Identity Documents Image Quality Assessment

Daniel Schulz, Jose Maureira
R&D Center SR-226
TOC Biometrics
Santiago, Chile
daniel.schulz, jose.maureira@tocbiometrics.com

Juan Tapia, Christoph Busch
Hochschule Darmstadt
da/sec-Biometrics and Internet Security Research Group
Hochschule Darmstadt, Germany
juan.tapia-farias, christoph.busch@h-da.de

***Abstract*—In this work, a No-Reference method for assessing quality of ID Card images is developed, using a combination of Face Image Quality Assessment and Text Quality Assessment, motivated by the fact that face and text are two of the main components on an ID Card. For evaluation, a new private dataset was created, consisting of 12,960 Chilean ID Cards, with their corresponding printed facial reference image for face verification. The results obtained with the proposed quality assessment method, show that as more low quality images are discarded, a simultaneous improvement in face and text verification performance metrics is reached.**

***Index Terms*—Image quality assessment, IQA, identity documents, ID cards, ID card image quality assessment**

I. Introduction

Today, the current COVID-19 pandemic has increased the interest in remote identity verification. The broad use of smartphones to access different services, for example, banking, e-commerce, fintech, etc, raises a critical demand to have a robust remote automatic verification systems.

ID Card, passport and driver licence images are analysed using computer vision and Optical Character Recognition (OCR) techniques for obtaining automatically and remotely the reference information about the cardholder. This information can be the face reference image for face verification against a selfie, the text information that can be parsed, signature, etc.

However, all these images are captured in unconstrained scenarios, for example, a user only need to have an smartphone camera and internet access to activate a bank account, so the captured images can present a lot of variations such as different background, illumination, geometrical deformation, focus, specular reflection, blurring, etc. This can cause problems in the subsequent stages, for example, if a blurry image is captured, OCR algorithm could fail reading some important data, like the person's name or the national ID number. Therefore, a image quality assessment stage after the capture process can improve the performance, ensuring the subsequent processes operate on an image with sufficient quality.

According to the literature [1]–[5], image quality algorithms have focused on two main branches: Face Image Quality Assessment (FIQA) which analyses face images for the purpose of biometrics applications, and Image Quality Assessment (IQA) oriented towards general purpose perceptual quality. This perceptual quality can be objective or subjective.

In the FIQA area, different works have been developed, for example, FaceQNet v1 [6], MagFace [5], SER-FIQ [3]), SDD-FIQA [4]. An FIQA approach for assessing the quality of the face present in an ID Card was used. Also, we evaluate the text quality, using an OCR method with prediction scores to generate a quality metric for text. Then, the Levenshtein (edit) distance [7] was used to compare a set of predefined strings that we know are present in all ID Card instances.

The main contribution of this paper is a novel method for assessing document image quality, based on a combination of FIQA and text quality measurements.

II. Related work

Image quality is impacted by fidelity of the capture process describing the similarity of the image representation to its source, tasks related to the image —such as facial recognition or OCR—, or perceived subjective quality based on the previous meanings [8]. Much of the work on face image quality and ID-card image quality focuses on FIQA or general-purpose IQA [1], [2]. Some standards such as such as ISO/IEC-29794-5-WD4-FaceQuality[1] and ICAO 9303-p10-2021[2] describe how biometric sample quality assessments, including FIQA, should be performed [9]. FIQA is used to ensure that face pictures in the ID documents have been adequately taken, ensuring high sample quality [10].

In the FIQA literature, Schlett et al. [1] surveyed over 50 works. Some of the methods employ measurements that are specific for evaluating faces, *i.e.* pose, and facial expression. Additionally, the scores yielded by FIQA algorithms are usually intended to predict facial recognition performance. These two factors may prevent FIQA algorithms from being usable for other types of IQA.

Zhou et al. [11] investigate the robustness of face detection algorithms on low-quality images at different levels of blur, noise and contrast. Specifically, they evaluate four representative face detection models, showing that both hand-crafted and deeply learned features are sensitive to low-quality input.

This research work has been partially funded by the TOC Biometrics R&D center SR-226, and the German Federal Ministry of Education and Research, in conjunction with the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 883356.

