Mathematical Problems of Computer Science 61, 7–14, 2024.
doi: 10.51408/1963-0111

UDC 004.932.2, 519.72

Analyzing the Quality of Distorted Images by the Normalized Mutual Information Measure

Mariam E. Haroutunian [1], David G. Asatryan[1,2] and Karen A. Mastoyan[1]

[1]Institute for Informatics and Automation Problems of NAS RA, Yerevan, Armenia
[2]Russian-Armenian University, Yerevan, Armenia
e-mail: armar@sci.am, dasat@iiap.sci.am, kmastoyan@gmail.com

Abstract

This research explores how different types of distorting algorithms impact the Full-Reference image quality assessment, particularly when subjective quality evaluations are incorporated. We draw upon the TID2013 database, which contains 3000 images distorted by 24 distinct algorithms, in conjunction with Mean Opinion Scores (MOS) for quality ratings. We compare the results of Normalized Mutual Information (NMI) for image quality score with W^2, based on Weibull distribution, the common PSNR similarity measure and MOS. We advocate for integrating of NMI into the repertoire of image quality assessment metrics.

Keywords: Image quality, Distortion types, Evaluation metrics, Normalized mutual information.
Article info: Received 25 March 2024; sent for review 26 March 2024; accepted 16 April 2024.

1. Introduction

Assessing the quality of images is a crucial process for various applications, including pattern recognition, classification, restoration, and others. The definition of quality, however, lacks an unambiguous formal consensus, leading to the requirement of specific interpretations of image quality and respective methods for its estimation. Three key methodologies exist for evaluating image quality. *Full-Reference* methods are based on the comparison between a distortion-free reference image and a test image, which is a distorted version of the original. The level of distortion serves as an indicator of the quality of the test image. The change in quality may either indicate a decrease or an increase, depending on the result of the distortion process [1, 2].

On the other hand, *No-Reference* methods assess the quality solely based on the analysis of the test image, taking into account its structural characteristics and other properties. *Reduced-Reference* methods fall in the middle, employing partial information about the original image in the assessment process.

There is abundant research literature focusing on these three types of image quality evaluation methods. These techniques can be bifurcated into two classes: *objective* and *subjective*. Objective methods utilize formal image theory and image processing techniques, whereas subjective methods rely on human visual system (HVS), based on expert quality assessments. The average of these subjective assessment results is the MOS [3, 4].

Several primary factors influence the assessment of image quality. The first one is the inherent quality of the test image and its depiction. The second are the distortions introduced into the image content through processes like image acquisition, visualization, transmission, etc. The third are the changes made in the image structure and parameters during image processing using various mathematical or computational methods.

Given the diversity of these factors, universal methods for assessing the quality of any image do not exist. Decisions on quality must consider the unique properties of the tested images and the employed methods, in combination with the available subjective assessments. Thus, there is continuous need for developing new quality criteria, similarity assessment methods, and methods for analyzing and comparing different approaches.

Several image databases with MOS assessments exist that have been collected through experimental procedures involving a large number of experts. For instance, 40 such databases are critically examined in [4]. The literature provides extensive references to studies on quality assessment through both objective and subjective methods [5]. In [6], the regularities of influence of the type of distorting algorithm on the result of evaluating the image quality by the Full-Reference method in the presence of subjective quality assessments were studied. As an example, the TID2013 database [7] with 3000 images distorted by 24 types of algorithms and subjective MOS quality ratings was used. An image quality score based on the Weibull distribution model and the usual Peak Signal-to-Noise Ratio (PSNR) similarity measure was applied. It was shown that the applied distorting algorithms are classified into two types - normal, leading to results consistent with the HVS, and "anomalous", the corresponding quality estimates of which are disordered or chaotic.

In this research, we investigate another approach to image quality assessment using the concept of NMI, which was introduced and studied in [8]. Its theoretical grounding in information theory [9] provides a robust and well-defined basis for measuring image similarity. Additionally, NMI's scale invariance makes it versatile and applicable to images of diverse resolutions. Furthermore, its non-parametric nature eliminates the need for prior assumptions about the image data, enhancing its adaptability to various image types. NMI quantifies the amount of information shared between the reference and the distorted images. This metric has shown potential between the reference and the distorted images [10].

As research on NMI and its applications in image quality assessment continues to evolve, exciting possibilities emerge. The development of deep learning-based NMI variants, for instance, holds promises for further enhancing accuracy and robustness in complex scenarios.

