range was divided into k bins , where k represents the no of lines of text in the image.
- Each gradient range was then looped through and the lines of text were separated from the rest.

6. Results

The proposed method was tested on a set of 144 epigraphs provided by the ASI. Since there does not exist an ideal or "noiseless" image for comparison, the values provided by the estimators do not provide a clear indicator to the image quality. However, a sufficiently high value of both these indicators indicate that the character integrity has been retained and that there is no loss in valuable information. The average values of SSIM and PSNR are 0.88 & 14.12 respectively. Human visual perception can easily identify the significant difference in quality between the results of the proposed method and that of an image denoised by standard noise removal techniques (Median filter, Gaussian Filter).

The algorithm is also time efficient and does not take more than 2 minutes to execute for an image with sufficient noise. The table below depicts the SSIM and PSNR values for a sample set of 10 images. The value of the estimators for the first four samples correspond to the images depicted in the following page.

Table 1 : Comparison with known algorithm

Method	Average SSIM	Average PSNR
Wiener Filter	0.9547	28.336
Median Filter	0.9410	26.033
CCD	0.8802	14.121

Table 2 : SSIM and PSNR for a few images

Image	SSIM	PSNR
Sample 1	0.87	14.27
Sample 2	0.85	13.25
Sample 3	0.92	16.37
Sample 4	0.87	13.97
Sample 5	0.88	14.21
Sample 6	0.95	17.54
Sample 7	0.94	16.32
Sample 8	0.88	13.55
Sample 9	0.81	12.04
Sample 10	0.92	15.32

A comparative study was performed to find the best possible solution with the set of novel algorithms developed for the same. Multi Scale Template Matching was first used as a baseline to benchmark the clear images. In MSTM , a template is created which consists of the noise in the epigraph. The size of this template is about 7x7 pixels. A nested median filter is applied to the image as it provides clearer demarcation between text and background. The template matching algorithm is run and a mask is created of all matches. All images above a certain predetermined threshold value are matched.A mask of all the matched noise is created and then image inpainting is applied.

The restriction to the method proposed in MSTM is the fact that the template size cannot be too large. A large template could cause character loss. Thus restricting its usage to remove small to medium size noise only. Using this method as a precursor to the proposed method has several advantages. MSTM is computationally efficient and removes small specs of noise from the inputs effectively, which could be falsely considered as connected components while running the proposed algorithm. In our experiments it was found that the no of connected components in the original image is reduced by 25% when MSTM is used. Due to this reason, the proposed method is in turn made more computationally efficient.

It was also observed that small noises are one of the reasons for false positives identified by the connected components algorithm. These small specs inherently join two large separate components which then seem as a single one to the algorithm. Thus using MSTM to remove such noise helps the proposed method to distinguish between components easily. The pipeline in MSTM involves the use of nested median filtering and template matching which removes noise such as textures of the rock, which otherwise would have been considered as a component by the algorithm. It was observed that the character loss for only connected component denoising and the ensemble method was minimal. Once the output denoised images are created a simple segmentation task was performed using the already marked connected components. The components were first colored using a gradient and the gradient range was then split. The no of bins to be split the gradient into was determined by finding the no of lines of text in the image. An example is provided below

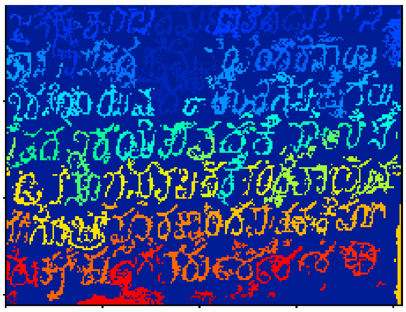

Fig 7: Gradient filled connected component

The segmentation is not free of errors due to extremely small connections caused by rocky surfaces in the images. These cause false positives and misclassify certain components. But the complete task of segmentation itself can be considered as future enhancements that can complete the algorithm.

