A Skin Tone Annotated Face Image Dataset for Studying Demographic Variability

1st Christoph Busch
Hochschule Darmstadt
Darmstadt, Germany
christoph.busch@h-da.de

2nd Paul Kibler
Hochschule Darmstadt
Darmstadt, Germany

3rd Fabian Stockhardt
Hochschule Darmstadt
Darmstadt, Germany
fabian.stockhardt@h-da.de

4th Christian Rathgeb
Hochschule Darmstadt
Darmstadt, Germany
christian.rathgeb@h-da.de

***Abstract*—Face recognition systems have achieved substantial performance improvements over the past decade. However, their effectiveness remains strongly dependent on the quality of the acquired biometric samples. To assess the quality of face images and in order to verify compliance with ICAO regulations, algorithms and standards have been developed. Recent investigations have shown that some of the proposed quality measures exhibit significant demographic variability related to skin tone. As a result, the development of unbiased quality assessment algorithms remains an open challenge.**

Progress towards reducing demographic bias in quality assessment requires face image datasets with reliable ground truth information. This work presents the DAST-Dataset, that contains portrait images from a large variety of capture subjects (light skin to dark skin). Face images are captured under controlled conditions and include ICAO compliant samples as well as systematically overexposed and underexposed images for each subject. In addition, for each subject multiple ground truth measurements for the skin tone are included. The dataset is intended to support the evaluation of demographic bias in face image quality measures and to facilitate the development and benchmarking of skin tone classification.

***Index Terms*—face recognition, face image quality, database, demographic variability, skin tone**

I. Introduction

Face recognition is widely adopted today and plays a key role in a broad range of applications, including authentication on personal smart devices (e.g., mobile phones), access control scenarios (e.g., border crossing), and forensic applications such as video surveillance. These applications constitute relevant operational biometric systems. To achieve high recognition accuracy, the quality of the acquired biometric samples is of central importance. High similarity scores can only be obtained when both the reference and probe samples meet sufficient quality requirements.

Before a sample is stored in a reference database or submitted to a comparison process, operators must verify whether the captured sample complies with the definition of a canonical face image. For two-dimensional face images, such a definition is specified in the International Standard ISO/IEC 39795-5:2019 in Annex D.1 [1]. To support this process, face image quality assessment (FIQA) algorithms have been developed that measure the utility of a given sample and eventually indicate a potential defect (e.g., the image is not focused) back to the biometric attendant or the capture subject as actionable feedback [2], [3]. With this information, a subsequent recapture attempt could be improved.

Previous studies indicate that many established quality measures exhibit little or no demographic variability (DV) with respect to demographic attributes such as gender or age, suggesting fair treatment across population subgroups [4]. However, certain quality measures derived from the luminance show pronounced DV as they depend directly on the skin tone [4]–[6]. Since the long-term success of face recognition systems depends on their acceptance by the target population, biometric algorithms, including FIQA, must treat different demographic groups fairly. Despite ongoing research efforts, the development of face image quality assessment algorithms that are robust to demographic variability remains an open challenge [7].

This work intends to support the further development of quality measures with the explicit objective of minimising demographic bias. To this end, we provide a dedicated database of face images complemented by reliable and systematically acquired ground truth information. We present the DAST-Dataset[1], which consists of portrait face images from subjects spanning a wide range of skin tones, from light to dark. All face images were captured under controlled conditions. For each subject, ICAO-compliant samples were acquired using appropriate illumination, along with intentionally non-compliant samples exhibiting slight and strong overexposure as well as slight and strong underexposure. Face images were taken with multiple capture devices. In addition, multiple skin tone ground truth measurements were obtained for each subject at different measurement points on the hands and face. These measurements were aggregated to assign a skin tone class to each subject.

The dataset will be released incrementally. Upon acceptance of this paper, an initial subset, *DAST-10*, comprising up to 10 subjects per skin tone class, will be made publicly available.