Within the scope of our investigation, we aim to rigorously examine and evaluate the performance of the NMI metric across a diverse array of image distortion types and levels. Our endeavor is directed towards discerning NMI's nuanced impact and effectiveness in capturing the similarities between datasets that have undergone different manifestations and intensities of image distortion. The research article employs a structured approach, encompassing research methodology and experimental results.

The paper is organized as follows. The next section introduces the considered measures. In Section 3 experimental results on the TID2013 database are discussed. The paper concludes in Section 4, summarizing key findings and advocating for a balanced consideration of both NMI and subjective evaluation methods.

2. Description of Considered Measures

First, we consider **MOS**, which is a subjective measure that represents the average opinion of human observers. It is useful for the evaluation of other measures.

- It quantifies the perceived quality of an image based on human evaluation.
- MOS scores are typically obtained through subjective experiments, where human observers rate the quality of images on a scale.
- Higher MOS scores indicate better perceived image quality.
- MOS is commonly used as a benchmark for objective image quality assessment algorithms.

PSNR is a widely used metric in image quality assessment, commonly applied in image processing and compression. It quantifies the fidelity of an image by comparing the maximum signal power (original image) to the noise power (introduced during representation, often as Gaussian noise). The key points are:

- (Objective Measure) PSNR provides an objective numerical assessment of image quality, enabling quantitative comparisons between different algorithms.
- (Decibel Scale) The use of decibels ensures a perceptually relevant representation of quality ratios.
- (Higher Values, Better Quality) Higher PSNR values signify better image quality with minimal noise interference [2, 13, 14].

While PSNR offers simplicity and objectivity in evaluating signal quality, it has limitations in accurately reflecting human perception. It may not be suitable for all types of signals or compression techniques. It is essential to consider its advantages and disadvantages carefully when using it for quality assessment in image and video processing applications.

$\mathbf{W^2}$ is an image quality metric. It measures the structural similarity between the original image and the image with additive Gaussian noise. W^2 values range from 0 to 1, where 1 indicates perfect structural similarity. Higher W^2 values suggest better image quality and preservation of structural information. W^2 is commonly used for image restoration and enhancement [15]. This image quality estimation is based on a Weibull distribution model of image gradient magnitude, the density of which is given by the formula

$$f(x;\lambda,\eta) = \frac{\eta}{\lambda}\left(\frac{x}{\lambda}\right)^{\eta-1}\exp\left[-\left(\frac{x}{\lambda}\right)^{\eta}\right], x \geq 0,$$

where $\eta > 0$ is the shape parameter, $\lambda > 0$ is the scale parameter. Distribution parameters are estimated from the totality of all gradient magnitudes using the Sobel operator. The similarity (proximity) of two images is estimated by the proximity of the estimates of the parameters of the Weibull distribution by the formula

$$W^2 = \frac{min\,(\eta_1, \eta_2)\,min\,(\lambda_1, \lambda_2)}{max\,(\eta_1, \eta_2)\,max\,(\lambda_1, \lambda_2)}.$$

The research in [6] presented that this measure is sensitive to those types of distortions that affect the structure and content of the image.

NMI is a measure of the distance between two images based on their joint probability distribution. It quantifies the amount of information shared between two images, considering both their individual distributions and their joint distribution. Higher NMI values indicate greater distance, whereas lower NMI values indicate less distance.

Mutual information is a fundamental concept in information theory. Given two random variables X and Y, the mutual information (MI) is defined as

$$I(X;Y) = H(X) \ + H(Y) - H(X,Y),$$

where $H(X)$ is the well-known notion of entropy [9]. MI is a non-negative quantity and can be used as a similarity or distance measure depending on various applications.

We consider the following **normalized** version of MI

$$NMI = 1 - \frac{I(X;Y)}{\max H\,(X)\,, H(Y)},$$

which is a **distance measure**. It was proved in [11] that this measure satisfies metric properties, in other words, it adheres to the criteria of a true metric, encompassing positive definiteness, symmetry, and triangle inequality. At its core, the metric property aligns with our intuitive understanding of distance, providing a foundational framework for quantifying spatial relationships. NMI values range from 0 to 1, with 0 indicating perfect similarity and 1 indicating no similarity at all. Beside from information theory, NMI is widely used also in image registration, image segmentation, and other applications [10], [12]. NMI is often used to evaluate clustering algorithms or comparing different clusterings of the same data [11]. NMI is based on the principles of information theory, which makes it theoretically grounded and well-suited for various applications in fields such as machine learning, pattern recognition, and data mining.

