Performance Evaluation of No-Reference Image Quality Metrics for Visible Wavelength Iris Biometric Images

Xinwei Liu[1,2], Christophe Charrier[1], Marius Pedersen[2], and Patrick Bours[2]
[1] Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France
[2] NTNU - Norwegian University of Science and Technology, Gjøvik, Norway
xinwei.liu@unicaen.fr

***Abstract*—Image quality assessment plays an important role in iris recognition systems because the system performance is affected by low quality iris images. With the development of electronic color imaging, there are more and more researches about visible wavelength (VW) iris recognition. Compared to the near infrared iris images, using VW iris images acquired under unconstrained imaging conditions is a more challenging task for the iris recognition system. However, the number of quality assessment methods for VW iris images is limited. Therefore, it is interested to investigate whether existing no-reference image quality metrics (IQMs) which are designed for natural images can assess the quality of VW iris images. In this paper, we evaluate the performance of 15 selected no-reference IQMs on VW iris biometrics. The experimental results show that several IQMs can assess iris sample quality according to the system performance.**

***Index Terms*—biometric, image quality assessment, visible wavelength iris, performance evaluation, image based attributes, multi-modality.**

I. Introduction

Due to the effectiveness proven by the deployed iris recognition systems, iris is one of the most commonly used modalities for biometric recognition. The traditional near infrared iris recognition systems are very constrained in order to ensure the quality of the acquired iris images. Recently, many research initiatives sought to increase distance and relax acquisition constraints, which extends the applicability of this technology to forensic domains where VW iris acquisition devices are used [1], [2]. The VW iris imaging systems lead to acquire degraded iris samples due to less constrained environments that makes the sample quality assessment a major issue.

Multi-modality biometric recognition technologies have became more popular in recent years [3]. However, biometric sample quality assessment methods that can be used for the evaluation of multi-modality sample quality are rarely considered. It is necessary to investigate if it is possible to develop quality metrics that can assess the quality of biometric image samples from multiple modalities. There are two kinds of quality attributes when assessing biometric sample quality: image-based and modality-based attributes. Image-based attributes are, for instance, contrast, sharpness etc. which are presented in all image-based biometric modalities (e.g. face, iris, palm print and so on). Modality-based attributes can be used for only one modality, such as iris-pupil contrast in iris biometric, or pose symmetry in face biometric. Using image-based quality attributes in quality assessment approaches make it possible to assess image-based multi-modality biometric sample quality [3]. There are many existing IQMs that have been developed for the evaluation of natural image's quality [4]. Based on the availability of a reference image, IQMs can be classified into full-reference, reduced-reference, and no-reference methods [5]. According to the properties of iris images, only no-reference IQMs might be suitable for the assessment of iris image quality. The goal of this paper is to investigate whether existing no-reference IQMs can assess VW iris image quality based on the biometric system performance.

In this paper, we selected 15 no-reference IQMs to be evaluated. A near infrared iris recognition algorithm is adapted to the VW iris sample in order to evaluate biometric system performance. Iris images from the GC^2 multimodality biometric database is used in this paper. The structure of the paper is described as follows. We first present related works and background. Then the experimental setup followed by the experimental results and their analysis are introduced. At last the conclusion and future work are presented.

II. State-of-the-Art

In the ISO/IEC standard: 29794-6 Information technology - Biometric sample quality - Part 6: Iris image data [6], the iris image quality is given as a predictor of biometric performance, such as the likelihood of achieving a correct match. Currently, image quality assessment approaches can be used to evaluate iris quality before iris recognition. It can helps to improve the quality of iris samples by either applying image enhancement methods to improve image quality, choosing different recognition systems depending on iris quality, or re-capturing the iris images. Therefore, it is necessary to assess iris image quality before the recognition process. There are many factors that can affect iris image quality, and the performance of biometric systems. It is important to take into account image quality attributes that influence iris quality. Both image-based and modality-based iris quality attributes are presented in [6]. Since we don't investigate modality based

This research is supported by the Conseil Regional Basse-Normandie Grant 14P02048 and Research Council of Norway through project number 221073: HyPerCept - Color and Quality in Higher Dimensions.

attributes, we only consider the following image-based iris quality attributes from [6]: noise, illumination intensity, image brightness, image contrast, focus, blur and sharpness [7].

