secunet

Facial Image Quality Assessment – State of the Art

15.11.2022

Johannes Merkle

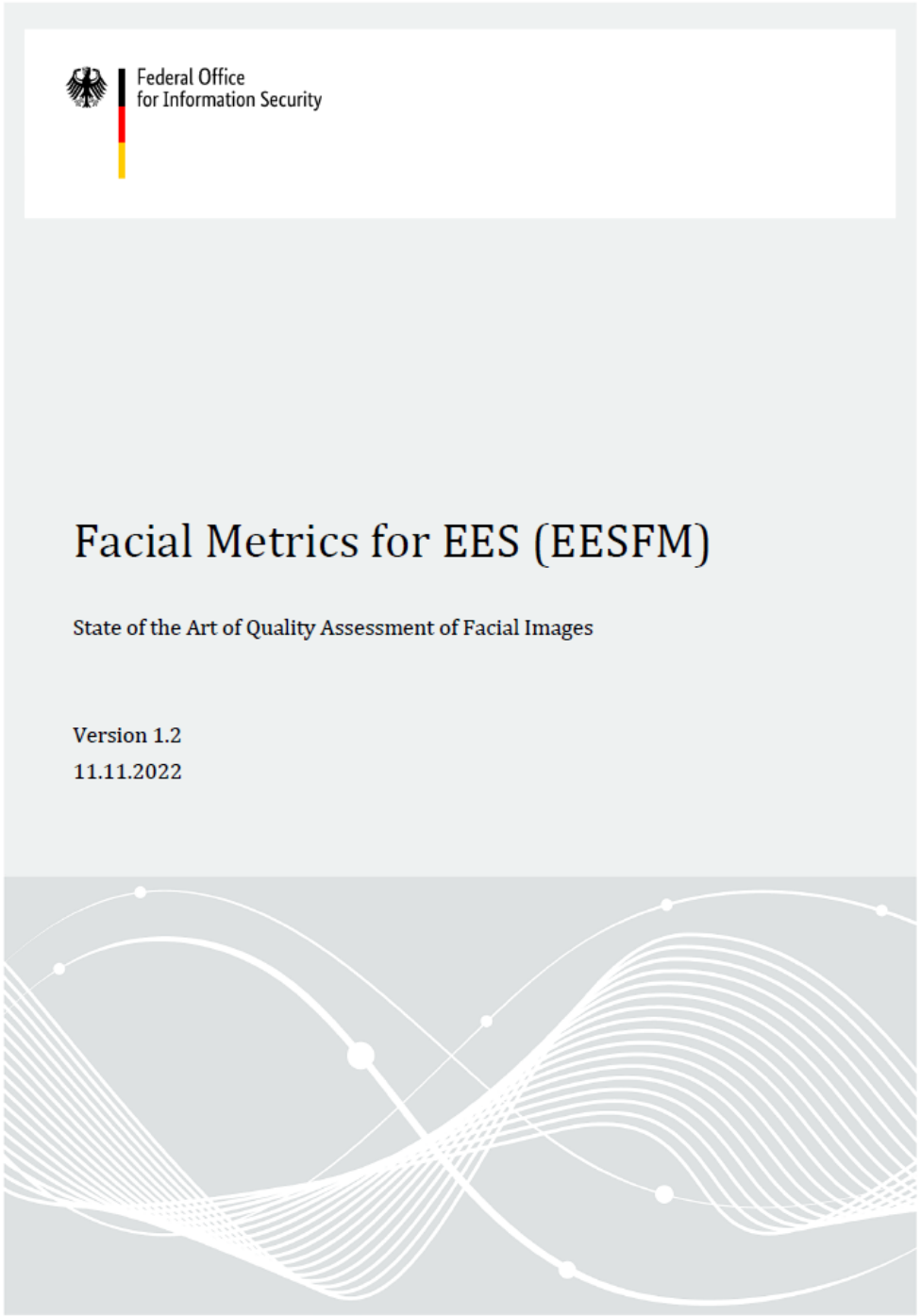
https://de.wikipedia.org/wiki/StyleGAN

Introduction

- Quality is paramount for biometric applications
- For facial images, various aspects need to be checked
- For many aspects and building blocks, algorithms have been developed
- The project OFIQ aims at compiling and developing suitable algorithms to create a comprehensive open source QA software
- First step was to determine the state of the art (SOTA) of QA for facial images

Introduction

- Report on SOTA will be published on arXiv.org
  - Moderator approval pending
  - Link will be announced on Thursday (if possible)



