# How to Optimize the Utilization of Image Quality Metrics in Computer Vision?

Aladine Chetouani
University of Orleans, PRISME laboratory
Orléans
France
aladine.chetouani@univ-orleans.fr

Mohammed El Hassouni
LRIT, Rabat IT Center, Mohammed V
Rabat
Morocco
mohammed.elhassouni@um5.ac.ma

## ABSTRACT

In this paper, we propose to show the importance to consider the image quality in Computer Vision (CV) applications. We also describe a proposed framework that not only take into account the quality but rather permits to select the more adapted measure for a given CV application. Here, the selection of the image quality metric is based on a degradation identification step using a Linear Discriminant Analysis (LDA) method. The proposed framework has been applied to a Full-Reference approach where the reference image is supposed to be available and for No-Reference approach where only the captured image is accessible. The method has been tested using the TID 2008 database, which is composed of 17 degradation types.

## KEYWORDS

Image Quality, Computer Vision, Degradation identification proceedings

**ACM Reference format:**

A. Chetouani, and M. El Hassouni. 2018. How to optimize the utilization of image quality metrics in computer vision ? In Proceedings of Mediterranean Conference on Pattern Recognition and Artificial Intelligence, Rabat, Morocco, March 27–28, 2018 (MedPRAI '18), 6 pages. https://doi.org/10.1145/3177148.3180080

## 1 INTRODUCTION

Computer Vision (CV) domain is a very wide research field which employs many techniques of image processing and pattern recognition to achieve its goals. A large number of applications can be cited such as in biometry for security (face, fingerprint, [1]iris, gait, etc.), robotics (localization, interaction, navigation, etc.), social interaction (autism, behaviour, etc.), medical imaging (reconstruction, Vision-guided, etc.), Human-Robot Interaction (HRI) (emotion, service, etc.) and so on.

For all these applications, images are considered as the principal input. So, the performance and the efficiency of these methods depend highly on the quality of this data. However, images are susceptible to be degraded by several necessary treatments, intrinsic to the application such as compression or transmission processes. Some external degradation types can also appear due to the acquisition environment or the sensors. These degradations can impact significantly the performance of a given CV method.

It is thus necessary to consider the possible encountered degradation types that can affect the input data in order to provide a feedback to users. In [1], the authors studied the impact of the quality of the image in the performance and proposed to improve the results by adding a quality measure in the authentication step. The objective here is to predict how well the used system will perform recognition. In [2], the authors showed that the performance of fingerprint application can be improved by 20% if the quality of the input data is considered. In [3], several tests have been done to show the importance of the image quality in face recognition. The obtained performances of different resolutions, noise, and blur levels have been compared and the results show that these factors impact the efficiency of the used method.

As mentioned in [4], the quality estimation process can be included in different steps of a given CV application: preprocessing (re-acquisition), recognition (as features), decision and so on. In all these cases, the evaluation of the quality of the data is very important and leads to improve the performance. Some interesting studies addressed also this problem [5][6].

In this article, we focus on this issue and we propose a new strategy in this area in order to better integrate the quality


MedPRAI '18, March 27–28, 2018, Rabat, Morocco

ACM ISBN 978-1-4503-5290-1/18/03…$15.00
https://doi.org/10.1145/3177148.3180097

information in the process and thus optimize the utilization of existing image quality metrics in CV applications.

This paper is organized as follows: section 2 presents the commonly encountered degradation types and the existing approaches to estimate its quality. The proposed framework is described in section 3. Some tests are exposed in section 4. The conclusion is then shown in section 5.

## 2 COMMON ENCOUNTERED DEGRADATION TYPES IN COMPUTER VISION

Such as several domains, CV applications have it's owned degradation types. Some of them are often encountered in image processing domain and some others are very specifics. We list below some of these types of degradation with its origin (not exhaustive):

- **Noise (Gaussian):** This degradation affects the image by adding information that alters the performance. The saliency information will be thus mixed and lost in the noise. It can be due to acquisition, environment or sensors.

- **Blur:** Blur affects edges and details in an image by smoothing its transmission. Incorrect focus, motion, and compression can generate it. In [4], the authors have mentioned that the motion blur and Gaussian noise can impact respectively the performance of a given system from 20% to 35%.

- **Blocking effect:** Blocking effect has been widely studied in image quality domain and it is known as vertical and horizontal artificial boundaries that appear in the image. It's generally due to a block-based method such as in compression step (JPEG compression) or during the transmission step.

- **Ringing effect:** So-called Gibbs 2D, this degradation affects textured regions and high contrasts in the image. It appears as edges around these specific regions. This is generally due to a quantification step such as in JPEG2000 compression.