Fingerprint, iris or face images can be considered as different subspaces evoluted at different places within the natural image space. Thus, using image-based quality attributes for biometric samples makes it possible to develop multi-modality biometric sample quality assessment method. Liu *et al.* [3] suggest to use five quality attributes when evaluating any kind of image-based biometric sample quality and they are based on the survey of state-of-the-art research works [6], [8]–[11]. We apply four of them in this paper and the four image-based quality attributes and their definitions are given as:

1) The **contrast** attribute has two aspects: local contrast and global contrast. The local contrast can be defined as the average difference between neighboring pixels' intensity. The global contrast is defined as the weighted sums of the overall local contrast for different resolutions. It is correlated to the 'iris-sclera contrast' and 'iris-pupil contrast' attribute in [6].
2) The **sharpness** attribute is defined as the clarity of biometric sample structure and details. It is correlated to the 'sharpness' attributes in [6].
3) The **luminance** attribute can be defined as the intensity of the biometric sample illumination.
4) The **artifacts** attribute is given as any undesired alteration in biometric sample introduced during its digital processing, such as noise, compression and so on.

These are the most important image-based attributes for the evaluation of iris image quality, and image-based multi-modality biometric sample quality.

Liu *et al.* [7] conducted a similar study to discover the performance of no-reference IQMs on VW iris images. However, they only use high quality VW iris images and all of their selected IQMs cannot access VW iris image quality based on the performance of iris recognition system. 13 no-reference IQMs were used in their work: AQI and AQIP are two metrics based on anisotropy; BIQI, BLIINDS2, BRISQUE, and ILNIQE2 are four generalized purposes blind metrics; CONTRAST which is a contrast metric; JNBM, DCTSP, SH, and SSH are four sharpness metrics; PWN is a metric for measuring noise; and SSEQ is a metric based on spatial and spectral entropies. In order to better evaluate the performance of no-reference IQMs for iris biometrics, we introduce four types (related to the four image-based quality attributes introduced above) and totally eight different distortions to the VW iris images. We also select several additional IQMs for the evaluation of their performance on iris images. Moreover, we will optimize the IQM which has better performance than the other selected ones.

III. Experimental Setup

A. VW Iris Image Database - GC^2 Multi-Modality Biometric Database

Since we focus on image-based quality attributes, we need to choose a specific iris database that only contains image-based distortions but not including modality-based degradations. Existing iris databases mostly contain both image-based and modality-based degradations. Therefore, we use a multiple modality biometric database named "**GC^2 Multi-Modality Biometric Database** [7]". This database has three biometric modalities: face, contactless fingerprint, and visible wavelength iris. Three cameras are used for the acquisition: 1) a Lytro [12] first generation Light Field Camera (LFC) (11 Megapixels), 2) a Google Nexus 5 embedded camera (8 Megapixels), and 3) a Canon D700 with Canon EF 100mm f/2.8L Macro Lens (18 Megapixels). There are 50 subjects in the database. We only use the iris images from this database in this paper. For the iris modality, 15 iris samples per eye per camera have been acquired. There are 4500 iris images in the database. In addition, we introduced different distortions to these original iris images as described below. Therefore, totally 180,000 degraded iris images are in the database.


(a) (b) (c) (d)

Fig. 1. Degraded iris samples for degradation level 5. The first column represents too high contrast iris image (upper) and too low contrast iris image (lower); the second column represents motion blurred iris image (upper) and Gaussian blurred iris image (lower); the third column represents high luminance iris image (upper) and low luminance iris image (lower), and the last column represents iris image contains poisson noise (upper) and JPEG compressed iris image (lower).

In order to obtain image-based distortions correlated to these four attributes, we need to artificially degrade iris images in the database. Inspired by the techniques used in CID:IQ image quality database [13] and a similar study in biometric sample quality assessment [14], we degrade iris images into five degradation levels (one to five, from little degraded to highly degraded) for each distortion as the following (all image processing is conducted by using Matlab R2016 a): low and high contrast distortions, motion blur and Gaussian blur distortions, low and high luminance distortions, poisson noise and JPEG compression artifacts distortions. Examples of degraded iris images for degradation level 5 (highly degraded) are shown in Fig. 1.