This research work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. Further this work was partially supported by EU Horizon under G.A. EINSTEIN (101121280) and CarMen (101168325

[1]DArmstadt Skin Tone (DAST) database: https://github.com/dasec/DAST-SkinTone-database

Data collection is ongoing, and future releases, *DAST-20* and *DAST-30*, will follow as additional subjects are captured. The dataset will be useful for evaluating FIQA algorithms, as well as for development of skin tone estimation algorithms and calibration thereof.

The remainder of this paper is structured as follows. Section II, will review related work on face image quality and the demographic variability of quality measures. In Section III the face image capture device and settings will be documented and the measurement of skin tone will be discussed with example data. In Section IV, we analyse the statistics of the *DAST-10* dataset and the meta data and will close some experiments on the collected data. Limitations are outlined in Section V and finally, Section VI concludes the paper.

II. Related Work

In order to control the quality of face images, capture requirements have been formulated in the international standards ISO/IEC 19794-4:2011 [8] and ISO/IEC 39794-5:2019 [1]such as sufficient spatial resolution [9], a full frontal perspective, adequate contrast, and proper illumination. If a captured face image does not comply with these requirements, biometric recognition performance can be significantly degraded, resulting in a reduced probability of correctly recognising the capture subject. In general, biometric performance is sensitive to pose variations, illumination changes, sensor conditions, and other disturbance factors that negatively impact image quality.

Recently, strong innovation has been observed in the development of FIQA algorithms. A major driving factor is the recent launch of the European Entry Exit System (EES) [10], which requires that EU member states conduct the biometric enrolment at border control points in accordance with Implementing Decision 2019/329 [11]. Despite these technical and regulatory advances, substantial research effort is required to develop FIQA algorithms that not only ensure sufficient quality of stored biometric samples but also yield comparable quality score distributions across different demographic groups.

A. Face Image Quality

A comprehensive overview of FIQA algorithms is provided in [12]. The vast majority of methods described in this survey focus on unified quality scoring approaches that aim to quantify the utility of an image for face recognition. Such algorithms should have predictive power, meaning that a low quality score indicates a low similarity score to be expected, if that image is used in a biometric comparison trial. Also complementary measures are needed that allow actionable feedback to the capture subject (e.g., guidance on pose correction) or to the biometric attendant (e.g., information on image sharpness or focus of the capture device).

These requirements were explicitly addressed in the standardisation process for a unified quality score (UQS) and complementary quality components, resulting in the publication of ISO/IEC 29794-5 [3]. The FIQA algorithms specified in this standard assess whether a face image complies with the definition of a canonical face image in accordance with the ICAO Machine Readable Travel Document (MRTD) specifications [13]. This definition is formalised in the Biometric Data Interchange Standard ISO/IEC 19794-5:2011 as *frontal image type* [8] and is further refined in the Extensible Biometric Data Interchange Standard ISO/IEC 39794-5:2019, Annex D.1 [1].

The UQS defined in ISO/IEC 29794-5:2025 is a holistic measure for the entire sample intended to be predictive of face recognition performance. It is represented as an integer value in the range of 0 to 100, with higher values indicating higher expected utility. Alongside the standard, the Open Source Face Image Quality (OFIQ)[2] project provides a reference implementation of the specified algorithms. This open-source software is suitable for deployment in both commercial and governmental biometric systems [2]. The algorithms implemented in OFIQ are mostly based on deep learning algorithms such as MagFace [14] and were selected based on their performance in the NIST Face Analysis Technology Evaluation (FATE), Part 11: Face Image Quality Vector Assessment – Specific Image Defect Detection [15].

In addition to the UQS, ISO/IEC 29794-5:2025 specifies a set of FIQA algorithms designed to provide actionable feedback during the capture process. These assess specific properties of the biometric sample, such as pose angle, and evaluate compliance with canonical face image requirements (e.g., near-frontal perspective with minimal pose deviation). Beyond subject-related characteristics, the standard also includes capture device–related quality measures that support system setup and calibration.