- **Illumination:** A no-uniform lighting can impact the performance of some CV applications. The saliency information will be visible in a part of the image and not visible in the other part. This degradation can be due to the sensor or to external reasons.

To quantify the quality of all these types of degradation, different image quality measures have been proposed in the literature [7]. The existing metrics can be divided into three approaches: Full-Reference (FR), No-Reference (NR) and Reduced-Reference (RR). In FR approach, the reference image is supposed available entirely, while in NR approach, only the degraded or the captured image is accessible. Metrics of this last one category (NR) are often degradation-dependent. In other words, a given NR metric should be used for a given degradation type. In RR approach, only some features of the reference image are available and are compared to the degraded image. Note that in the CV domain, some applications have the reference images such as in authentication applications, while for some others only the captured image (considered here as the degraded image) is given as in some medical applications.

## 3 PROPOSED FRAMEWORK

As explained above, the performances of CV applications are sensitive to these degradations and thus should be taken into account. In order to enhance their efficiencies, we propose to optimize the utilization of the existing image quality measures for such kind of applications. It is important to mention that we do not propose as previous works [3,4] to use a given metric to take into account the quality information but rather propose to insert an intermediate step that aims to identify the type of degradation contained in the image. Once this identification is done, the more effective measure to estimate the detected degradation type can be then selected. By this way, the quality estimation is more optimal and more precise than simply using a unique measure whatever the degradation type. Indeed, as mentioned in [8], a given FR quality measure provides good estimation for some degradation types and poor results for some others.

**Table 1: Ranking of some FR metrics according to the PCC**

| IQM ranking | Degradation type | | |
|---|---|---|---|
| | **Blur** | **JPEG** | **JPEG2000** |
| **1** | **VIFP** | PSNRH | PSNRH |
| **2** | VIF | PSNRM | PSNRM |
| **3** | WSNR | VIF | NQM |
| **4** | VSNR | WSNR | WSNR |
| **5** | PSNRM | NQM | **VIFP** |
| **6** | PSNRH | **VIFP** | VSNR |
| **7** | SSIM | VSNR | UQI |
| **8** | UQI | SSIM | VIF |
| **9** | NQM | PSNR | PSNR |
| **10** | IFC | XYZ | SSIM |
| **11** | PSNR | UQI | IFC |
| **12** | XYZ | IFC | XYZ |

In order to better explicit it, Table I shows a ranking of 12 common used FR measures (VIFP / VIF [9], WSNR [10], PSNRM [11], PSNRH [12], SSIM [13], UQI [14], NQM [15], IFC [16], XYZ [17], VSNR [18]). The ranking is here based on the Pearson Correlation Coefficients (PCC) using the TID 2008 database [19] (i.e. we sort the PCC values). These PCC values are computed between the objective scores (index quality) and the subjective scores, also so-called MOS (Mean Opinion Score). This last one is obtained by some psycho-visual experimentations. As we can see, the VIFP measure obtains the best ranking for Blur distortion and, occupies the position 6 and 5 for JPEG and JPEG2000 compression methods, respectively.

Note that in the case where only the captured image is supposed available, NR measures should be used. However, this kind of metrics is often degradation-based [20,25,24]. There are some NR metrics, which can be applied to measure several types of degradation [28,29]. Nevertheless, their performances depend highly on the database. Moreover, the best performances are not obtained by this kind of metrics for all degradation types [15].

**Table 2: Obtained correlations for different degradation types.**

| | Pearson Correlation | Spearman Correlation |
|---|---|---|
| **Noise** | 0.52 | 0.57 |
| **Blur** | 0.34 | 0.44 |
| **Blocking** | **0.93** | **0.91** |
| **Ringing** | 0.40 | 0.40 |

In Table 2, we show the PCC and Spearman correlation values obtained for different degradation types using a given NR measure (blocking effect) [20]. A good result is given only for one degradation type (Blocking). In other words, most of the existing NR metrics can be used in CV applications only if we know the degradation type contained in the image. In order to better use the quality information and to have a universal framework that permits to estimate the quality whatever the CV applications (FR or NR approaches), we propose the following framework (see Fig. 1). According to the approach (FR or NR), some features are first extracted from the captured image (NR) or from the reference image and its captured version (FR). A Linear Discriminant Analysis (LDA) method is then used as a classifier. Mahalanobis distance has been used as criteria. Once the degradation type is identified, the more appropriate measure is then chosen to estimate the quality of the data. Depending on the CV application, this quality index is exploited to improve its performance as mentioned in [4].