3. Experimental Results

The selected database is TID2013 [7]. This database contains 3000 images obtained from 25 originals, distorted by 24 different types of five levels each (for example, see Fig. 1). The authors of the database conducted an extensive experiment on the visual assessment of the quality of database images using a point system by a large number of people from different countries. As a result of processing these data, each of the 3000 images is assigned a numerical MOS score.

All necessary quantities are calculated using the developed software system, and the results are entered into Excel tables. The base data are the results related to the original and five distorted samples of a particular image.

For each such data set, three evaluation methods were employed: NMI, PSNR, W^2, and compared with MOS. These methods were utilized to assess and analyze the quality of the images in the database.

Fig. 1. The first, third, and fifth levels of distorted images from Image Number 10 in the TID 2013 database using the 'Sparse Sampling and Reconstruction' method.

In numerous scenarios, NMI demonstrated similar sensitivity in comparison with alternative measures, for example on Fig. 2 five levels of Additive noise in color distortion are demonstrated. The values of all measures for each level are given in Table 1.

Fig. 2. 5 Levels of Additive Noise in Color Distortion from the TID2013 Database: Image Number 15.

Table 1. Experimental results for the 15th image from TID2013 database (Additive noise in color)

NMI	PSNR	W^2	MOS
0.14	42.33	0.87	6.09
0.17	39.45	0.78	5.82
0.22	36.47	0.66	5.64
0.27	33.61	0.53	4.89
0.34	31.39	0.38	4.64

In some cases, NMI demonstrates higher efficiency. For example, in the case of distortion with the Non eccentricity pattern noise method, the W^2 values are close to one, which means that it performs poorly in terms of human evaluation and human understanding, and the NMI values are close to human evaluation (Table 2, Fig. 3).

Fig. 3. Reference image and level 5 distorted image (Non eccentricity pattern noise)

Table 2. Experiment results for the 8th image from TID2013 database (Non-eccentricity pattern noise).

NMI	PSNR	W^2	MOS
0.06	43.33	1	5.65
0.10	41.30	1	5.43
0.16	39.08	0.99	4.87
0.20	37.82	0.99	4.75
0.24	36.92	0.99	4

4. Conclusion

Our experimental results revealed interesting insights into the performance of metrics across different types of distortions. We found that NMI, being a normalized distance measure, showed promising results for various distortion types. Particularly, it exhibited close alignment with human subjective evaluations in almost all cases, indicating its potential as an effective image quality assessment metric.

Moreover, NMI's theoretical foundation in information theory and its versatility in capturing differences between images of diverse resolutions contribute to its robustness and applicability in image quality assessment tasks. NMI consistently demonstrated its efficacy for a wide range of distortions.

In conclusion, our findings advocate for integrating NMI into the repertoire of image quality assessment metrics, complementing traditional measures like PSNR and W^2. By leveraging NMI's inherent advantages and considering its performance in conjunction with subjective evaluations, we can enhance the accuracy and reliability of image quality assessment methods, catering to diverse application scenarios in image processing, computer vision, and beyond.

References

[1] A. George and S. J. Livingston, "A survey on full reference image quality assessment algorithms", *International Journal of Research in Engineering and Technology*, vol.2, no.12, pp. 303-307, 2013.

[2] 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.

[3] T.-J. Liu, Y.-C. Lin, W. Lin and C.-C.J. Kuo, "Visual Quality Assessment: Recent Developments, Coding Applications and Future Trends", *APSIPA Trans. on Signal and Information Processing*, vol. 2, 2013. DOI: https://doi.org/10.1017/ATSIP.2013.5

[4] G. Zhai and X. Min, "Perceptual image quality assessment: a survey", *Science China Information Sciences*, vol. 63, no. 11, 211301, 2020. DOI: 10.1007/s11432-019-2757-1

[5] J. Wang, Z. Wang, L. Lu and A. C. Bovik, "Subjective Quality Assessment of Stereoscopic Omnidirectional Image", *Lecture Notes in Computer Science, in book Advances in Multimedia Information Processing PCM*, pp. 589-599, 2018.