B. Demographic Variability of Face Image Quality Assessment

It is important to minimise measurable demographic bias prior to the deployment of a face recognition system. [16]. The same holds true for FIQA algorithms that are integrated in a capture process [4]. An overview of the effect of algorithmic bias in biometric systems and a survey on the recent literature is given in [17]. Beyond investigations regarding differential performance related to demographic variables such as the gender, ethnicity and facial morphology, increasing attention has also been paid to performance differences associated with skin tone. This aspect is especially relevant in face analysis, as low image brightness in the facial region may result either from poor illumination conditions (e.g., underexposure) or from the capture subject's darker skin pigmentation. Distinguishing between these factors requires reliable measurement of skin tone.

Skin tone is commonly mapped to categorical representations to facilitate DV analysis and reporting. Established schemes include the Fitzpatrick Skin Types (FST) [18], the more recently proposed Monk Skin Tone Scale (MST) [19], and the Colorimetric Skin Tone Scale (CST) [20]. Categorical representations enable standardised reporting of DV, which is often visualised using violin plots to illustrate quality score

[2]OFIQ on GitHub: https://github.com/BSI-OFIQ/OFIQ-Project

distributions for the overall population and for demographic subgroups. Moreover, categorical labelling allows the computation of quantitative DV metrics for FIQA, as proposed in [21] and applied in [22].

The quality measures defined in ISO/IEC 29794-5 have been investigated by Kabbani et al. on publicly accessible datasets [4] and more recently by Utcke et al. on face images stemming from operational data [6]. These works became contributions to the technical report ISO/IEC WD3 25722 [5].

From the quality measures that are contained in ISO/IEC 29794-5:2025, a subset was considered as sensitive to face skin tone and therefore specifically associated with high DV. This subset includes the measures *Luminance mean, Luminance variance, Under-exposure prevention, Over-exposure prevention, Natural color*, which indeed have shown strong DV [4], [6], [22]. However, Dörsch et al. have shown that skin tone-related bias can be reduced if skin tone-balanced datasets are used during the training of FIQA algorithms [7].[3]

Unfortunately, open accessible datasets with labeled skin tone information are either extremely limited in the number of contained subjects like the dataset MST-E[4] or are of extremely low resolution for most samples, which constitutes the dominating defect [23]. With the provision of the DAST-datasets we are confident to provide a database for DV analysis of FIQA algorithms.

While good progress was obtained using synthetic or semi-synthetic data [24] the need for real face images with ground truth labels is obvious. The dataset that we provide can also validate skin tone estimation capabilities of current state-of-the-art, which will enable one to transfer more accurate skin tone labels to larger already existing dataset.

III. Data Acquisition

Data collection is conducted at a fixed photo studio located on a university campus. This restricts participant recruitment primarily to students and local populations, resulting in limited geographic and demographic diversity. Although skin tone classes are intentionally populated, their distribution might not reflect natural population statistics. Each participating volunteer was informed about the purpose of the data collection and signed a consent form.

A. Face Image Capture Device

Two commercial off-the-shelf cameras were used to capture the portrait face images:

- A Canon EOS 50D equipped with a Canon EF-S 18–200 mm f/3.5–5.6 IS lens, set to a focal length of approximately 90 mm.
- A Sony ILCE-7M2 equipped with an FE 90 mm F2.8 Macro G OSS lens.

The resulting image resolutions were 3,168×4,752 px and 4,000×6,000 px for the Canon and Sony cameras, respectively.

[3]FIQA algorithms with reduced DV can be included in the revision of ISO/IEC 29794-5: https://www.iso.org/standard/92472.html

[4]MST-E Dateset with only two subjects per class: https://skintone.google/mste-dataset

TABLE I: Settings for face image acquisition.