Pre-processing Steps

- Many quality measures require the following pre-processing steps
  - Restrict computations to face region
  - Locate certain parts of face and subject

- Face detection: Find face(s) in image
- Facial landmark computation: Determine exact positions of certain face parts
- Subject segmentation: Determine regions of certain parts of subject in image

Unified Quality Score

- Prediction of the utility of the facial image for face recognition
- Should depend on all factors (and only on those) that can influence recognition performance
- Mapping to a single scalar value
  - Analogous to NFIQ2 score for fingerprints
  - No actionable feedback to photographer or subject

Unified Quality Score

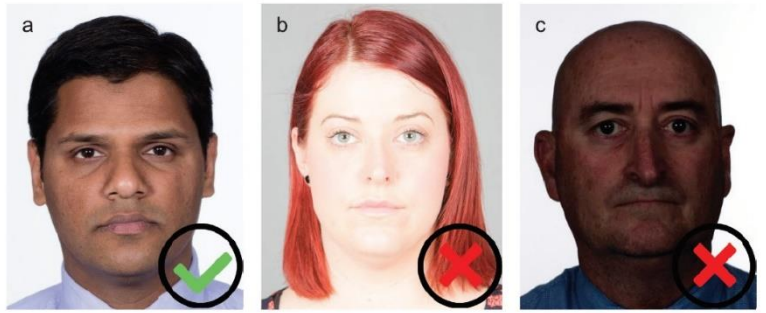
- SOTA Algorithms based on CNNs
  - SSD-FIQA, PFE
  - MagFace (also face recognition)
  - FaceQAN (based on adverserial samples, very slow)

Capture-related Aspects

- Illumination
- Natural colors
- Sharpness
- Compression
- Radial distortions

Illumination

- Illumination issues can comprise:
  - Face is too bright or too dark
  - Narrow frequency distribution of brightness
  - Uneven spatial distribution of brightness (e.g. shadows, hot spots)
- Possible causes:
  - Under exposure / over exposure
  - Unsuitable lighting
  - Post-processing (e.g. image editing)



a) Correct exposure b) Over exposure c) Under exposure

a) Side illumination b) Top illumination c) Bottom illumination

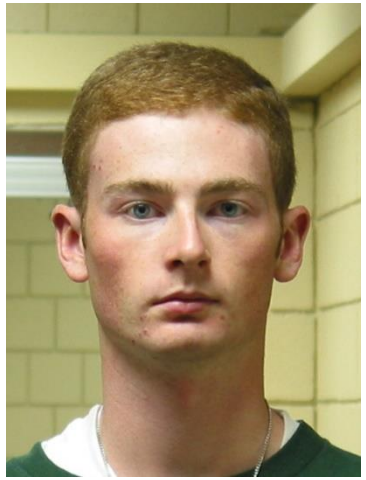
- Illumination problems can reduce visibility of texture details
- However: SOTA face recognition algorithms CNNs are quite robust w.r.t illumination

Algorithms for illumination assessment

- Analysis of luminance histograms (e.g. WD5 of ISO/IEC 29794-5)
  - Statistical moments (mean, variance, skewness, kurtosis)
  - Count pixels with very low / very high value
  - Dynamic Range
    - Entropy of luminance histogram
  - Illumination uniformity
    - Measure intersection of luminance histograms in zones on left and right cheek
- Assessment by CNN: $FIIQA_{DCNN}$

Sharpness / Spatial Resolution

- Fine details / textures of the face should be clearly visible
- Different types of issues and their origins:
  - **De-focus**: Incorrect focus of the capture system.
  - **Motion blur**: Movement of the subject or capture system
  - **Post-processing:** E.g. resizing, printing and scanning



Sharpness / Spatial Resolution

- Potential negative impact on face recognition possible
  - However: Many SOTA face recognition CNNs take inputs of size 112x112
- May be more important for other scenarios, e.g.
  - Manual inspection
  - Manipulation detection (e.g. morphing attacks)

Algorithms for Assessment of Sharpness

- Features for edge or texture detection (e.g. Laplace, Sobel, LBP)
- Statistical approaches (eigenvalues of image's covariance matrix)
- Comparison with blurred version of image
- Distribution of local intensity variation / gradients
- Analysis in frequency domain
- Classification by CNN

Subject-related Aspects

- Only one face present
- Location, size and completeness of face
- Background Uniformity
- Occlusions of Face
- Eyes open and mouth closed
- Head Pose
- Expression Neutrality
- Reflection on Eyeglasses
- Shoulder Pose
- No Head Coverings

