ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/335255036

Subjective Versus Objective Face Image Quality Evaluation For Face Recognition

Conference Paper · May 2019

DOI: 10.1145/3345336.3345338

CITATIONS
4

READS
134

3 authors, including:

Ali Khodabakhsh
Norwegian University of Science and Technology

27 PUBLICATIONS **338** CITATIONS

SEE PROFILE

Christoph Busch
Darmstadt University of Applied Sciences

693 PUBLICATIONS **9,417** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Early detection of Alzheimer's disease from speech signal View project

Project BioMobile II View project

All content following this page was uploaded by Ali Khodabakhsh on 27 October 2020.

The user has requested enhancement of the downloaded file.

Subjective Versus Objective Face Image Quality Evaluation For Face Recognition

Ali Khodabakhsh
Norwegian University of Science and Technology
Gjøvik, Norway
ali.khodabakhsh@ntnu.no

Marius Pedersen
Norwegian University of Science and Technology
Gjøvik, Norway
marius.pedersen@ntnu.no

Christoph Busch
Norwegian University of Science and Technology
Gjøvik, Norway
christoph.busch@ntnu.no

ABSTRACT

The performance of any face recognition system gets affected by the quality of the probe and the reference images. Rejecting or recapturing images with low-quality can improve the overall performance of the biometric system. There are many statistical as well as learning-based methods that provide quality scores given an image for the task of face recognition.

In this study, we take a different approach by asking 26 participants to provide subjective quality scores that represent the ease of recognizing the face on the images from a smartphone based face image dataset. These scores are then compared to measures implemented from ISO/IEC TR 29794-5. We observe that the subjective scores outperform the implemented objective scores while having a low correlation with them. Furthermore, we analyze the effect of pose, illumination, and distance on face recognition similarity scores as well as the generated mean opinion scores.

CCS CONCEPTS

• **Computing methodologies → Biometrics**.

KEYWORDS

face image quality, face recognition, subjective image quality, objective image quality

ACM Reference Format:
Ali Khodabakhsh, Marius Pedersen, and Christoph Busch. 2019. Subjective Versus Objective Face Image Quality Evaluation For Face Recognition. In *Proceedings of* . ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/12345.67890

1 INTRODUCTION

Face recognition systems are becoming ubiquitous and are being utilized in many applications including but not limited to surveillance, forensics, and authentication. These systems rely on a pre-captured image (called enrolment image) as a reference for comparison to an image at hand (called probe image) for identity recognition. In real-world scenarios, the captured face images for probe and reference would be affected by many factors such as lighting, pose, distance from the camera, the camera sensor, etc [5, 8]. These conditions are known to degrade the performance of a face recognition system [17]. One way of tackling this issue is the use of face image quality metrics to filter images with low quality for enhancement [11], rejection, or recapture, and by doing so, improve the overall performance of the face recognition system. There are numerous statistical and learning-based methods that provide quality scores given an image for the task of face recognition [3, 20]. In contrast, there are only a few studies on the human perception of face image quality as a metric for face recognition and its relations to the objective measures.

Adler and Dembinsky [1] compare the subjective face image quality by eight humans to 6 face recognition algorithms. Their results indicate a correlation within two groups, but a low correlation between the groups. However, Hsu et al. [7] show consistency between quality scores from two humans to their proposed quality metrics. In a more recent work, Best-Rowden and Jain [2] design a predictor based on the crowdsourced human assessments of face image quality. Their results show a high correlation with the recognition performance of the face recognition systems outperforming baseline objective measures.

In this study, we focus on the specific scenario of smartphone-based face recognition, with applications such as access control and transaction authentication [14] as a continuation of Wasnik et al. [19]. We follow ISO/IEC TR 29794-5 [10] guidelines for measuring face image quality. The study of the subjective perception of face image quality would give us insights on the similarities and differences of perceptive face image quality and objective face image quality metrics in relation to face recognition system performance. In addition, we analyze the effect of pose, lighting, and distance on the performance of face recognition, subjective face image quality, and objective face image quality in comparison. The rest of the article is organized as follows: Section 2 describes the metrics and data collection, Section 3 discusses the findings, and Section 4 concludes the article.

2 DATA AND METHODOLOGY

The following describes the selected dataset for this study, the objective quality metrics adapted from ISO/IEC 29794-5, and the subjective test protocol and user-interface. This is followed by the description of the performance metrics used in this study.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 1-58113-000-0/00/0010…$15.00
https://doi.org/10.1145/12345.67890