[1] https://isotc.iso.org/livelink/livelink?func=ll&objId=22179738&objAction=Open

[2] https://www.icao.int/publications/documents/9303_p10_cons_en.pdf

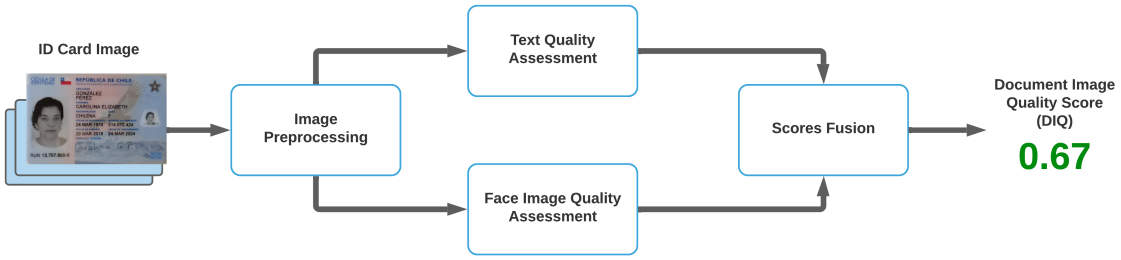

Figure 1. Block diagram for the Document Image Quality estimation.

Meng et al. [5] proposed new method called MagFace, with a category of losses that learn an universal feature embedding, whose magnitude can measure the quality of the given face image. Under the new loss, it can be proven that the magnitude of the feature embedding monotonically increases if the subject is more likely to be recognised.

Terhorst et al. [3] proposed a novel unsupervised face quality assessment concept called "SER-FIQ" by investigating the robustness of stochastic embeddings. This solution measures the quality of an image based on its robustness in the embedding space.

Ou et al. [4] proposed a method that generates quality pseudo-labels by calculating the Wasserstein Distance between intra and inter classes similarity distributions. With these quality pseudo-labels, they are capable of training a regression network for quality prediction. Extensive experiments on benchmark datasets demonstrate that the proposed SDD-FIQA surpasses the state-of-the-art.

In our previous work [12], a system for predicting subjective image quality on ID-card images was developed. This goal was accomplished studying multiple features extracted for ID image quality assessment and predicting subjective quality scores. An ID-card subjective IQA dataset was generated, surveying 15 subjects on the quality of 204 images.

For text, we focused on OCR and scene text recognition. In [13] a novel end-to-end trainable framework named Semantic Reasoning Network (SRN) for accurate scene text recognition was proposed with a global semantic reasoning module (GSRM) to capture global semantic context through multi-way parallel transmission.

Shi et al. [14] proposed an architecture integrating feature extraction, sequence modeling and transcription into an unified framework. It combines Convolutional Neural Networks(CNN) and Recursive Neural Network (RNN) to extract sequential visual features from a text image, and then fed them into a decoder to predict the best character category.

Xie et al. [15] proposed a new loss function called aggregation cross-entropy, aimed to the sequence recognition. It optimises the statistical frequency of each character along the time dimension, improving the efficiency.

III. Method

The method proposed in this work implements a fusion of FIQA and text scores, aimed to obtain a No-Reference quality score for ID Cards. This is motivated by the fact that two of the main components in an ID Card layout are the face and the text. The former aimed to perform face verification, and the later to parse the textual information, that can be validated against a government institution, for example. For this purpose, we used public implementations for the MagFace FIQA method [5], SDD-FIQA [4] and EasyOCR for text reading[3].

A block diagram of the quality estimation method is shown in Figure 1.

A. Text Image Quality Score

For measuring the text quality, an approach based on OCR was used. As an initial step, we scale the previously cropped images to enclose the ID Card only to a fixed resolution of 1024×768 px. It was applied a histogram equalisation using the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm [16] to the V channel in the HSV space. Finally, we apply an OCR method, take all the detections, filter them by the bounding box height (20 px minimum), and then average all the text prediction scores obtained by the OCR, as is defined in equation 1:

$$score_{ocr} = \frac{1}{|S|} \sum_{s \in S} score(s) \tag{1}$$

where S is the set of detected strings whose bounding box height is greater than 20 pixels, and $|S|$ is their cardinality. Figure 2 shows an example of the detected strings, and the bounding boxes encompassing them. The colour of the bounding boxes represents the text prediction score, using a heat map with jet colormap, where blue represent low scores and red represent high scores.

B. Face Image Quality Assessment Score

For extracting FIQA scores, we used two state-of-the-art methods: MagFace [5] and SDD-FIQA [4]. As a first step, we cropped the faces, using the MultiTask Cascaded Convolutional Network (MTCNN) method [17] for face detection and alignment. Then, the cropped face is processed by the FIQA method, obtaining a No-Reference score. Scores ranges from 0 to 40 in the case of MagFace, and from 0 to 100 for SDD-FIQA. For both methods, a low score represent poor quality, and a high score represent better predicted quality.