Fig 8 : Detected line 1 and line 2 of the image

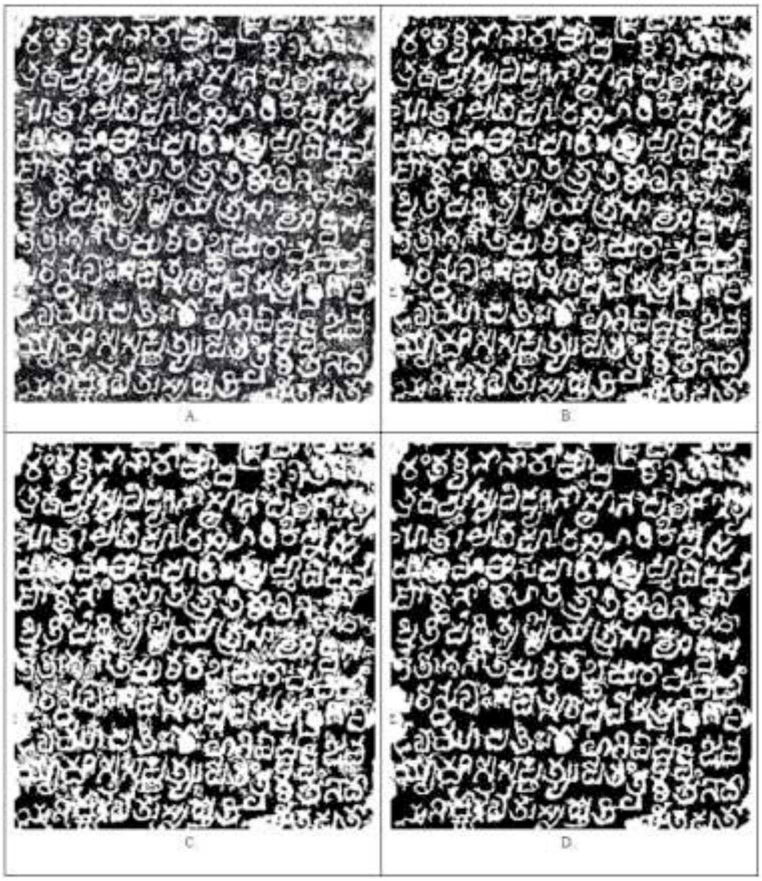

Fig 9 : A. Input , B. MSTM output , C. CCD output , D. Hybrid approach output

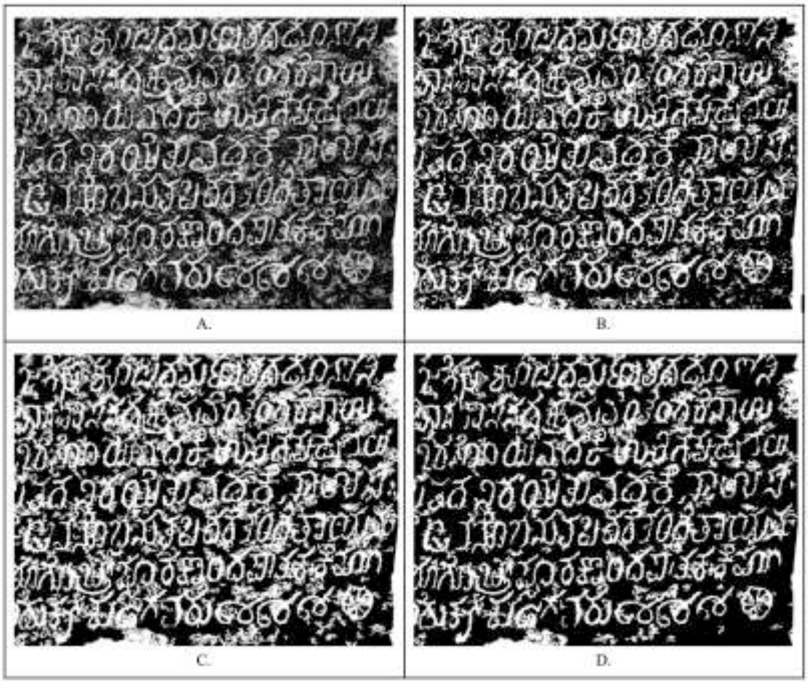

Fig 10 : A: Input , B: MSTM output , C: CCD output , D: Hybrid approach output

Fig 11 : A: Input , B: MSTM output , C: CCD output , D: Hybrid approach output

7. Conclusion

The objective of this research paper is to complete the initial phase of a process that starts with simple epigraphical estampages and finishes at character recognition. The precision of the OCR can be improved with higher quality "less noisy" images while also making sure of preserving character integrity and structure. This paper thus focuses on creating a fully automated denoising algorithm which is time and space efficient in order to aid in further steps of processing such as segmentation and recognition.The proposed method makes use of projections along with connected component labeling for denoising. The results have been evaluated using PSNR and SSIM metrics and are compared with standard denoising filters such as the median and wiener filters. It can be observed by the HVS that there is a noticeable difference in quality between the results from CCD and those of the standard filters,

8. Acknowledgment

We would like to thank the Archaeological survey of India for providing us with the dataset. We would also take this opportunity to thank Dr Mamatha H.R who has provided us with great insights and guidance.

9. References

[1] A. K. Bhunia, A. Das, P. P. Roy and U. Pal "A Comparative Study of Features for Handwritten Bangla Text Recognition", Proc. in International Conf. on Document Analysis and Recognition, pp. 636-640, 2015.

[2] R. Saabni, A. Asi, and J. El-Sana, "Text line extraction for historical document images," Pattern Recognition Letters, vol. 35, no. 0, pp. 23–33, 2014.

[3] W. Chang, C. Chiu, J. Yang, Block-based connected-component labeling algorithm using binary decision trees, Sensors, 15 (2015), pp. 23763-23787.

[4] He, L.; Ren, X.; Gao, Q.; Zhao, X.; Yao, B.; Chao, Y. The connected-component labeling problem: A review of state-of-the-art algorithms. Pattern Recognit. 2017, 70, 25–43, doi:10.1016/j.patcog.2017.04.018.