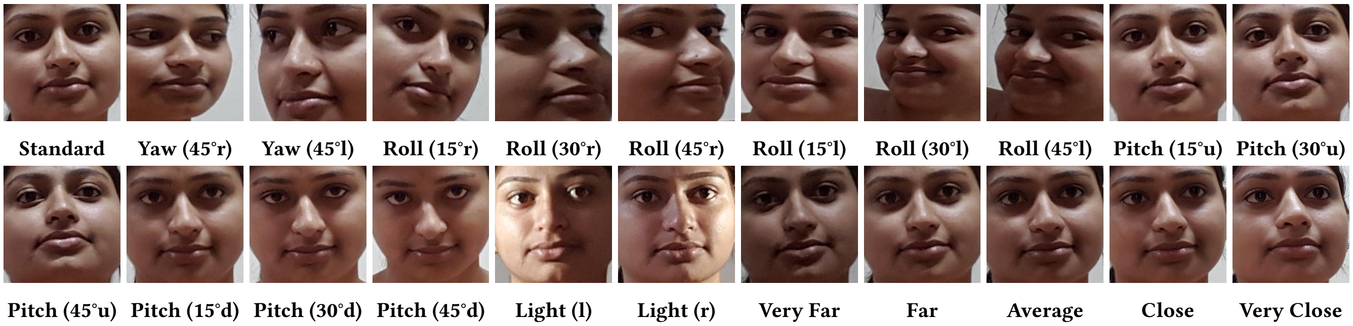


Figure 1. An example set of images from Samsung dataset for one of the subjects after face registering and cropping. The characters r, l, u, and d in captions indicate right, left, up, and down directions.

2.1 Dataset

The dataset used in this study is adapted from [19]. This dataset was collected for the task of face image quality assessment for face recognition in mobile applications. The dataset consists of images collected from 101 subjects in two sessions. The sensor used was the frontal cameras of the iPhone 6 Plus and Samsung Galaxy S7 smartphones.

For each subject, 22 portrait images were captured in the following conditions: one in a standard manner for enrolment, two images with yaw angle of 45° in left and right directions, six roll and six pitch images with angles ranging from -45° to +45° in steps of 15°, two images containing different light source position, and five at different distances. Figure 1 provides an example of these images for a subject after registering and cropping by a face detection algorithm. In this study, only the first session data, containing 22 samples from 101 subjects on each phone, is used.

All the images used in this study went through face detection and registering using the NEUROtechnology VeriLook 5.4 Face SDK[18] which is a commercial face recognition system. The images were cropped and resized to 120 x 120 pixel dimensions.

2.2 Quality Metrics

Similar to [19], the objective quality measures were adapted from the ISO/IEC TR 29794-5. Eight of these metrics that were used in this study are described briefly as follows:

2.2.1 Blur. The blur metric uses edge width as described in [12]. The authors hypothesize that blurry images would have wider edges, and thus the average width of edges in an image can be used for calculating the blur. Higher values indicate more blurry images, thus is expected to have a lower quality.

2.2.2 Sharpness. Given the gradient over an image, the sharpness can be calculated as the sum of gradients over the whole image. The higher the sharpness value, the sharper the image is hypothesized to be, and thus is expected to have a higher quality.

2.2.3 Exposure. Absolute Central Moment (ACM) [16] can be used as a basis for calculation of exposure. Higher exposure is considered to be correlated with higher quality.

2.2.4 Brightness. The brightness of an image can simply be calculated as the average intensity over the image. Higher brightness is considered to correlate with higher quality.

2.2.5 Contrast. The image contrast can be calculated as the root mean square error (RMSE) of the pixel intensities in comparison to the brightness value. An image with a higher contrast is expected to have a higher quality.

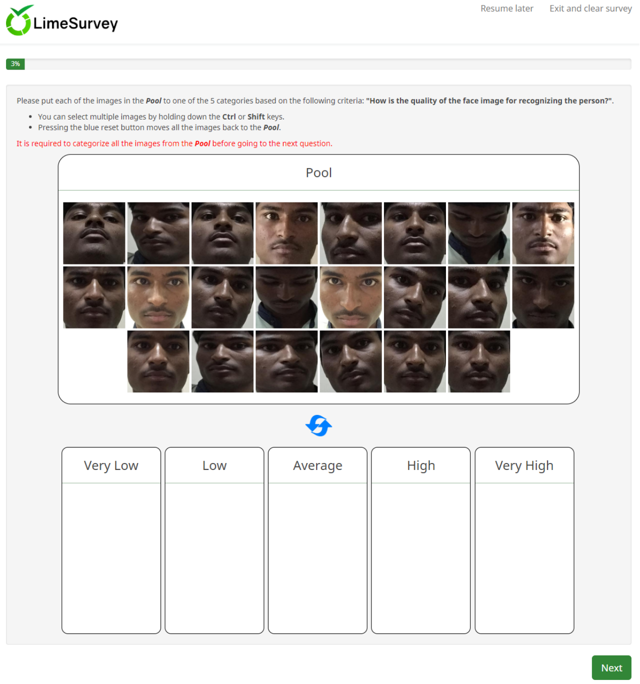
2.2.6 Global Contrast Factor. Following [13] global contrast factor (GCF) is calculated over the image by measuring local contrast at different resolutions. Similar to contrast, higher GCF values are expected to point at higher quality images.

2.2.7 Pose Asymmetry. Following [10], the pose asymmetry can be calculated by splitting the normalized face image in half and comparing the left side and mirrored-right side together. To avoid the effect of lighting, the half face images were filtered using local binary pattern (LBP) filters. The difference is then calculated between the two parts and is averaged in a suitable manner. Higher pose asymmetry would indicate a difference in the pose in each half, and thus lower expected quality.