B. No-Reference IQMs and their classification

Based on the survey and the availability of the source codes, we selected 15 no-reference IQMs for the performance evaluation. In addition to the IQMs used by Liu *et al.* [7], we select several new IQMs: CONTRAST2 which is a contrast metric; dipIQ is a generalized purposes metric; and JPEG which is a metric measuring JPEG comparison artifacts. These IQMs have high correlation with the image-based quality attributes [3], [7]. We classify these IQMs into two categories:

TABLE I

CLASSIFICATION OF THE SELECTED IQMS

IQMs	Distortion-specific	Generalized purposes
NSS	CONTRAST [15]	BIQI [16], BLIINDS2 [17], BRISQUE [18], ILNIQE2 [19]
Non NSS	JNBM [20], DCTSP [21], SH [22], CONTRAST2 [23], JPEG [24], PWN [25]	AQI [26], AQIP [26], dipIQ [27], SSEQ [28]

1) distortion specific, and 2) generalized purposes holistic IQMs. In each category, we separate IQMs into two groups: Natural Scene Statistics (NSS) based and non NSS based IQMs. The classification of the selected IQMs is in Table I.

C. Iris Recognition System

We adapt a near infrared iris recognition system to VW iris data in this paper: OSIRIS (Open Source for IRIS) version 4.1 [29]. The OSIRIS reference system is an open source iris recognition system developed in the framework of the BioSecure project [29]. OSIRIS is composed of four modules: segmentation, normalization, feature extraction and matching. Those modules are classical for iris recognition and follow the main steps proposed by Daugman [30].

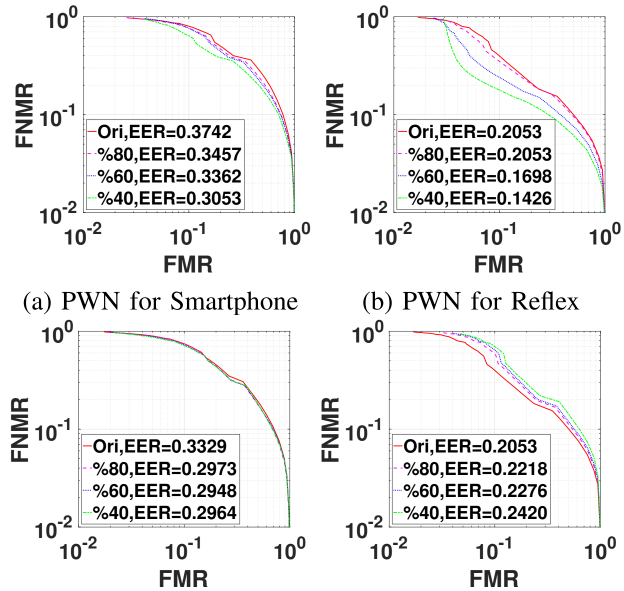
D. Approaches for the Evaluation of Iris Recognition System Performance

To evaluate the performance of iris recognition systems, many measures exist. An IQM is useful if it can at least give an ordered indication of an eventual performance [9]. Rank-ordered Detection Error Trade-off (DET) characteristics curve is one of the most commonly used and widely understood method used to evaluate the performance of quality assessment approaches. The DET curve used here plots False None Match Rate (FNMR) versus False Match Rate (FMR). Grother *et al.* [9] proposed to use quality-bin based approaches to evaluate the image quality assessment methods. They believe if a certain percentage of low quality samples are excluded from the dataset, the biometric system performance would become better and the Equal Error Rate (EER) (when FMR and False FNMR are equal) would decrease. We use it as one method to represent the performance of no-reference IQMs. Because the scale of the quality score for each IQM is different and the linearity of the score is unknown, thus, we omit the percentile low quality samples and keep 80%, 60%, and 40% of highest quality samples from each subject for each IQM [31].

IV. EXPERIMENTAL RESULTS

A. DET curve and EER

Here we obtain DET curve and EER as two indicators to examine the performance of IQMs. The interesting DET curves with EER for data with and without omitting low quality VW iris samples for three cameras by using selected