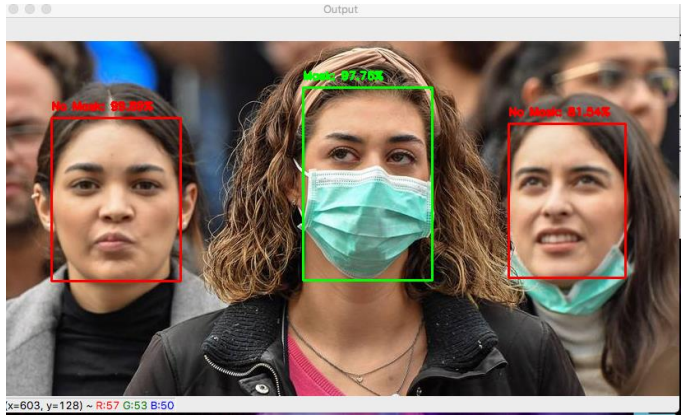
Occlusions of Face

- Face shall not be occluded by objects
  - E.g. face masks, sunglasses, hair, scarfs, veils, hands, mobile phones
- Negative impact on accuracy of biometric algorithms
  - In particular for eyes, mouth and nose

Detection of Occlusions of the Face

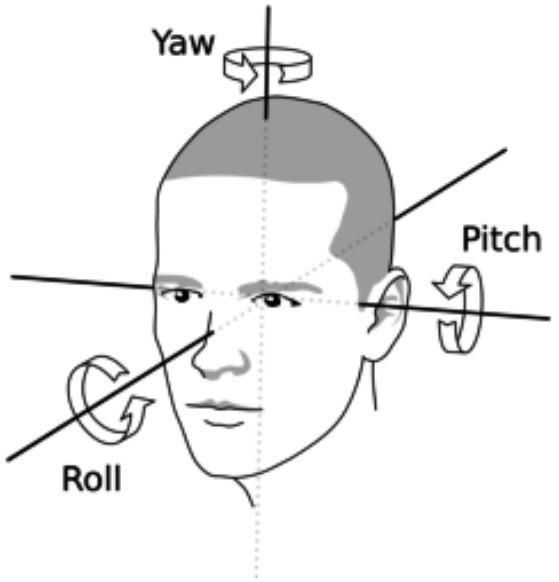
- QA algorithm should indicate, which face parts (eyes, nose, mouth) are not visible
  - Actionable feedback

- Many CNNs can detect certain type of occlusions
  - Face mask detection
  - Detection of sunglasses
- Few algorithms for the detection of arbitrary occlusions
  - No free implementations found
  - De-occlusion CNNs or face segmentation could be used, but are large
  - CNN is currently trained by secunet



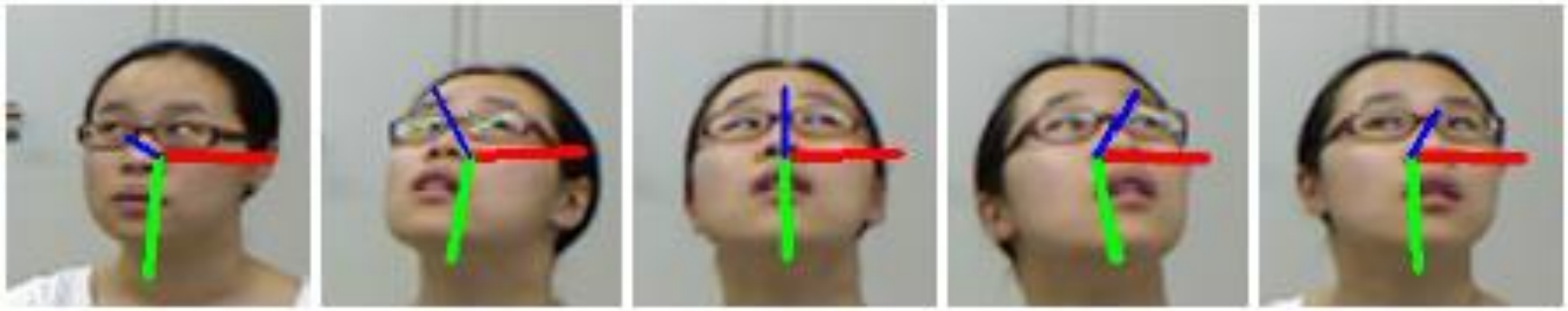
Head Pose

- Frontal pose is required in reference images by ISO/IEC 19794-5:2011 and 39794-5:2019
- SOTA face recognition CNNs are quite robust to limited variations of pitch/yaw (< 20°)
  - Other algorithms and manual inspection may benefit from frontal poses
  - Roll is corrected in pre-processing (alignment)