2.2.8 Light Asymmetry. This metric is similar to pose asymmetry, except that the LBP filtering is not done to take into account the lighting asymmetry in the difference. Higher light asymmetry results are hypothesized to affect face recognition performance negatively.

2.3 Protocol and User Interface

To conduct the subjective test, a categorical image quality classification experiment is designed. The test implemented a simple drag-and-drop interface for classifying each face image into one of the corresponding bins. The criteria was defined as *How is the quality of the face image for recognizing the person?*. The bins were ranging from Very Low (1) to Very High (5). Each set contained only images from one subject, and the same task is then repeated for every other 100 subjects. The order of the face images, as well as the subjects in the images, were randomized to remove any effect of ordering on subjective opinions. The open source LimeSurvey[15] survey web application was used for the implementation of the user interface. A sample page of the test is shown in Figure 2.


Please put each of the images in the ***Pool*** to one of the 5 categories based on the following criteria: **"How is the quality of the face image for recognizing the person?"**.

- You can select multiple images by holding down the **Ctrl** or **Shift** keys.
- Pressing the blue reset button moves all the images back to the ***Pool***.

It is required to categorize all the images from the ***Pool*** before going to the next question.

Figure 2. The test interface. The participants were asked to put the face images in the five quality bins based on the following criteria: *How is the quality of the face images for recognizing the person?*.

The test was done in a controlled setup, where the display was calibrated using *i1 match* photo-spectrometer. The monitor had a brightness level of 120 Cd/m^2 at a color temperature of 6500° Kelvin. The room light was dimmed to approximately 20 Lux. The participants were asked to sit in an arms-length distance from the monitor. The participants were asked to take a 10-minute break at the middle of the test, which was expected to take around one hour in total, to avoid fatigue.

2.4 Performance Evaluation

A face quality metric is expected to provide lower quality scores to samples that are expected not to result in a match for a mated comparison trial by the biometric system. As a result, one can expect that the removal of the lowest quality samples would improve the performance of the biometric system. To follow the terminology described in ISO/IEC 19795-1 [9], the false non-match rate (FNMR) is measured.

To calculate the biometric scores, the ArcFace[4] deep learning method is used to extract embeddings representing each face image. The image taken in a *standard* condition for each subject is used as the enrolment sample and the similarity between the two images was calculated by the cosine distance between their embeddings.

FNMR is defined as the rate of the completed biometric mated comparison trials that result in a false non-match at the decided operating point. Based on the above description, the following two methods can be used for performance evaluation of a quality metric:

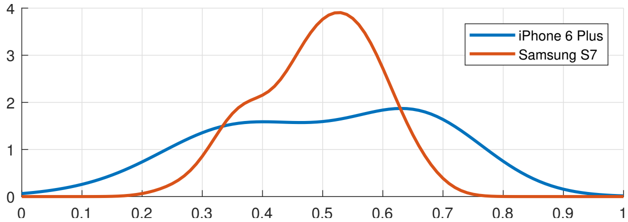


Figure 3. Kernel density on the pairwise correlation between subjective opinions on each dataset. More consistency is visible in the Samsung dataset.

2.4.1 Error Rejection Curve. Error rejection curve (ERC)[6] measures the FNMR in contrast to the number of low-quality samples rejected based on the quality metric. An example can be seen in Figure 5. In an ERC curve, a fixed FNMR rate is used at the start of the process (for example 10%) to tune the operating point of the biometric system. Ideally, the low-quality samples should have the most impact on the FNMR, and thus removing them should reduce the FNMR rapidly and towards zero. On the other hand, if a quality metric does not correlate with the quality of the face image for the face recognition system, the removing of the lowest quality samples would not affect the FNMR value in a consistent manner. As more low-quality samples get rejected base on their quality, the fewer samples remain for calculation of FNMR, resulting in a noisy estimate of it.

2.4.2 Area Under Curve. To have a single value describing the performance of a quality metric, the area under the ERC curve can be used. The ERC area under the curve (AUC) can be calculated by integration on the whole ERC curve or a specific range of it (i.e. the Partial Area Under Curve). This can be done by focusing on the initial part of the curve and thus avoiding the noisy end part of the ERC curve. Given a specific FNMR fixed starting value x, the theoretical minimum AUC for the full ERC curve would be x^2, and a non-performing quality metric will have a value near x.

3 RESULTS AND DISCUSSION

Subjective scores from 13 participants per phone types were collected. The participants were colleagues with different nationalities from the Information Security and Communication Technology and Computer Science departments at NTNU. The participants each spent on average two 45 minutes sessions with a 10-minute break in between. Responding to a single set of face images took on average 54 seconds (std = 30). The highest education level completed by participants were Bachelor (1), Master (18), and Ph.D. (7). Out of 26 participants, 20 were male and 6 were female. The average age was 30 (std = 3). Out of 26 participants, 17 claimed familiarity with image quality assessment.