[6] D. Asatryan, M. Haroutunian, G. Sazhumyan and G. Hakobyan, "Procedure for analyzing the quality, structure and subjective rating of distorted images by the Full-Reference technique", *Intern. Scientific Journals of Scientific Technical Union of Mechanical Engineering "Industry 4.0", Mathematical Modeling*, vol. 6, no. 4, pp. 100-102, 2022.

[7] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, Battisti F., et al. "Image database tid2013", *Signal Processing: Image Communication*, vol. 30, pp. 57-77, 2015.

[8] A. Kraskov, H. Stogbauer, R. G. Andrzejak and P. Grassberger, "Estimating mutual information", *Physical Review*, E 69, 066138, 2004.

[9] T. M. Cover and J. A Thomas, *Elements of Information Theory*, Second Edition, Wiley-Interscience, 2006.

[10] F. E. Ruiz, P. S. Pérez and B. I. Bonev, "Information theory in computer vision and pattern recognition", *Springer Science and Business Media*, 2009.

[11] N. X. Vinh, J. Epps and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance", *Journal of Machine Learning Research*, vol. 11, pp. 2837-2854, 2010.

[12] J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, "Mutual-information-based registration of medical images: A survey", *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986-1004, Aug. 2003.

[13] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 3rd edition, Pearson Education, 2008.

[14] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment", *In The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1398-1402, 2003.

[15] D. Asatryan, "Image blur estimation using gradient field analysis", *Computer Optics*, vol. 41, no. 6, pp. 957-962, 2017. DOI: 10.18287/2412-6179-2017-41-6-957-962.

Աղավաղված պատկերների որակի վերլուծություն նորմալացված փոխադարձ ինֆորմացիայի միջոցով

Մարիամ Ե. Հարությունյան[1], Դավիթ Գ. Ասատրյան[1,2] և Կարեն Ա. Մաստոյան[1]

[1]ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ, Երևան, Հայաստան
[2] Հայ-Ռուսական համալսարան, Երևան, Հայաստան
e-mail: armar@sci.am, dasat@iiap.sci.am, kmastoyan@gmail.co

Ամփոփում

Այս հետազոտության մեջ մենք ստուգանմուշի օգտագործմամբ ուսումնասիրում ենք, թե ինչպես են տարբեր տեսակի աղավաղող ալգորիթմները ազդում պատկերի որակի ամբողջական գնահատման վրա, մասնավորապես, երբ ներառված են սուբյեկտիվ որակի գնահատականները: Մենք օգտվում ենք Tid2013 տվյալների բազայից, որը պարունակում է3000 պատկեր, որոնք աղավաղված են 24 տարբեր ալգորիթմներով, կարծիքների միջին միավորների (MOS) հետ համատեղ: Մենք համեմատում ենք նորմալացված փոխադարձ ինֆորմացիայի արդյունքները (NMI) պատկերի որակի գնահատման համար Վեյբուլի բաշխման վրա հիմնված W^2 արդյունքների, հայտնի PSNR նմանության չափի և MOS-ի հետ: Մենք պնդում ենք NMI- ի ինտեգրումը պատկերի որակի գնահատման չափման մեծությունների ցանկում:

Բանալի բառեր՝ Պատկերի որակ, աղավաղման տեսակներ, գնահատման չափումներ, նորմալացված փոխադարձ ինֆորմացիա:

Анализ качества искаженных изображений с помощью нормализованной меры взаимной информации

Мариам Е. Арытюнян[1], Давид Г. Асатрян[1,2] и Карен А. Мастоян[1]

[1]Институт проблем информатики и автоматизации НАН РА, Ереван, Армения
[2]Российско-Армянский университет, Ереван, Армени
e-mail: armar@sci.am, dasat@iiap.sci.am, kmastoyan@gmail.com

Аннотация

В этой статье методом сравнения с эталоном мы исследуем, как различные типы алгоритмов искажения влияют на оценку качества изображения с полной ссылкой, особенно при включении субъективных оценок качества. Мы опираемся на базу данных TID2013, которая содержит 3000 изображений, искаженных 24 различными алгоритмами, в сочетании со средними оценками мнений (MOS) для рейтингов качества. Мы сравниваем результаты нормализованной взаимной информации (NMI) для оценки качества изображения с W^2, на основе распределения Вейбулла, общего показателя сходства PSNR и MOS. Мы выступаем за интеграцию NMI в репертуар показателей оценки качества изображения.

Ключевые слова: Качество изображения, типы искажений, метрики оценки, нормализованная взаимная информация.