(a) PWN for Smartphone (b) PWN for Reflex

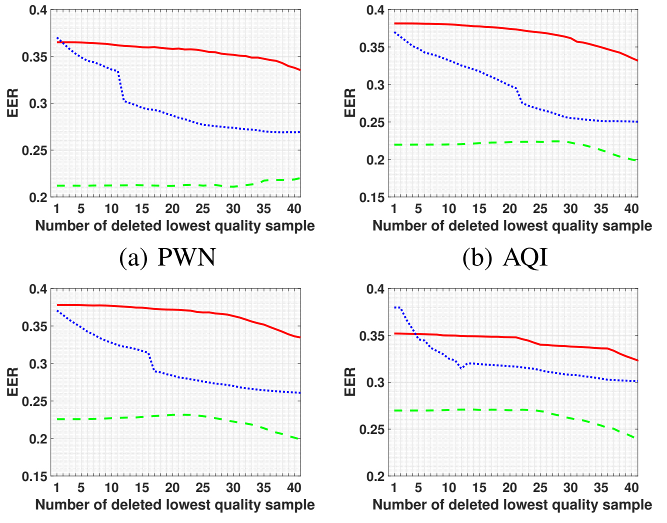
(c) BLIINDS2 for LFC (d) dipIQ for Reflex

Fig. 2. Examples of DET curves with EER for comparison score with and without omitting low quality samples.

IQMs are given in Fig. 2. For each sub-plot in Fig. 2, the red continuous line represents the original DET curve; the magenta '- -' line represents the DET curve when we keep 80% highest quality iris samples; the blue ':' line represents the comparison score when we keep 60% highest quality iris samples; and the green '-.' line represents the comparison score when we keep only 40% highest quality iris samples in the database for the experiment. If a DET curve is closer to the bottom-left point, it means that this set of data lead to a higher iris recognition performance. Meanwhile, the lower EER value the better system performance.

From Fig. 2 (a) and (b) we can see that, DET curves shift closer to bottom-left point when we keep 80%, 60%, and 40% highest quality samples by using the assessment results from PWN to omit low quality samples taken by smartphone and reflex camera, respectively. Especially we can see very obvious gap between each lines in Fig. 2 (b). It means that such IQMs can assess VW iris image quality and it is correlated with the performance of iris recognition algorithm. However, the DET curves have no obvious shift and the EER values have no significant changes by using the assessment results from BLIINDS2 when we omit low quality samples for LFC (see Fig. 2 (c)). In Fig. 2 (d), DET curves shift closer to top-right point and EER values increase when we keep 80%, 60%, and 40% highest quality samples by using the assessment results from dipIQ to omit low quality samples taken by reflex camera. This means that such IQMs have reversed correlation with the performance of iris recognition algorithm.

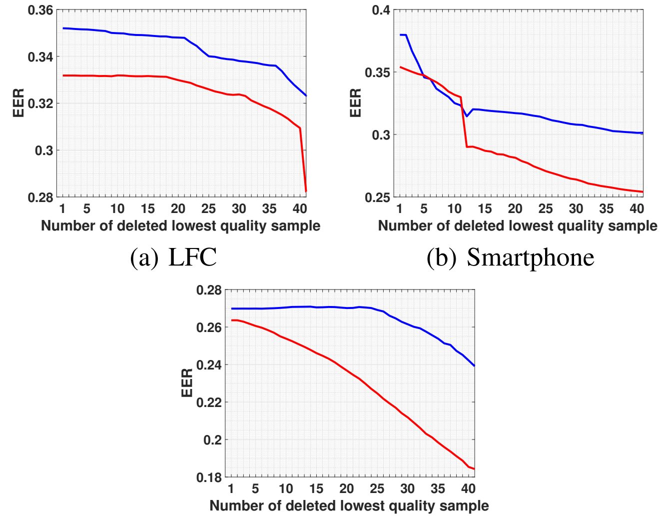
We also use EER values for all three cameras by omitting lowest quality iris sample one by one until only one highest quality iris sample left from each subject as another indicator to assess the performance of selected IQMs. Here we only



(a) PWN (b) AQI (c) AQIP (d) ILNIQE2

Fig. 3. Examples of EER values with omitting low quality iris samples one by one until the best quality sample left for 3 cameras. The red solid lines represent LFC camera, the blue ':' lines represent smartphone, and the green '–' lines represent reflex camera.