EL	Flash settings	Camera settings
EL1	small SB 1/8 +0.3 large SB 1/8 BG 1/64	SS 1/100 iso 100 F16
EL2	small SB 1/8 +0.3 large SB 1/8 BG 1/64	SS 1/100 iso 100 F22
EL3	small SB 1/8 +0.3 large SB 1/8 BG 1/64	SS 1/100 iso 100 F11
EL4	small SB 1/4 +0.3 large SB 1/4 BG 1/32	SS 1/125 iso 100 F11
EL5	small SB 1/2 +0.3 large SB 1/2 BG 1/16	SS 1/160 iso 100 F11

SB: Softbox, BG: Background, SS: Shutter speed

The capture setup is illustrated in Figure 1. Subjects were seated on a chair at a defined distance from the camera in accordance with ISO/IEC 39794-5:2019, Annex E [1].

All images were captured in the sRGB color space and stored using lossless compression in Tagged Image File (TIF) format. White balance calibration was performed prior to the first capture session of each day using a standard white balance card. Illumination in the photo studio was controlled using professional photoflash equipment, allowing precise manipulation of the capture conditions and enabling the acquisition of both normally exposed and overexposed face images. Underexposure was achieved by reducing the effective exposure through aperture adjustment. Each photoflash facing the capture subject was equipped with a softbox to ensure compliant and homogeneous illumination in accordance with ISO/IEC 39794-5:2019, Annex E [1].

Fig. 1: Illustration of the face image capture conditions.

By combining camera and flash settings, as summarised in Table I, five distinct exposure levels (ELs) were realised:

- **EL1**: strongly underexposed, approximately 25% of the nominal exposure (two stops underexposed)
- **EL2**: underexposed, approximately 50% of the nominal exposure (one stop underexposed)
- **EL3**: nominal exposure (reference exposure level)
- **EL4**: overexposed, approximately 200% of the nominal exposure (one stop overexposed)
- **EL5**: strongly overexposed, approximately 400% of the nominal exposure (two stops overexposed)

As illustrated in Figure 1, two softboxes were used to reduce specular reflections and achieve homogeneous facial illumination. The larger softbox was a Nanlite SBPR90Q circular softbox with a diameter of 90 cm, while the smaller softbox was a Sirui Quick Release softbox measuring 60×90 cm. In total three photoflashes were used each with a max power of 400ws. The background consisted of a standard white paper backdrop with an average RGB color value of (241, 234, 216).

B. Skin Tone Measurement Device

Skin tone measurements were obtained using the handheld DSM-4 colorimeter, which provides accurate measurements of skin color, individual typology angle (ITA), and pigmentation[5]. The colorimeter measures a skin area of approximately 50 mm^2 under standardised D65 illumination generated by four light-emitting diodes. Skin color is quantified using diffuse reflectance spectroscopy.

For each volunteer, five skin tone measurements were collected. Facial measurements were taken at the forehead and at the left and right cheeks (zygomatic region). In addition, measurements were obtained from the interosseous region of the left and right hand.

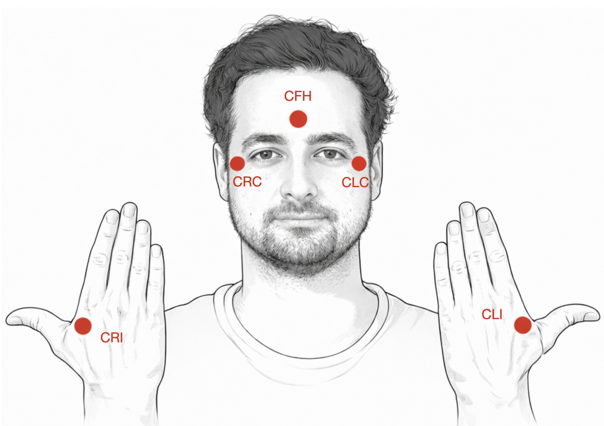

Fig. 2: Locations of the colorimeter measurements.