Head Pose

- SOTA algorithms:
  - Deep learning methods (e.g. 3DDFA-V2, MOS)
    - Often multi-task CNNs: 3D facial landmarks + pose estimation
  - 3D Morphable Model (3DMM) parameters regression



- Relatively small CNNs freely available

Expression Neutrality

- Neutral expression required for reference images by ISO 19794-5 and ISO 39794-5
- SOTA face recognition systems are robust against slight variations in expression
  - However, extreme expressions can lead to false rejection
  - Other algorithms and manual inspection may benefit from neutral expressions

Expression Neutrality

- SOTA algorithms are CNNs for emotion estimation
  - Multi-class classifier (e.g. Joy, Sadness, Anger, Fear, Surprise, Disgust, Neutral)
  - Requires mapping to single value expressing how neutral a facial expression is
    - Training of classifier, e.g. SVM, on CNN's output vector

secunet

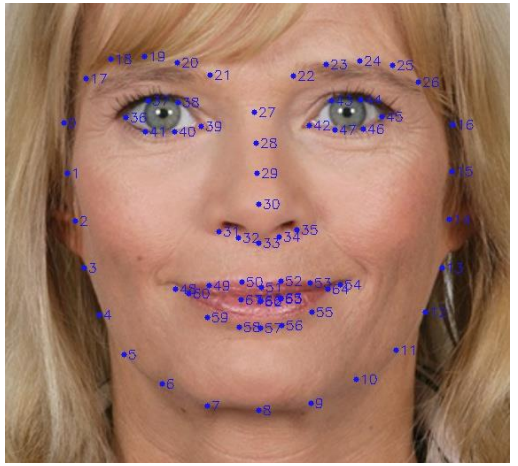
Face Detection

- Output bounding box(es) of face(s)

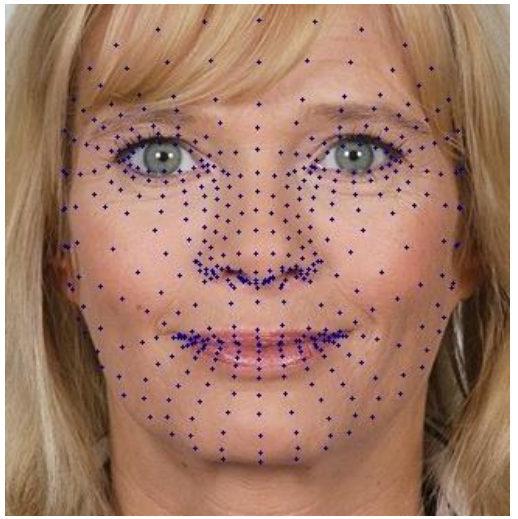
- SOTA algorithms based on CNNs
  - Can detect multiple (even tiny) faces in image
  - Very robust w.r.t. head poses, illumination and occlusion
  - Various CNN architectures, e.g. FPN, SSD, YOLO
    - Wide range of model sizes

Facial Landmark Computation

- Determine coordinates of key points of face parts
  - E.g. eyes, eyebrows, nose, lips, chin
  - Pre-requisite: face detection
  - Often used for alignment of faces for CNNs



- SOTA algorithms
  - Mostly CNNs (e.g. AWing, PFLD), but dlib still quite good
  - Very robust w.r.t. head pose, expression, illumination and occlusion
  - Number of landmarks depend on training set (e.g. 68, 98)
  - Some Algorithms output 3D face mesh (e.g. 3DDFA, mediapipe)
  - Wide range of model sizes