illustrate the four IQMs (PWN, AQI, AQIP, and ILNIQE2) that can assess iris quality for three cameras based on their performance from DET curves to discover the change of EER values. The x-axis in Fig. 3 represents the number of omitted lowest quality samples unit. There are 40 units per captured sample image per subject (eight distortions in five levels). Each unit has 1500 images (15 captured sample image per eye per subject). The y-axis represents the EER value. If the EER value has a smooth decreasing tendency when we omit lowest quality samples one by one, it means that the IQM used for generating the quality scores can predict the iris recognition algorithm well which represents the high performance of such IQM. The red solid lines represent LFC camera, the blue ':' lines represent smartphone, and the green '–' lines represent reflex camera. In Fig. 3 we can see that, by using the assessment results from the selected four IQMs to omit one lowest quality sample unit each time by using LFC and smartphone (red solid lines and blue ': lines'), the EER curves have decreasing tendency. However, from the green '–' lines in Fig. 3, the EER values from AQI and AQIP for reflex camera increase in the beginning and decrease in the end. There is no obvious change of the EER values for PWN in the beginning, but the EER values increase in the end. Only EER values from ILNIQE2 in Fig. 3) (green '–' line) have no increasing tendency and decrease when 25 low quality units are omitted. From the observation above we can summarize that, based on EER values with omitting low quality iris samples one by one until the best quality sample left for three cameras, ILNIQE2 can assess iris image quality for three cameras. The rest of the IQMs can either assess iris quality for only one or two cameras, or have low ability to assess iris quality based on



(a) LFC (b) Smartphone (c) Reflex

Fig. 4. Comparison of EER by omitting lowest quality sample one by one using ILNIQE2 for each subject between the original method and the re-trained method.

the system performance.

V. Re-training ILNIQE2 on VW iris database

From previous results, one specific metric, namely ILNIQE2 shows interesting results in terms of correlation between the provided quality scores and the performance results. Since this quality index has been trained on general purpose natural images, it would be interesting to investigate if results can be improved by retraining it on VW iris images. A recent research conducted by Liu *et al.* [32] shown that, by re-training the no-reference IQMs on biometric dataset can improve their performance on biometric samples. To perform the retraining, the UBIRIS .v2 database [33] has been selected, which the iris images were captured on non-constrained conditions (at-a-distance, on-the-move and on the visible wavelength). We use 241 images (one sample image per subject) in session one from the UBIRIS .v2 database to re-train the ILNIQE2 IQM. These 241 images are all high quality VW iris images because the ILNIQE2 metric only requires pristine images for training. The re-trained metric is then used to re-conduct the experiment removing lowest quality samples one by one from each subject. The plots of EER values for three cameras are shown in Fig. 4. The blue lines represent the original ILNIQE2 method, and the red lines represent the re-trained ILNIQE2 method. From Fig. 4 we can see that, after the re-training process, the overall performance of the IQM is improved because the red lines are under the blue lines. It means that the overall EER values from the re-trained method are lower than the original method. In addition, the improvements for LFC and reflex camera are greater than smartphone, especially for reflex camera. By using the original ILNIQE2 to omit lowest quality samples from the database, the EER values are not smoothly decreasing until 25 units of lowest samples are removed for reflex camera. However, by using the re-trained method, the line becomes smoother and has a overall decreasing tendency. Finally, the EERs reach close to 0.18 in the end (compared to 0.24 when using the original ILNIQE2). The difference of EERs between the original and the re-trained method for reflex camera is obvious.

VI. Conclusion

In this paper, we evaluated the performance of 15 selected no-reference IQMs for VW iris biometric images on GC^2 Multi-Modality Biometric Database. Two indicators are used to reflect the performance of IQMs according to the iris recognition algorithms: DET curve and EER value. We illustrated the results by comparing between indicators with and without omitting certain percentage of low quality iris samples. In addition, re-training an IQM by using only VW iris database has been done. From the experimental results we can conclude that, before the re-training process, ILNIQE2 has a better performance than the other selected IQMs to assess the quality of iris images based on the EER values for two cameras: LFC and reflex camera. The re-trained ILNIQE2 metric has better performance than its original version and now it can assess iris quality for all three cameras. Therefore, it is possible to use existing no-reference IQMs to assess the VW iris sample quality, moreover, the optimization process can further improve the performance of IQMs. One way to improve the performance of selected IQMs is to train them on iris databases, because the performance of IQMs on iris images may affected by the database used for training. The above mentioned findings can be used for the development of robust quality metrics for VW iris image quality, and furthermore, for multiple biometric modalities image quality assessment.

References

[1] J. R. Matey, D. Ackerman, J. Bergen, and M. Tinker, "Iris recognition in less constrained environments," in *Advances in Biometrics*. Springer, 2008, pp. 107–131.