## 4 TESTS AND RESULTS

Some results in terms of degradation identification are here presented for both FR and NR approaches. Due to the diversity of the input data of CV applications, we opted to use a dataset that contains different real images with different types of degradation. Indeed, the goal is to cover a wide range of applications (surveillance, tracking, localization, face, etc.). In this work, the TID 2008 [19] dataset has been selected for its diversity (face, texture objects, etc.) in terms of degradation types (see Table 3 and Fig 2). This database is composed of 1700 images with 100 images per degradation type. Note that for the NR approach, the existing measures are developed only for few degradation types (blocking, ringing, noise, and blur). The degradation identification method for this last one has been thus tested only for these four types of degradation.

**Table 3: Degradation types of the TID 2008.**

| Degradation | Type |
|---|---|
| 1 | **Additive Gaussian noise** |
| 2 | Additive noise in color components |
| 3 | Spatially correlated noise |
| 4 | Masked noise |
| 5 | High-frequency noise |
| 6 | Impulse noise |
| 7 | Quantization noise |
| 8 | **Gaussian Blur** |
| 9 | Image denoising |
| 10 | **JPEG compression** |
| 11 | **JPEG2000 compression** |
| 12 | JPEG transmission errors |
| 13 | JPEG2000 transmission errors |
| 14 | Noneccentricity pattern noise |
| 15 | Local block-wise distortions of different intensity |
| 16 | **Mean shift (intensity shift)** |
| 17 | Contrast change |

**Figure 2: Sample of the TID 2008 database.**

In the following, we show the results from both FR and NR approaches by presenting the extracted features and the obtained results in terms of percentage of good classification. For both, the learning procedure is repeated 1000 times. For each time, 75% of the database is used for the training step (for each degradation type), while 25% is used for the testing with no overlap. We also present the results with different sizes.

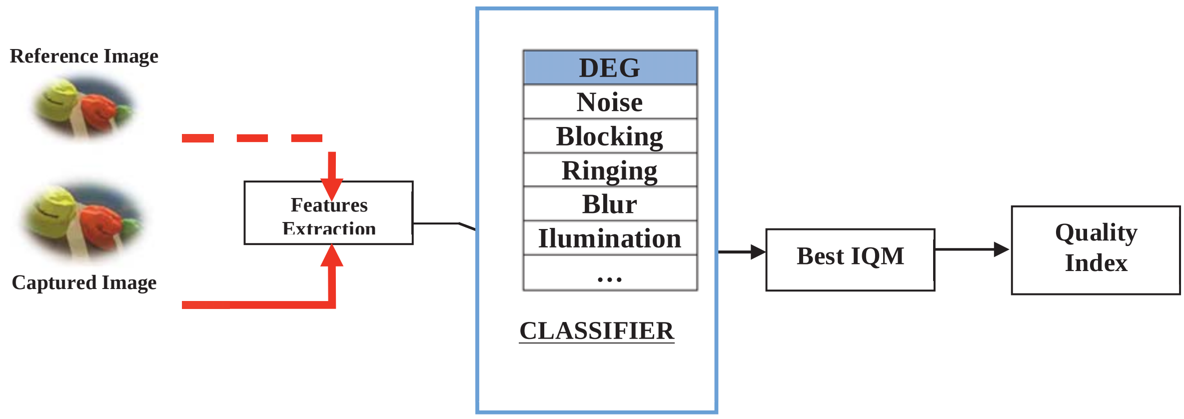


**Figure 1: Flowchart of the proposed degradation identification method.**

## 4.1. Full Reference Approach

In Table 4, we show the selected features (12 features) where some of them are based on Human Visual System (HVS) characteristics, while some others are based on mutual information or extract structural data.

Fig.3 shows the obtained confusion matrix. The mean percentage of good classification is around 94.55% with some confusions have been noted. The more high confusions (10.42%) are between the classes 12 (JPEG Transmission errors) and 13 (JPEG2000 Transmission errors). The poor classification is obtained for the class 12 (JPEG Transmission errors) with a good percentage of classification equal to 81.46%. We also test our degradation identification method for different sizes of the subsets. The variation is around 4.5%.