3.1 Subjective Opinion Scores

The opinion scores were closely correlated between the participants on each dataset as shown in Figure 3. Based on the collected categorical opinion score per image, the mean opinion score (MOS) was generated as the quality score to be compared with the objective measures. Along with MOS, the standard deviation of the opinion

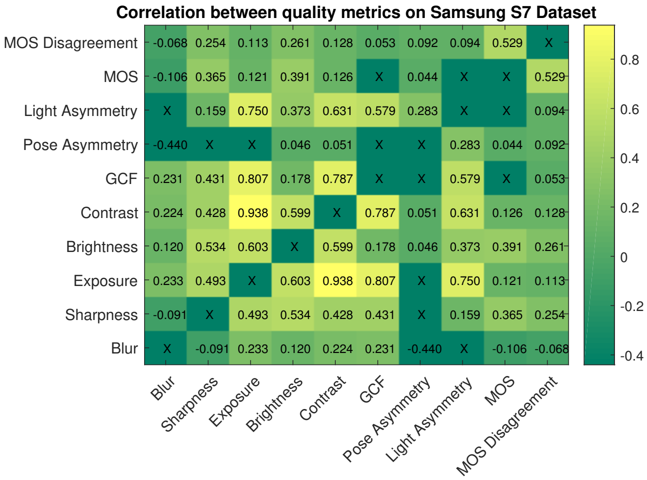


Figure 4. Pearson correlation coefficient between pairs of quality metrics on Samsung dataset. The 'X' symbol indicated the correlation coefficients which had a higher than 0.05 p-value.

scores is also calculated as a measure of disagreement between the individuals. This feature is called MOS disagreement in the rest of the article.

The resolution of the front-facing camera of the two phones are 5 MP and 1.2 MP for the Samsung Galaxy S7 and the iPhone 6 Plus respectively. Despite the big difference in the resolutions, similar results were obtained as shown later in the next section, and thus only the results from the phone with the higher resolution camera (Samsung Galaxy S7) is used in explanations of the performances of the systems. The same figures for iPhone dataset are added to the end of this article for reference (Figure 7).

The correlation between all the eight ISO metrics along with the two MOS metrics is shown in Figure 4. Based on the correlation values, the metrics can be categorized to four groups with high intra-group and low inter-group correlation: (1) MOS and MOS disagreement, (2) Sharpness, Exposure, Brightness, Contrast, GCF, and Light Asymmetry, (3) Blur, and (4) Pose Asymmetry. Contrast, GCF, Light Asymmetry, and Exposure measures are highly correlated. These results also indicate that the MOS features are not correlated to ISO quality metrics, except for a very low correlation to the brightness and sharpness metrics.

3.2 Quality Metrics vs Face Recognition Performance.

Table 1 shows AUC values for ERC curves for all the evaluated metrics. Values around 0.1 for AUC and 0.02 AUC20 (Partial AUC over the first 20%) would signify an ineffective image quality metric. Based on this table it can be concluded that MOS has the highest effectiveness as a quality metric, followed by MOS disagreement. Out of objective metrics, brightness and contrast show a consistent better-than-chance performance in their AUC values. Exposure, GCF, light asymmetry show similar behavior, but only on one of the datasets.

Table 1. Area under curve (AUC) and partial AUC over the first 20% (AUC20) for ERC plots on both datasets for different quality methods. FNMR is set to 10%. ISO/IEC combined indicate mean of scores from the eight ISO quality metrics in the best performing mode and MOS combined is the combination of MOS and MOS disagreement. The (+) sign signifies that higher values correspond to better quality, while (-) signifies the contrary. The (d) sign signifies higher likelihood to average distribution as a measure of quality. The theoretical best value for both AUC and AUC20 is 0.01, while a random quality metric would have an AUC and AUC20 of around 0.1 and 0.02 respectively. Results from commercial of the shelf (COTS) system is added from Wasnik et al.[19] for reference.