[3] https://github.com/JaidedAI/EasyOCR

Figure 2. Example of an ID Card image with low FIQA scores and high OCR score. The low FIQA score is caused by the reflection in the left portion of the face.

C. Text and Face qualities combination

For obtaining an ID Card quality score, we first scale the FIQA scores to the interval $[0, 1]$ for being combined with the text scores. Then, we perform a linear combination between the FIQA Score and the Text Score, as shown in equation 2. Both scores are in the interval $[0, 1]$, using a weight factor α that ranges from 0, where full OCR score is used, up to 1.0, where full FIQA score is used.

$$final\_score = \alpha \times score_{fiqa} + (1 - \alpha) \times score_{ocr} \quad (2)$$

An interesting case is shown in Figure 2, where a high text score ($score_{ocr} = 0.87$) is obtained, but in the other hand, a low FIQA score ($score_{MagFace} = 20.98, score_{SDDFIQA} = 36.26$) was obtained. These scores are explained by the reflection present in the middle left portion of the image, that affects the quality estimated by the FIQA methods.

D. Text Verification Metric

A text verification scheme as an evaluation metric for text data was performed. For this purpose, first was defined a set of common strings present in all documents instances. In our case, we have only one type of ID Card (CHL2), so this set is bounded to a few known strings. These strings are in Spanish, for example, 'NOMBRES' (Names), 'APELLIDOS' (Last-names), 'NACIONALIDAD' (Nationality), etc. Then, it calculates a similarity measure between them and each of the strings detected by the OCR for each string. Finally, all similarities are averaged to obtain a final score.

Formally, we have the set of searched strings, $S = \{s_1, s_2, ..., s_N\}$, and the set of strings returned by the OCR method $O = \{o_1, o_2, ..., o_M\}$. Firstly, for each string $s \in S$, the Levenshtein distance $ld(s, o)$ between searched s and each of the OCR strings $o \in O$ is calculated. Then, it was divided by the length of the searched string $|s|$, clipping at 1. Then, look for the closest string, where the $d(s, o)$ is minimal, as seen in equation 3. In the end, the final distance is calculated, averaging all the individual distances for the searched strings $s \in S$.

$$d(s, o) = \min\left(1, \frac{ld(s, o)}{|s|}\right) \quad (3)$$

$$d(s) = \min_{o \in O} d(s, o) \quad (4)$$

$$d_{ocr} = \frac{1}{N} \sum_{s \in S} d(s) \quad (5)$$

IV. Experiments and Results

A. Database

In this paper, a new private dataset was created, called ID-CARD-CLv2, with images captured in a remote verification system. It consists of 12,960 front images of Chilean national identity cards, distributed in in 6,459 females and 6,501 males, with resolutions varying between 2634×1280 and 181×268 pixels. Also, it contains 12,960 images of face selfies, corresponding to the same subjects from the ID-cards images. All images were captured using several smartphone and tablet models. Examples of the faces, corresponding to the ID Card (top) and selfie (bottom) are shown in Figure 3. Figure 4 shows examples of Chilean ID Cards. Due to data protection regulations, we cannot publicly release the dataset.

Figure 3. Examples from ID-CARD-CLv2 dataset. Upper row show faces from ID Cards, and bottom row show selfies. Each column represents the same subject.

Figure 4. Examples of ID-card images from ID-CARD-CLv2 dataset. A black tag was added in order to cover sensible data.