**Table 4: Selected features for FR approach.**

| IQM | Domain/Based on | IQM | Domain/Based on |
|---|---|---|---|
| **PSNR** | Spatial/Pixel | **VSNR** | Spatial/HVS |
| **SSIM** | Spatial/Structural | **XYZ** | Spatial/HVS |
| **UQI** | Spatial/Structural | **WSNR** | Fourier/HVS |
| **IFC** | Spatial/Structural | **PSNR-HVS** | DCT/HVS |
| **VIFP** | Spatial/Mutual information | **PSNRHVS-M** | DCT/HVS |
| **NQM** | Spatial/HVS | **VIF** | Wavelet/Mutual information |

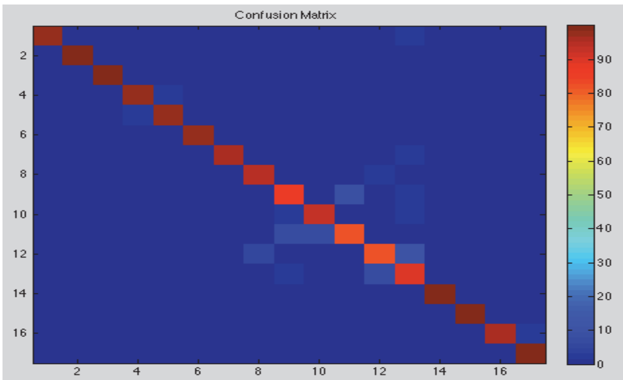


**Figure 3: Obtained confusion matrix.**

**Table 5: Obtained percentage of good classification for different sizes of the training subset.**

| | Size of the training subset | | |
|---|---|---|---|
| | **35%** | **50%** | **75%** |
| **Percentage of good classification** | 90.0544 | 93.6406 | 94.55 |

## 4.2. No Reference Approach

For this approach, 8 features have been selected (see Table 6). Some of them are based on the frequency transforms, while some others are computed in the spatial domain.

**Table 6: Selected features for NR approach.**

| IQM | Domain/Based on | IQM | Domain/Based on |
|---|---|---|---|
| **1** | Wavelet [21] | **5** | Spatial [20] |
| **2** | Spatial [22] | **6** | Wavelet [25] |
| **3** | Frequency [23] | **7** | Spatial [26] |
| **4** | Contrast [24] | **8** | DCT [27] |

In Table 7, we present the obtained matrix confusion for the four considered degradation types. We obtain 90.20% as a mean percentage of good classification. Some confusions are noted, especially between blur and JPEG2000. These errors can be easily explained by the fact that for a certain bitrate, blur appears also in JPEG2000 compressed images as we can see in Fig. 3.

**Table 7: Obtained confusion matrix for NR approach.**

| True class | Estimated Class: Noise | Blur | JPEG | JPEG2000 |
|---|---|---|---|---|
| **Noise** | **91.936** | 0.256 | 7.808 | 0 |
| **Blur** | 0 | **85.632** | 3.660 | 10.708 |
| **JPEG** | 0.364 | 0 | **98.996** | 0.640 |
| **JPEG2000** | 0 | 10.560 | 5.1880 | **84.252** |

Therefore, our method permits to well detect the degradation type contained in the image for both FR and NR approaches. Now, we can select the more appropriate measure for the identified degradation and use it as additional information in CV applications to improve its performance.

a)

b)

**Figure 4: Degraded Images: a) Blur degradation, b) JPEG2000 compression degradation.**

We also test our degradation identification method for different sizes of the subsets (see Table 8). As the FR approach, the performance varies slightly.

**Table 8: Obtaine percentage of good classification for different sizes of the training subset**

| | Size of the training subset | | |
|---|---|---|---|
| | **25%** | **50%** | **75%** |
| **Percentage of good classification** | 88.3327 | 89.938 | 90.20 |

## 5 CONCLUSION AND PERSPECTIVES

In this paper, we have presented an interesting framework for improving the performance of CV methods. This framework is based on a degradation identification step. This approach permits to consider the quality of the input data (image) of CV methods not by using a given measure, but rather by selecting the more appropriate measure. The obtained index quality index can be then integrated into the process.

## References


[1] P. Grother, and E. Tabassi, "Performance of Biometric Quality Measures", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.29, no.4, pp. 531-543, 2007.

[2] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia and A. K. Jain., "Incorporating image quality in multi-algorithm fingerprint verification", International Conference on Biometrics (ICB), Springer LNCS-3832, pp. 213-2207, 2006.

[3] A. Dutta, R.N.J. Veldhuis, and L.J. Spreeuwers, "*The Impact of Image Quality on the Performance of Face Recognition*", Symposium on Information Theory in the Benelux, pp. 141-148, 2012.

[4] S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric Quality: A Review of Fingerprint, Iris, and Face", *EURASIP Journal on Image and Video Processing, 2014.*

[5] K. Kryszczuk, J. Richiardi, and A. Drygajlo, "Impact of combining quality measures on biometric sample matching", IEEE International Conference in Biometrics: Theory, Applications, and Systems, pp. 1-6, 2009.