Quality Measures	iPhone 6 Plus Dataset (ERC)		Samsung S7 Dataset (ERC)	
	AUC	AUC20	AUC	AUC20
Direct				
Blur (-)	0.1119	0.0190	0.1195	0.0202
Sharpness (+)	0.1208	0.0191	0.1386	0.0193
Exposure (+)	0.0817	0.0197	0.1013	0.0199
Brightness (+)	0.0780	0.0193	0.0781	0.0175
Contrast (+)	0.0714	0.0188	0.0903	0.0194
GCF (+)	0.0881	0.0195	0.1333	0.0205
Pose Asymmetry (-)	0.1201	0.0208	0.1086	0.0207
Light Asymmetry (-)	0.1040	0.0216	0.0858	0.0202
MOS (+)	**0.0332**	**0.0151**	**0.0354**	**0.0161**
MOS Disagreement (+)	0.0719	0.0174	0.0724	0.0168
Reversed				
Blur (+)	0.1118	0.0194	0.0891	0.0191
Sharpness (-)	0.1064	0.0186	0.1032	0.0175
Exposure (-)	0.1132	0.0212	0.1008	0.0204
Brightness (-)	0.1173	0.0209	0.1467	0.0209
Contrast (-)	0.1288	0.0215	0.1155	0.0207
GCF (-)	0.1131	0.0207	0.0853	0.0185
Pose Asymmetry (+)	0.0908	0.0189	0.0818	0.0197
Light Asymmetry (+)	0.0788	0.0202	0.1066	0.0208
MOS (-)	0.2006	0.0223	0.1896	0.0222
MOS Disagreement (-)	0.1446	0.0215	0.1630	0.0205
Likelihood				
Blur (d)	0.0899	0.0186	0.0973	0.0197
Sharpness (d)	0.0818	0.0183	0.0679	0.0176
Exposure (d)	0.1036	0.0206	0.0947	0.0203
Brightness (d)	0.1157	0.0205	0.0944	0.0186
Contrast (d)	0.1026	0.0202	0.0987	0.0199
GCF (d)	0.1034	0.0198	0.0845	0.0191
Pose Asymmetry (d)	0.0969	0.0192	0.1101	0.0207
Light Asymmetry (d)	0.1216	0.0213	0.1133	0.0204
MOS (d)	0.0840	0.0190	0.0906	0.0167
MOS Disagreement (d)	0.0895	0.0189	0.0754	0.0178
Combined				
ISO/IEC combined	0.0639	0.0193	0.0824	0.0197
MOS combined	0.0515	0.0172	0.0477	0.0164
COTS	**0.015**	**0.00**	**0.0285**	**0.0020**

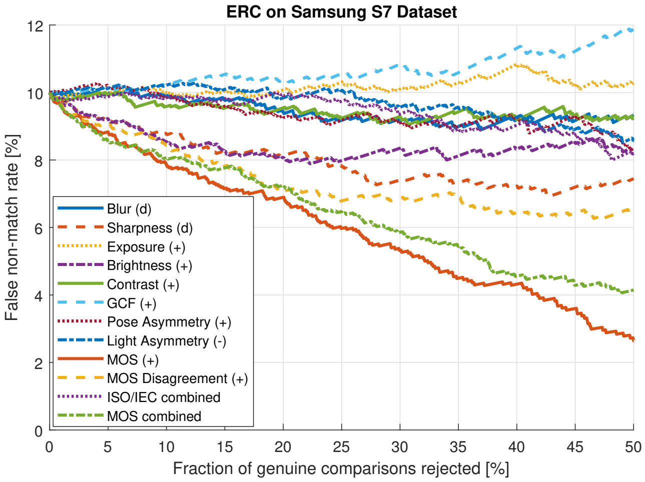


Figure 5. Error rejection curve corresponding to area under curve values at Table 1 for Samsung dataset. The +/- sign at the legend signifies the direction in which the quality value was considered better. The (d) signifies the likelihood to the average distribution.

The second part of the table consists of the same quality measures, yet the direction in which we consider the measure to point at a high quality is reversed. The motivation behind this is that for some metrics, too high of a specific quality measure might point at a faulty image (for example if the brightness of an image is too high). Among the reversed metrics, sharpness and pose asymmetry consistently outperform their direct counterparts. Furthermore, blur, GCF, and light asymmetry show a similar behavior but only on one of the datasets. This counter-intuitive result suggests that some metrics have a non-monotonic relation to image quality, thus it is not a specific direction that signifies lower quality, but the extremeness. It is also worth mentioning that MOS disagreement performed much better than MOS agreement (aka reversed MOS disagreement). In other words, a higher MOS disagreement results in higher quality. This can be interpreted to that participants agree on which images are low-quality while disagreeing on the quality of higher quality images.

The extremeness of a value can be measured by the negative likelihood to the overall distribution, and as such, the likelihood value can be used as a pointer to higher quality. The third part of the table shows the AUC values using likelihood as a measure of quality. As expected, for specific measures (blur and sharpness), likelihood outperforms both direct and reverse metrics. From these three parts, it can be concluded that for blur and sharpness the likelihood metric, and for pose asymmetry the reversed metric work better. For exposure, brightness, contrast, MOS, and MOS disagreement, the direct measure works the best, while for GCF and light asymmetry the results show no specific pattern.

The last table includes the combination of the best ISO and MOS metrics along with the performance of commercial of the shelf (COTS) system from Pankaj et al.[19]. The combinations were done by averaging the quality metrics after zero normalization. MOS combined is the combination of MOS and MOS disagreement metrics, and the ISO/IEC combined is the combination of reversed pose asymmetry along with likelihood blur and sharpness with the other five ISO/IEC metrics. Combination methods do not outperform their best-performing metrics, suggesting the naive combination method to be not suitable.

The ERC curve for the best performing ISO/IEC and MOS metrics along with their combined methods (ISO/ICE combined and MOS combined) are visualized at Figure 5. The quality metrics can be roughly categorized to two categories, the ones that follow the curve of an ineffective quality measure (constant and varying near the initial FNMR value), and the ones that gradually decrease towards zero. The metrics in the second category are brightness (+), sharpness (d), MOS, MOS disagreement, and MOS combined.