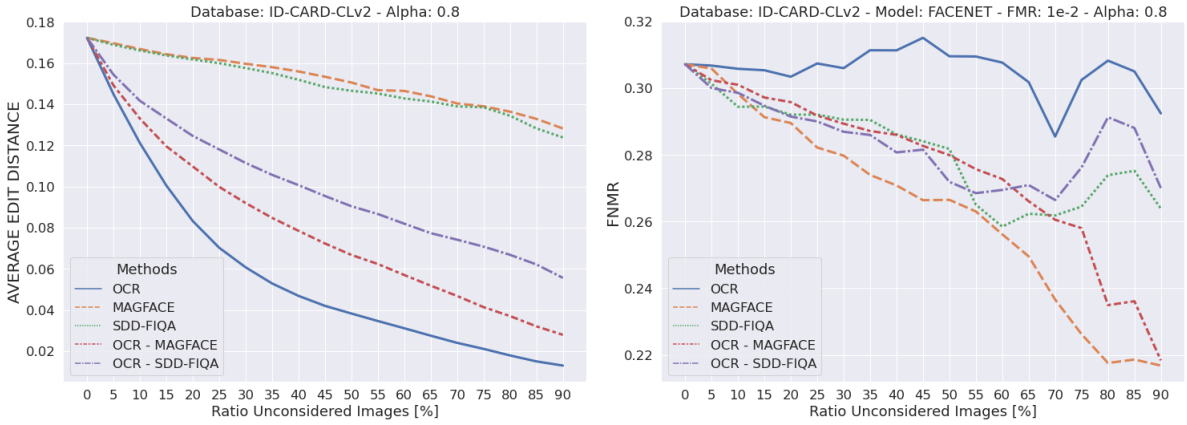

Figure 5. Performance evaluation discarding a ratio of the lower estimated quality images using $\alpha = 0.8$: Left: Text Verification, Right: Face verification.

B. Experiment 1 - Face verification

This experiment evaluates how the face verification performance changes as a function of the estimated ID Card image quality, defined in section III-C. For this goal, images with the lowest quality score are removed progressively in 5% increments, ranging from 0% to 90% of the lower quality score images, and evaluate the face verification performance on each subset. Performance is measured using the Error versus Reject Curve (ERC), which show the effectiveness of rejecting low-quality facial images in terms of FNMR reported at a fixed FMR, 0.01 in our case. The face verification algorithm used was FaceNet [18], with cosine distance.

C. Experiment 2 - Text verification

In this experiment, it was evaluated how the text verification performance, as defined previously in section III-D, changes accordingly to the ID Card estimated quality, described in section III-C. For this goal, images with the lowest ID Card quality score are removed progressively in 5% increments, ranging from 0% to 90% of the lower quality score images. Then, the text verification performance was evaluated, as defined previously, on each image belonging to the subsets with the remaining data. For obtaining an aggregated metric for each subset, we averaged the distances obtained using the text verification performance defined in section III-D

D. Results analysis

Table I shows the summary results for the face and text verification, with ratios of unconsidered images ranging from 0% up to 50%. Different alpha values were explored from 0.0 up to 1.0, in steps of 0.2, generating different curves showing how the performance changes as a function of the ratio of unconsidered images. It is essential to highlight that this analysis considers the trade-off between the FNMR for face verification and the Average edit distance for text verification. The best results will be closely related to the operational point selected according to the number of images discarded. An operating point that shows a good trade-off for both tasks is when $\alpha = 0.8$, as can be seen in Figure 5, especially for the OCR-MagFace method. In this case, for both tasks, the OCR-MagFace curves are closer to the best results obtained for each task, which are the full OCR for text verification and full MagFace for face verification. For quantitatively chosen the operational point, a metric for the curves in the whole domain needs to be developed, for example, the minimal area under the curve.

In Figure 5 right, when using only the text quality estimation ($\alpha = 0$, Solid blue line), performance does not improve for the face verification task, even for some ratios, it has worse results than the case where no images are discarded. These observations lead us to conclude that is not convenient to used text quality estimation in the context of face verification.

In Figure 5 left, the performance for the full FIQA method ($\alpha = 1$, dashed orange line for MagFace, dotted green line for SDD-FIQA), applied to the text verification task, shows an steady decay in the average edit distance, obtaining consistently better results when more images are discarded according to their FIQA quality. This implies that applying a FIQA method for discarding some low score images, can improve the results for text verification.

V. Conclusions

In this work, we implemented No-Reference approach for ID Card quality estimation, using a combination of FIQA and Text Quality. This is motivated by the fact that face and text are two of the main components in an ID Card. FIQA methods used were MagFace and SDD-FIQA, and for Text Quality, we used the average prediction score given by an OCR method, in our case, EasyOCR. Experiments were performed on a private dataset, consisting of 12,960 Chilean ID Cards (And selfies for face verification). Results show that this fusion approach, using an appropriate α, works well for the face verification and text verification tasks, improving the results as more images with low quality score are discarded. The results obtained can be used to discard low quality ID Cards in a remote Face Verification system, for example. Future work can be focused on evaluating other combinations of FIQA and Text Recognition methods, or using non-linear fusion schemes,

Table I
SUMMARY RESULTS FOR THE FACE AND TEXT VERIFICATION RESULTS.