[5] Suzuki K., Horiba I., Sugie N. linear-time connected-component labeling based on sequential local operations. Comput. Vision Gr. Image Underst. 2003;89:1–23. doi: 10.1016/S1077-3142(02)00030-9

[6] Hongyao, D., & Xiuli, S. (2009). License Plate Characters Segmentation Using Projection and Template Matching. 2009 International Conference on Information Technology and Computer Science. doi:10.1109/itcs.2009.116

[7] S.Y. Huang, Y.K. Lee, G. Bell, and Z. Ou. A Projection-based Segmentation Algorithm for Breaking MSN and YAHOO CAPTCHAs, Proceedings of the World Congress on Engineering, 2008, July 2-4, London, U.K.

[8] Xia, H., & Liao, D. (2011). The study of license plate character segmentation algorithm based on vertical projection. 2011 International Conference on Consumer Electronics, Communications and Networks (CECNet). doi:10.1109/cecnet.2011.5768714.

[9] R. Kreis, "Issues of spectral quality in clinical H-magnetic resonance spectroscopy and a gallery of artifacts", NMR in Biomedicine, vol. 17, no. 6, pp. 361-381, 2004.

[10] Kwok NM, Jia X, Wang D, Chen SY, Fang G, Ha QP. Visual impact enhancement via image histogram smoothing and continuous intensity relocation. Comput Electric Eng 2011;37(5):681–94.

[11] Z. Z. Wang, "A new approach for segmentation and quantification of cells or nanoparticles," IEEE Trans. Ind. Informat., vol. 12, no. 3, pp. 962–971, Jun. 2016

[12] Z. Z. Wang, "Monitoring of the GMAW weld pool from the reflected laser lines for real-time control," IEEE Trans. Ind. Informat., vol. 10, no. 4 pp. 2073–2083, Nov. 2014

[13] I. Avcibas, B. Sankur and K. Sayood, "Statistical evaluation of image quality measures", Journal of Electronic Imaging, vol. 11, no. 2, pp. 206-223, 2002

[14] J. E. Farrell, Image quality evaluation in colour imaging: vision and technology. MacDonald, L.W. and Luo, M.R. (Eds.), John Wiley, pp. 285-313, 1999.

[15] M. Cadik and P. Slavik, "Evaluation of two principal approaches to objective image quality assessment", 8th International Conference on Information Visualisation, IEEE Computer Society Press, pp. 513-551, 2004.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.