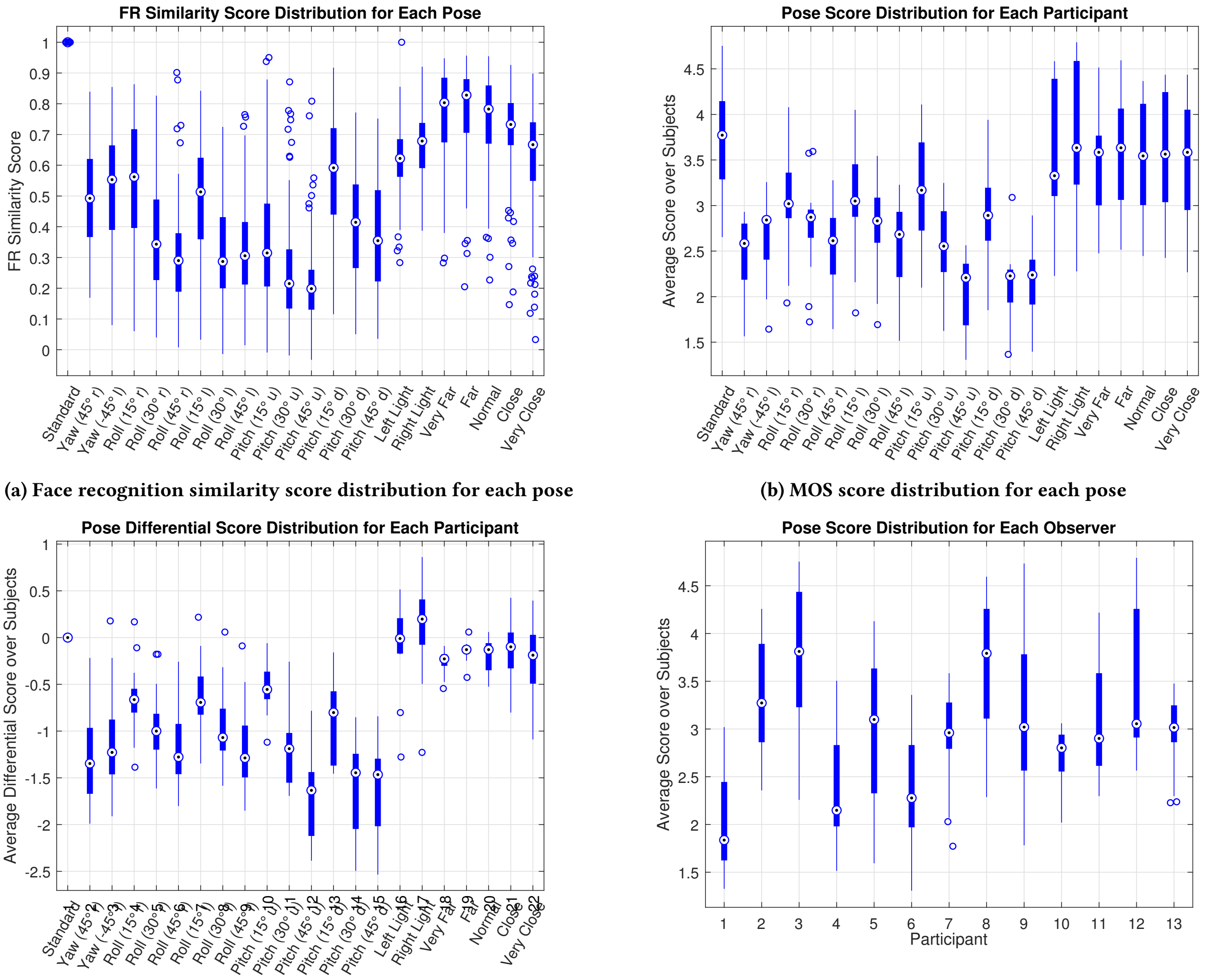
Out of the five well-performing quality metrics, the best performing three are the MOS-based scores. This signifies the correlation between the subjective opinion about the quality and the performance of the face recognition system.

3.3 Capturing Condition

In this study, the factors that contribute to the variation in the face recognition and participant scores can be limited to the following three: The subject in the image, the participant, and the capturing condition. To observe the variation that each of these factors contributes, the standard deviation is calculated after averaging the effect of the other two factors. For Samsung dataset, the standard deviation is 0.20, 0.59, and 0.62 for the subject, participant, and pose respectively. For iPhone dataset, these values are 0.16, 0.80, and 0.74 in order. It can be concluded that the subject in the image has the least effect on the subjective score variation. To further study the variation caused by the other two factors, the score variations for face recognition system and MOS are illustrated in Figure 6.

In Figure 6a the score distribution for the face recognition system is illustrated for each pose. It can be seen that after the standard image (which was used as the template and, as a result, had the perfect similarity score of 1) the highest similarity scores were achieved over frontal face images with light or distance variations. The yaw images at 45° and roll or pitch of 15° did reduce the similarity scores compared to the frontal images, yet the confidence intervals overlap partially with frontal images. A significantly lower face similarity score resulted with roll and pitch of above 30°. This shows the robustness of the face recognition system to lighting conditions, and distance, while sensitivity over roll and pitch angles. Furthermore, the biometric system is more sensitive to upward pitch angles compared to the downward angles.