Ratio Unconsidered Images	α	Face Verification FNMR@FMR 0.01		Text Verification Average Edit Distance	
		OCR-MagFace	OCR-SDD-FIQA	OCR-MagFace	OCR-SDD-FIQA
0%	-	0.307	0.307	0.172	0.172
10%	0.0	0.306	0.306	0.121	0.121
	0.2	0.305	0.304	0.121	0.121
	0.4	0.306	0.303	0.121	0.123
	0.6	0.305	0.302	0.121	0.123
	0.8	0.301	0.299	0.133	0.142
	1.0	0.298	0.294	0.167	0.166
20%	0.0	0.303	0.303	0.083	0.083
	0.2	0.305	0.301	0.083	0.084
	0.4	0.303	0.298	0.084	0.087
	0.6	0.301	0.297	0.090	0.098
	0.8	0.296	0.291	0.110	0.125
	1.0	0.290	0.292	0.162	0.162
30%	0.0	0.306	0.306	0.061	0.061
	0.2	0.306	0.303	0.061	0.061
	0.4	0.303	0.295	0.063	0.065
	0.6	0.302	0.290	0.069	0.079
	0.8	0.289	0.287	0.092	0.111
	1.0	0.280	0.291	0.160	0.158
40%	0.0	0.311	0.311	0.047	0.047
	0.2	0.312	0.304	0.047	0.047
	0.4	0.307	0.296	0.048	0.052
	0.6	0.300	0.289	0.054	0.066
	0.8	0.286	0.281	0.078	0.101
	1.0	0.271	0.286	0.156	0.152
50%	0.0	0.310	0.309	0.038	0.038
	0.2	0.306	0.310	0.038	0.038
	0.4	0.303	0.301	0.039	0.042
	0.6	0.294	0.284	0.043	0.055
	0.8	0.280	0.272	0.067	0.090
	1.0	0.266	0.282	0.151	0.147

that can improve results for both Face Verification and Face Verification tasks. Also, the proposed ID Card quality score can be used as a target for training a Convolutional Neural Network, generating a semi-referential approach.

REFERENCES

[1] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Comput. Surv.*, dec 2021. [Online]. Available: https://doi.org/10.1145/3507901

[2] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, no. 11, pp. 1–52, 2020.

[3] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020, pp. 5651–5660.

[4] F.-Z. Ou, X. Chen, R. Zhang, Y. Huang, S. Li, J. Li, Y. Li, L. Cao, and Y.-G. Wang, "SDD-FIQA: Unsupervised face image quality assessment with similarity distribution distance," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[5] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "Magface: A universal representation for face recognition and quality assessment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2021, pp. 14 225–14 234.

[6] J. Hernandez-Ortega, J. Galbally, J. Fierrez, and L. Beslay, "Biometric quality: Review and application to face recognition with FaceQnet," *arXiv preprint arXiv:2006.03298*, 2021.

[7] V. I. Levenshtein *et al.*, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet physics doklady*, vol. 10, no. 8. Soviet Union, 1966, pp. 707–710.

[8] F. Alonso-Fernandez, J. Fierrez, and J. Ortega-Garcia, "Quality Measures in Biometric Systems," *IEEE Security and Privacy*, vol. 10, no. 6, pp. 52–62, 2012.

[9] ISO/IEC JTC1 SC37 Biometrics, "ISO/IEC 29794-1:2016 Information technology — Biometric sample quality — Part 1: Framework," International Organization for Standardization, Geneva, CH, Standard, Sep. 2016.

[10] P. Grother, M. Ngan, and K. Hanaoka, "Face recognition quality assessment concept and goals," *National Institute of Standards and Technology (NIST)*, 2019.

[11] Y. Zhou, D. Liu, and T. Huang, "Survey of face detection on low-quality images," in *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 2018, pp. 769–773.

[12] C. Yáñez and J. Tapia, "Image quality assessment on identity documents," in *2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2021, pp. 1–5.

[13] D. Yu, X. Li, C. Zhang, T. Liu, J. Han, J. Liu, and E. Ding, "Towards accurate scene text recognition with semantic reasoning networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 113–12 122.

[14] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 11, pp. 2298–2304, 2016.

[15] Z. Xie, Y. Huang, Y. Zhu, L. Jin, Y. Liu, and L. Xie, "Aggregation cross-entropy for sequence recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6538–6547.

[16] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.

[17] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[18] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June-2015, 2015.