Skin tone measurements using the colorimeter were performed at five predefined anatomical locations, as illustrated in Figure 2:

1) **CFH**: Colorimeter measure forehead of the face
2) **CLC**: Colorimeter measure forehead of the left cheek
3) **CRC**: Colorimeter measure forehead of the right cheek
4) **CLI**: Colorimeter measure left hand (interosseus muscle[6])
5) **CRI**: Colorimeter measure right hand (interosseus muscle)

Measurements at these locations are inspired by earlier work [20], [23] and enable combined and location-specific analysis of skin color, as pigmentation may vary depending on anatomical region and exposure to sunlight. The subjects' skin was not cleaned prior to measurement in order to reflect natural skin conditions. The colorimeter was calibrated once per day prior to the first measurement using the calibration plate provided by the manufacturer. All measurements were recorded in the CIELAB color space.

For hygiene and consistency, the skin-contacting surfaces of the colorimeter were disinfected with an alcohol wipe after each participant.

[5]DSM-4 Colorimeter: https://www.cortex.dk/skin-analysis/colorimeter-dsm-4/

[6]Interossei muscles: https://en.wikipedia.org/wiki/Palmar_interossei_muscles

C. Skin Tone Classes

For the assignment of data subjects to preliminary skin tone classes, we follow the mapping proposed by Cook et al. [20], which defines 10 mutually exclusive classes. The corresponding class centroids are publicly available[7].

The class assignment is performed by computing the average CIELAB values (L*, a*, b*) obtained from the three facial measurement locations CFH, CLC, and CRC. Each subject is then assigned to the class whose centroid exhibits the smallest distance to the averaged color vector.

D. Example Data

All captured images were manually cropped to the facial region of interest of approx. 2,300×2,500 px. The distributed dataset contains the cropped images.

The Figure 3 shows as example 12 face image that were taken for each subject. The four images in the middle row are with normal (good) illumination. Images above are over-exposed and images on the bottom are under-exposed. The left column contains the images from Canon-EOS-50D and the right column represents images from SONY-ILCE-7M2 capture device.

IV. Dataset Analysis

A. Dataset Statistics

The number of subjects per gender and CST class of the captured database is summarised in Table II. It can be seen, that the lightest CST class contains only two subjects, while the darkest CST class is empty. Further, it can be observed that the database contains more male than female subjects.

TABLE II: Number of subjects per gender and skin tone class

	CST classes									
Gender	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Male	0	4	8	9	9	8	6	9	8	0
Female	2	6	2	1	1	2	4	1	0	0

B. Demographic Variability of Unified Quality Measure

The violin plot in Figure 4 visualises the distribution of the UQS across demographic groups. The width of each violin reflects the kernel density of the scores, while the internal lines indicate the median and interquartile range.

UQSs are generally high due to the controlled acquisition conditions. It was observed that even for over- and under-exposed face images UQSs are still high, since facial traits are largely visible. Nonetheless, a clear variation of quality score distributions across classes is observed, indicating DV

[7]Classes in CST: https://github.com/TheMdTF/mdtf-public/blob/master/datasets/skintone/data/colorimetric_scale.csv