In Figure 6b a similar pattern emerged, yet the confidence intervals are wide and it can only be concluded that the frontal poses (standard, light variation, and distance variation) have MOS that are significantly higher than pose variations of above 30°. To reduce the variation and thus the confidence interval, the effect of inter-participant opinion differences can be reduced by using Differential MOS (DMOS) in reference to the standard image per subject. Figure 6c illustrates DMOS over poses. It can be observed that lower angles from frontal correspond to a higher DMOS and the pitch variations cause the highest drop in DMOS scores. Yaw variations also contribute to a higher drop in DMOS compared to that of face



(a) Face recognition similarity score distribution for each pose

(b) MOS score distribution for each pose

(c) Differential MOS score distribution for each pose

(d) Score distribution of each participant

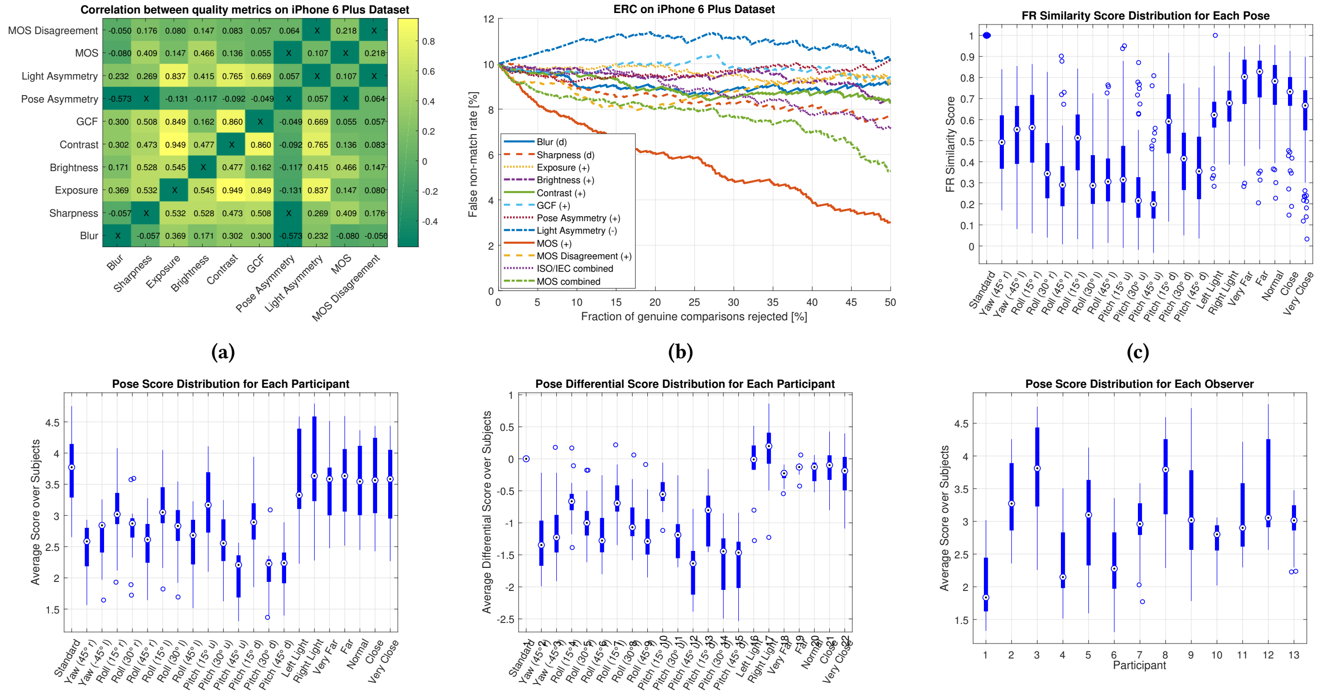
Figure 6. All figures incorporate Samsung S7 dataset and MOS scores are averaged on subjects. For face recognition and differential MOS, the standard image is used as reference.

recognition similarity score. The difference between the effect of upward and downward pitch angle is not present in DMOS scores.

Figure 6d shows the score distribution provided by each participant on each capturing condition. It can be seen that participants have a wide opinion difference on average quality, yet the variation of scores is close to 1 point and is similar between participants.

4 CONCLUSION

In this study, we compared the subjective face image quality to objective metrics in the context of face recognition systems. Subjective opinion scores were collected from 26 participants in a controlled experimental setup. The results show that subjective face image quality is well aligned with the face recognition system performance while showing low correlation to the objective metrics. The results also show that not all quality metrics have a monotonic relation to image quality for face recognition. In addition, participants showed agreement on the quality of the low-quality samples, while they disagreed on the high-quality samples, and the disagreement by itself acted as a good predictor of face image quality. Even though evaluating disagreement is not feasible in an operational scenario, the relation can be relevant in posteriori analyses. Furthermore, we studied the effect of the various pose, lighting, and distance conditions on the face recognition system and the subjective opinion scores, concluding that pitch has the highest impact while the distance from camera and lighting had little impact on the performance.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 7. Figures in order correspond to Figure 4, 5, 6a, 6b, 6c, and 6d, generated on the iPhone Dataset.