[6] J.R. Beveridge, P.J Phillips, G.H. Givens, B.A. Draper, M.N. Teli, and D.S Bolme, "When high-quality face images match poorly", IEEE International Conference on in Automatic Face & Gesture Recognition and Workshops, pp. 572-578, 2011.

[7] M. Pedersen and J. Y. Hardeberg, " Survey of full-reference image quality metrics," Høgskolen i Gjøviks rapportserie ISSN: 1890-520X, 2009.

[8] A. Chetouani, "Full Reference Image Quality Assessment: Limitation", International Conference on Pattern Recognition, pp. 833-837, 2014.

[9] H.R. Sheikh and A.C Bovik, "Image information and visual quality", IEEE Transactions on Image Processing, Vol.15, pp. 430-444, 2006.

[10] J. B. Martens, "Multidimensional modeling of image quality", in Proceedings of the IEEE, Vol. 90, pp133-153, 2002.

[11] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti and M. Carli, "New full-reference quality metrics based on HVS", International Workshop on Video Processing and Quality Metrics, 2006.

[12] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola and V. Lukin, "On between-coefficient contrast masking of DCT basis functions", International Workshop on Video Processing and Quality Metrics, 2007.

[13] Z. Wang, E.P. Simoncelli and A.C Bovik, "Multi-scale structural similarity for image quality assessment", Asilomar Conference on Signals, Systems and Computers, Vol. 2, pp. 1398-1402, 2003.

[14] Z. Wang and A.C. Bovik, "A universal image quality index", IEEE Signal Processing Letters, Vol. 9, pp. 81-84, 2002.

[15] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans and A.C. Bovik, "Image Quality Assessment Based on a Degradation Model", IEEE Transactions on Image Processing, Vol. 9, pp. 636-650, 2000.

[16] H.R. Sheikh, A.C. Bovik and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", IEEE Transactions on Image Processing, Vol.14, pp. 2117-2128, 2005.

[17] B. W. Kolpatzik and C. A. Bouman, "Optimized Universal Color Palette Design for Error Diffusion", Journal Electronic Imaging, Vol. 4, pp. 131-143, 1995.

[18] D.M. Chandler and S.S. Hemami, "VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images", IEEE Transactions on Image Processing, Vol. 16, pp. 2284-2298, 2007.

[19] N. Ponomarenko, M. Carli, V. Lukin, K. Egiazarian, J. Astola and F. Battisti, "Color Image Database for Evaluation of Image Quality Metrics", International Workshop on Multimedia Signal Processing, Australia, pp. 403-408, 2008.

[20] Z. Wang, A.C. Bovik and B.L. Evans, "Blind measurement of blocking artefacts in images," IEEE International Conference on Image Processing, Vol. 3, pp. 981-984, 2000.

[21] H. Tong, M. Li, H. Zhang, and C. Zhang, "Blur detection for digital images using wavelet transform", IEEE International Conference on Multimedia and Expo, Vol. 1, pp. 17-20, 2004.

[22] F. Crête, "Estimer, mesurer et corriger les artefacts de compression pour la télévision", Université Joseph Fourier, 2007.

[23] A. Chetouani, A. Beghdadi and M. Deriche, "A new free reference image quality index for blur estimation in the frequency domain", IEEE Symposium on Signal Processing and Information Technology, 2009.

[24] R. Ferzli and J. L. Karam, "A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur", IEEE Transactions on Image Processing, Vol. 18, pp. 717- 728, 2009.

[25] H. R. Sheikh, A. C. Bovik, and L. K. Cormack, "No-Reference Quality Assessment Using Natural Scene Statistics: JPEG2000", *IEEE Transactions on Image Processing*, Vol. 14, 2005.

[26] D. Van de Ville and M. Kocher, "SURE-Based Non-Local Means", IEEE Signal Processing Letters, Vol. 16, pp. 973-976, 2009.

[27] M.J. Buclkey, "Fast computation of a discretized thin-plate smoothing spline for image data", Biometrika, Vol. 81, pp 247-258, 1994.

[28] A. Mittal, A. K. Moorthy and A. C. Bovik, "Referenceless Image Spatial Quality Evaluation Engine", Asilomar Conference on Signals, Systems and Computers, 2011.

[29] M. A. Saad, A. C. Bovik and C. Charrier, "Model-Based Blind Image Quality Assessment: A natural scene statistics approach in the DCT domain", IEEE Transactions Image Processing, pp. 3339-3352, vol. 21, no. 8, 2012.