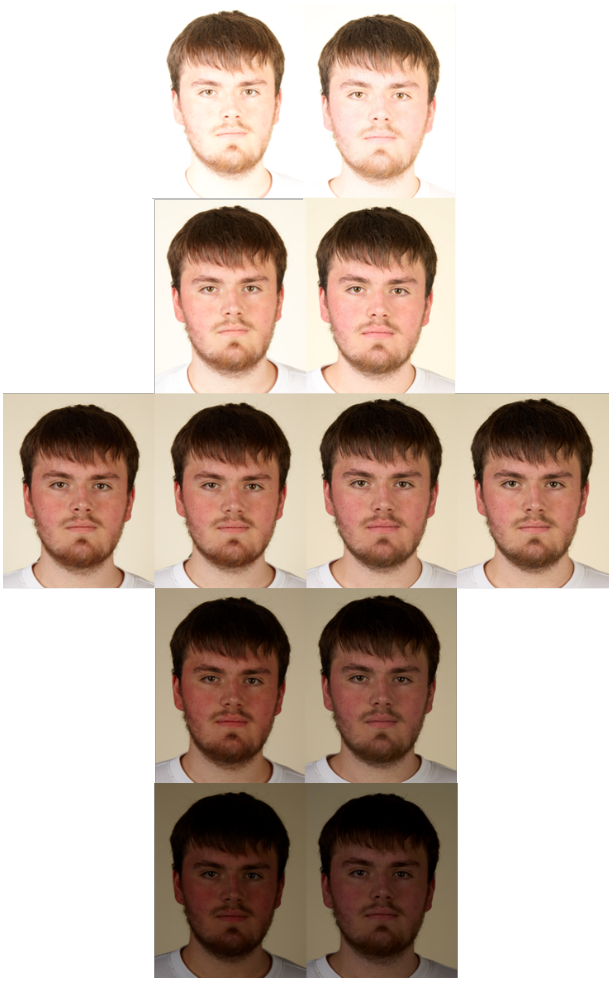

Fig. 3: Example of 12 face images per one subject in different illumination conditions.

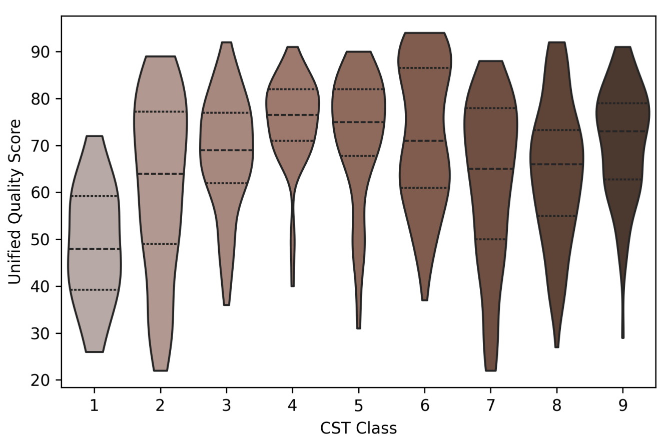

Fig. 4: Distribution of unified quality scores across CST classes.

with respect to skin tone. While mid-range classes achieve the highest median scores, darker skin tone classes exhibit lower medians and increased dispersion. As mentioned earlier, Class 1 contains only two subjects and is therefore not considered statistically representative. Slightly higher UQSs were obtained for normally exposed face images, compared to over- and under-exposed face images. These results highlight that UQSs remain sensitive to skin tone and motivate the need for developing fair FIQA algorithms.

C. Correlation with Skin Tone Estimation

In addition to the measure skin tone from the face, we also estimated the skin tone for each face image in each capture condition (EL1 to EL5). Skin tones were estimated using a public implementation[8] of a modern skin tone classification algorithm [25].

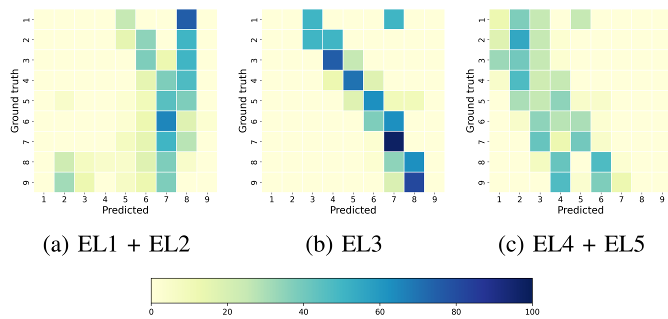

Fig. 5: Correlation of estimated CST classes with measured ones.

Figure 5 shows the correlation between predicted and measured (ground truth) CST classes for nominal as well as under- and over-exposed face images. It can be observed that the CST classes merely estimated from the face images can strongly deviate from the actual measured CST class, in particular for over- and under-exposed capture conditions. This is likely due to the fact that the skin tone estimation algorithm performs a face parsing procedure to extract skin regions, while the overall brightness of the image is not analysed.