[2] C. Fancourt, L. Bogoni, K. Hanna, Y. Guo, R. Wildes, N. Takahashi, and U. Jain, "Iris recognition at a distance," in *International Conference on Audio-and Video-Based Biometric Person Authentication, NY, USA, July*. Springer, 2005, pp. 1–13.

[3] X. Liu, M. Pedersen, and C. Charrier, "Image-based attributes of multi-modality image quality for contactless biometric samples," in *3rd International Conference on Signal Processing and Integrated Networks (SPIN), New Delhi, India, Feb.*, 2016, pp. 106–111.

[4] M. Pedersen and J. Y. Hardeberg, "Survey of full-reference image quality metrics," *Foundations and Trends in Computer Graphics and Vision*, vol. 7, no. 1, pp. 1–80, 2012.

[5] 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.

[6] ISO/IEC, "29794-6 information technology - biometric sample quality - part 6: Iris image data," 2012-10-10.

[7] X. Liu, M. Pedersen, C. Charrier, and P. Bours, "Can no-reference image quality metrics assess contactless visible wavelength iris sample quality?" in *Image Processing (ICIP), 2017 IEEE International Conference on, Beijing, China, September 2017*. IEEE, 2017, pp. 1–5.

[8] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "Quality measures in biometric systems," *IEEE Security & Privacy*, pp. 52–62, 2012.

[9] 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.

[10] S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric quality: a review of fingerprint, iris, and face," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 34, 2014.

[11] M. El-Abed, C. Charrier, and C. Rosenberger, "Quality assessment of image-based biometric information," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 15, 2015.

[12] "Lytro," *https://www.lytro.com/*, visited on 22/11/2016.

[13] X. Liu, M. Pedersen, and J. Y. Hardeberg, "CID: IQ–a new image quality database," in *International Conference on Image and Signal Processing, June, 2014, Cherbourg, France*. Springer, 2014, pp. 193–202.

[14] X. Liu, M. Pedersen, C. Charrier, P. Bours, and C. Busch, "The influence of fingerprint image degradations on the performance of biometric system and quality assessment," in *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*. IEEE, 2016, pp. 1–6.

[15] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, pp. 838–842, 2015.

[16] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[17] M. A. Saad, A. C. Bovik, and C. Charrier, "Dct statistics model-based blind image quality assessment," in *2011 18th IEEE International Conference on Image Processing, Brussels, Belgium, September*. IEEE, 2011, pp. 3093–3096.

[18] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[19] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579–2591, 2015.

[20] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb)," *IEEE transactions on image processing*, vol. 18, no. 4, pp. 717–728, 2009.

[21] Y. Han, X. Xu, and Y. Cai, "Novel no-reference image blur metric based on block-based discrete cosine transform statistics," *Optical Engineering*, vol. 49, no. 5, pp. 050 501–050 501, 2010.

[22] A. Leclaire and L. Moisan, "No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 145–172, 2015.

[23] K. Gu, W. Lin, G. Zhai, X. Yang, W. Zhang, and C. W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE transactions on cybernetics*, vol. 47, no. 12, pp. 4559–4565, 2017.

[24] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1.

[25] T. Zhu and L. Karam, "A no-reference objective image quality metric based on perceptually weighted local noise," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, pp. 1–8, 2014.

[26] S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *JOSA A*, vol. 24, no. 12, pp. B42–B51, 2007.

[27] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[28] L. Liu, B. Liu, H. Huang, and A. C. Bovik, "No-reference image quality assessment based on spatial and spectral entropies," *Signal Processing: Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.

[29] N. Othman, B. Dorizzi, and S. Garcia-Salicetti, "Osiris: An open source iris recognition software," *Pattern Recognition Letters*, p. 7, 2016.

[30] J. Daugman, "How iris recognition works," in *The essential guide to image processing*. Elsevier, 2009, pp. 715–739.

[31] N. Poh, S. Bengio, and A. Ross, "Revisiting doddington" s zoo: A systematic method to assess user-dependent variabilities," IDIAP, Tech. Rep., 2006.

[32] X. Liu, M. Pedersen, C. Charrier, and P. Bours, "Performance evaluation of no-reference image quality metrics for face biometric images," *J. Electron. Imaging 27(2), 023001 (2018), doi: 10.1117/1.JEI.27.2.023001*, 2018.

[33] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.