5 ACKNOWLEDGMENTS

We thank Dr. Pankaj Wasnik and Prof. Raghavendra Ramachandra for providing the dataset for the experiments, and for their invaluable comments and guidance.

6 REFERENCES

[1] A. Adler and T. Dembinsky. 2006. Human Vs. Automatic Measurement of Biometric Sample Quality. In *2006 Canadian Conference on Electrical and Computer Engineering*. 2090–2093. https://doi.org/10.1109/CCECE.2006.277715

[2] L. Best-Rowden and A. K. Jain. 2018. Learning Face Image Quality From Human Assessments. *IEEE Transactions on Information Forensics and Security* 13, 12 (Dec 2018), 3064–3077. https://doi.org/10.1109/TIFS.2018.2799585

[3] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. 2014. Biometric quality: a review of fingerprint, iris, and face. *EURASIP Journal on Image and Video Processing* 2014, 1 (02 Jul 2014), 34. https://doi.org/10.1186/1687-5281-2014-34

[4] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou. 2018. Arcface: Additive angular margin loss for deep face recognition. *arXiv preprint arXiv:1801.07698* (2018).

[5] W. Funk, M. Arnold, C. Busch, and A. Munde. 2005. Evaluation of image compression algorithms for fingerprint and face recognition systems. In *Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop*. 72–78. https://doi.org/10.1109/IAW.2005.1495936

[6] P. Grother and E. Tabassi. 2007. Performance of Biometric Quality Measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 4 (April 2007), 531–543. https://doi.org/10.1109/TPAMI.2007.1019

[7] R. V. Hsu, J. Shah, and B. Martin. 2006. Quality Assessment of Facial Images. In *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*. 1–6. https://doi.org/10.1109/BCC.2006.4341617

[8] ISO/IEC. [n. d.]. *ISO/IEC DIS 39794-5 Information technology – Extensible biometric data interchange formats – Part 5: Face image data*. ISO/IEC.

[9] ISO/IEC. 2006. *ISO/IEC 19795-1 Information technology – Biometric performance testing and reporting – Part 1: Principles and framework*. ISO/IEC.

[10] ISO/IEC. 2010. *ISO/IEC TR 29794-5 Information technology - Biometric sample quality - Part 5: Face image data*. ISO/IEC.

[11] X. Liu, M. Pedersen, C. Charrier, and P. Bours. 2018. Can image quality enhancement methods improve the performance of biometric systems for degraded face images?. In *2018 Colour and Visual Computing Symposium (CVCS)*. 1–5. https://doi.org/10.1109/CVCS.2018.8496511

[12] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. 2002. A no-reference perceptual blur metric. In *Proceedings. International Conference on Image Processing*, Vol. 3. III–III. https://doi.org/10.1109/ICIP.2002.1038902

[13] Kresimir Matkovic, László Neumann, Attila Neumann, Thomas Psik, and Werner Purgathofer. 2005. Global Contrast Factor-a New Approach to Image Contrast. *Computational Aesthetics* 2005 (2005), 159–168.

[14] Ajita Rattani and Reza Derakhshani. 2018. A Survey Of mobile face biometrics. *Computers & Electrical Engineering* 72 (2018), 39 – 52. https://doi.org/10.1016/j.compeleceng.2018.09.005

[15] Carsten Schmitz et al. 2012. LimeSurvey: An open source survey tool. *LimeSurvey Project Hamburg, Germany. URL http://www. limesurvey. org* (2012).

[16] M. V. Shirvaikar. 2004. An optimal measure for camera focus and exposure. In *Thirty-Sixth Southeastern Symposium on System Theory, 2004. Proceedings of the.* 472–475. https://doi.org/10.1109/SSST.2004.1295702

[17] T. Sim, S. Baker, and M. Bsat. 2002. The CMU Pose, Illumination, and Expression (PIE) database. In *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*. 53–58. https://doi.org/10.1109/AFGR.2002.1004130

[18] NEUROtechnology VeriLook. [n. d.]. 5.4,"VeriLook 5.4 Face SDK,".

[19] P. Wasnik, K. B. Raja, R. Ramachandra, and C. Busch. 2017. Assessing face image quality for smartphone based face recognition system. In *2017 5th International Workshop on Biometrics and Forensics (IWBF)*. 1–6. https://doi.org/10.1109/IWBF.2017.7935089

[20] Pankaj Wasnik, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. 2018. An Empirical Evaluation of Deep Architectures on Generalization of Smartphone-based Face Image Quality Assessment. In *IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*.

Authors' background

Your Name	Title	Research Field	Personal Website
Ali Khodabakhsh	Phd candidate	Biometrics	ali.khodabakhsh.org
Marius Pedersen	full professor	Image Quality	www.ansatt.hig.no/mariusp
Christoph Busch	full professor	Biometrics	christoph-busch.de

View publication stats