V. Limitations

We stress that all biometric samples of the introduced database are captured within a single acquisition session per subject under controlled capture conditions. Consequently, the dataset does not capture temporal variability in appearance, illumination, or skin tone measurements. Short-term physiological and environmental factors (e.g., makeup, skin temperature influenced by season, perspiration) can affect both appearance and skin tone measurements and cannot be compensated by a single capture session. This limits the suitability of the dataset for studying longitudinal effects and for evaluating robustness against session-to-session variation, which is a key factor in biometric performance.

[8]Cheng Ma - Skin Tone Classifier: https://github.com/ChenglongMa/SkinToneClassifier

The controlled capture conditions intend to isolate the impact of skin tone on DV. However, skin tone is known to be entangled to other appearance factors that can cause DV, which limits the interpretability of observed results. Moreover, all images are captured using identical exposure parameters across subjects. While this ensures consistency, it does not account for skin tone–dependent optimal exposure. DV observed across skin tone classes might as well reflect suboptimal capture conditions rather than algorithmic bias.

Skin type is primarily characterised using colorimetric measurements and lightness-based metrics. However, skin characteristics relevant to face image quality and recognition extend beyond skin tone alone and include factors such as texture, reflectance properties, and physiological differences. Note that different demographic groups may share similar lightness values while exhibiting distinct skin properties not captured by the dataset.

Due to the above mentioned limitations, care must be taken to avoid interpreting unequal performance as evidence of unfairness without sufficient causal analysis.

VI. Conclusion

With this work we release the first version of the DAST dataset containing 80 data subjects each represented with 12 high resolution face images. CST skin tone classes 2 to 9 are covered by 10 subjects per class, resulting in a larger database size compared to those of previous works. In addition, a subset of the images show defects with regards to over-exposure or under-exposure. The database allows for DV analysis of face image quality algorithms and for calibration of algorithms that estimate skin tone or skin color. Once the paper is accepted the database can be obtained for academic and commercial purpose upon request.

Acknowledgment

The author thank the volunteers that contributed to this skin tone database. Further we thank Thomas Wilmer for the data protection review of our data collection and review of the consent form.

References

[1] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 39794-5:2019 Information technology - Extensible biometric data interchange formats - Part 5: Face image data*, International Organization for Standardization, 2019.

[2] J. Merkle, C. Rathgeb, B. Herdeanu, B. Tams, D. Lou, A. Dörsch, M. Schaubert, J. Dehen, L. Chen, X. Yin, D. Huang, A. Stratmann, M. Ginzler, M. Grimmer, and C. Busch, "Open source face image quality (OFIQ) - implementation and evaluation of algorithms," https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/OFIQ/Projektabschlussbericht_OFIQ_1_0.pdf, September 2024, last accessed: 2025-02-22.

[3] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 29794-5:2025 Information Technology - Biometric Sample Quality - Part 5: Face Image Data*, International Organization for Standardization, 2025.

[4] W. Kabbani, K. Raja, R. Raghavendra, and C. Busch, "Demographic variability in face image quality measures," in *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. IEEE, September 2024.

[5] ISO/IEC JTC1 SC37 Biometrics, "ISO/IEC TR 25722, demographic variability of face image quality measures," International Organization for Standardisation, Tech. Rep., 2025.

[6] S. Utcke, S. Kramp, S. Engelhardt, W. Kabbani, C. Rathgeb, T. Schlett, J. Merkle, B. Tams, J. Dehen, and C. Busch, "Demographic variability of face image quality measures in operational data," *IEEE Access*, 2026.

[7] A. Dörsch, C. Rathgeb, M. Grimmer, and C. Busch, "Detection and mitigation of bias in under exposure estimation for face image quality assessment," in *Proc. Intl. Conf. of the Biometrics Special Interest Group (BIOSIG)*. IEEE, September 2024, pp. 1–5.

[8] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 19794-5:2011. Information Technology - Biometric Data Interchange Formats - Part 5: Face Image Data*, International Organization for Standardization, 2011.

[9] W. Funk, M. Arnold, C. Busch, and A. Munde, "Evaluation of image compression algorithms for fingerprint and face recognition," in *Proc. IEEE Information Assurance Workshop*. IEEE Press, June 2005.

[10] European Council, "Regulation 2017/2226 of the european parliament and of the council of 30 november 2017 on establishing an entry/exit system (EES) to register entry and exit data and refusal of entry data of third-country nationals," November 2017.

[11] ——, "Commission implementing decision 2019/329 of 25 february 2019 laying down the specifications for the quality, resolution and use of fingerprints and facial image for biometric verification and identification in the entry/exit system (EES)," February 2019.

[12] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch, "Face image quality assessment: A literature survey," *ACM Computing Surveys (CSUR)*, December 2021.

[13] International Civil Aviation Organization NTWG, "Machine readable travel documents – part 3 – specifications for electroncally enabled MRtds with biometric identification capability," https://www.icao.int/sites/default/files/publications/DocSeries/9303_p3_cons_en.pdf, International Civil Aviation Organization (ICAO), 2021, last accessed: 2026-02-05.

[14] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, March 2021.

[15] J. Yang, P. Grother, M. Ngan, K. Hanaoka, and A. Hom, "Face analysis technology evaluation (FATE) part 11: Face image quality vector assessment - specific image defect detection," National Institute of Standards and Technology, NIST Interagency Report 8485, 2024.

[16] P. Grother, M. Ngan, and K. Hanaoka, "Face recognition vendor test (FRVT) part 3: Demographic effect," National Institute of Standards and Technology, NIST Interagency Report 8280, December 2019.

[17] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch, "Demographic bias in biometrics: A survey on an emerging challenge," *Trans. on Technology and Society (TTS)*, vol. 1, no. 2, pp. 89–103, June 2020.

[18] T. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Archives of Dermatology*, vol. 124, no. 6, pp. 869–871, June 1988.

[19] E. Monk, "The monk skin tone scale (MST)," https://osf.io/preprints/socarxiv/pdf4c.pdf, Harward University, 2023, last accessed: 2024-10-28.

[20] C. Cook, J. Howard, L. Rabbitt, I. Shuggi, Y. Siroting, J. Tipton, and A. Vemury, "Colorimetric skin tone scale for improved accuracy of human skin tone annotations," *ACM Journal on Responsible Computing*, vol. 2, pp. 1–6, 2025.

[21] A. Dörsch, T. Schlett, P. Munch, C. Rathgeb, and C. Busch, "Fairness measures for biometric quality assessment," in *Pattern Recognition. ICPR 2024 Intl. Workshops and Challenges*. Springer Nature Switzerland, 2025, pp. 287–298.

[22] A. Dörsch, J. Merkle, B. Tams, G. G. Alvarez, P. Munch, C. Busch, and C. Rathgeb, "Demographic differentials in face image quality: Evaluation and comparison on real and synthetic data," in *Proc. Intl. Conf. of the IEEE/CVF on Computer Vision (ICCV) Workshops*, October 2025, pp. 5874–5882.

[23] G. Pangelinan, X. Merino, S. Langborgh, K. Vangara, J. Annan, A. Beaubrun, T. Weekes, and M. King, "The chroma-fit dataset: Characterizing human ranges of melanin for increased tone-awareness," in *IEEE Intl. Conf. on on Applications of Computer Vision Workshops (WACVW)*. IEEE, 2024, pp. 1170–1178.

[24] I. Joshi, M. Grimmer, C. Rathgeb, C. Busch, F. Bremond, and A. Dantcheva, "Synthetic data in human analysis: A survey," *IEEE in Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 46, pp. 4957–4976, 2024.

[25] R. A. Rejón Pia and C. Ma, "Classification Algorithm for Skin Color (CASCo): A new tool to measure skin color in social science research," *Social Science Quarterly*.