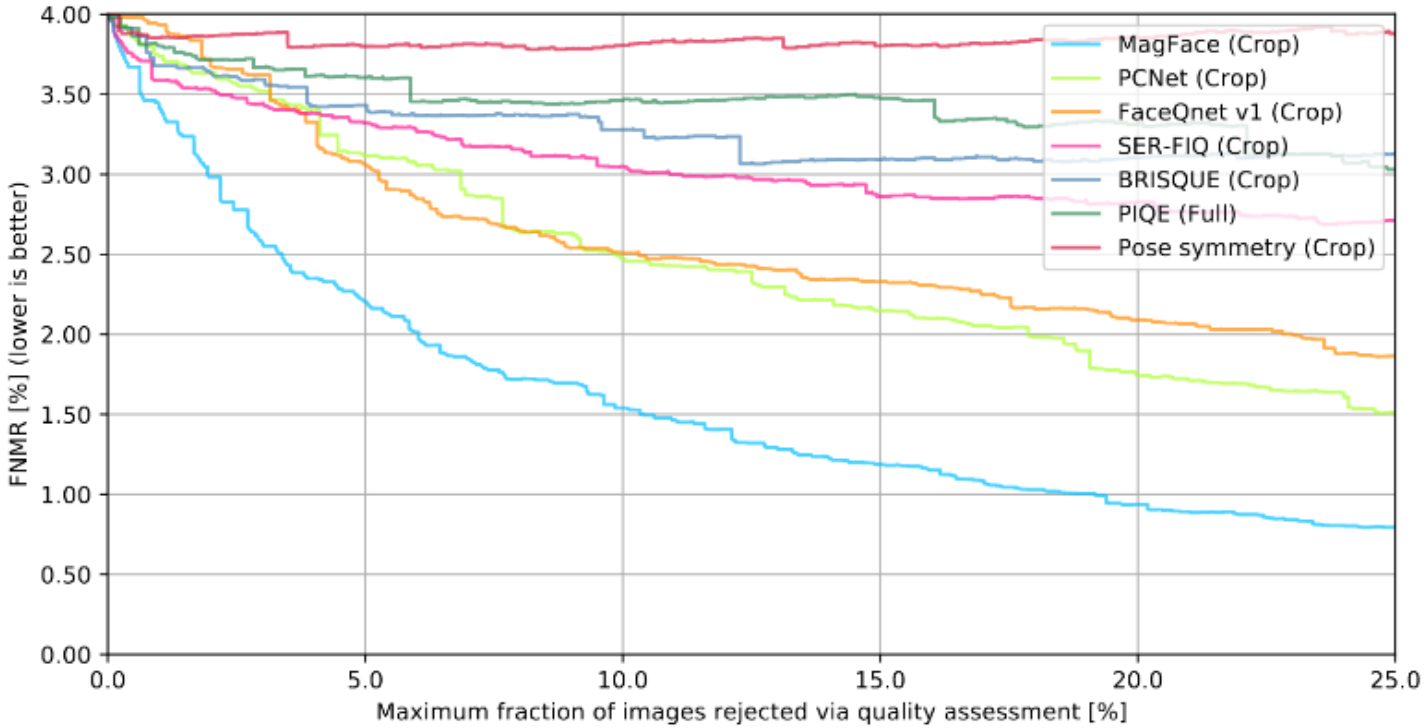


Subject Segmentation

- Determine regions of parts of subject in image
  - E.g. face parts, hair, body, eyeglasses, hats
- SOTA algorithms based on “face parsing” CNNs
  - E.g. FaRL-B, DML-CSR, EHANet
  - Segmented parts depend on training set (and labels) used
  - Wide range of model sizes

Unified Quality Score

- Evaluation by correlation with match scores
  - E.g. Error vs. Discard (EDC), a.k.a Error vs. Reject Curves (ERC)



Illumination

- Illumination problems can reduce visibility of texture details
  - However: SOTA face recognition algorithms CNNs are quite robust w.r.t illumination
- Challenges for QA algorithms:
  - Should work for all skin colors
  - Should not be misguided by hair (e.g. dark beards)

Natural colors

- Colors of face in image may deviate from its natural colors
  - Color tint or too high / low saturation
- Possible causes:
  - Inept illumination (light source, reflections)
  - Malfunctions or misconfiguration of camera or software
  - Image editing
- No published results on impact of on face recognition performance
  - Negligible impact for greyscale conversion

a) Correct colour b) Colour deviations c) Excessive saturation in hue, saturation and luminance (HSL) colour space

Algorithms to detect unnatural color

- Analysis of color channel histograms
  - E.g. in CIELAB or HSV space
- Deep-learning approaches
  - White balancing

- Challenges :
  - Should work for all skin colors
  - Should not be misguided by hair (e.g. beards)

Compression

- Lossy image compression can reduce visibility of texture
  - Negative impact on face recognition only for very strong compression (talk of Torsten Schlett)
  - Other algorithms are more sensitive, e.g. MAD
- Common compression algorithms: JPEG and JPEG2000
  - Result in different type of artifacts
- Detection approaches:
  - Compression ratio based on image resolution and file size
  - Image-based detection of compression artifacts (JPEG only)
    - Hand-crafted analyses of DCT coefficients
    - Deep learning-based methods

Radial Distortions

- Fish eye effect
  - Typically results from too small camera-subject distance, e.g. for selfies
- Results in incorrectly represented geometry of face
  - Assumed to have negative effect on face recognition
- For detection, EXIF data can be used (if present)
  - ISO/IEC 29794-5:2010
- No algorithms for detection are known
  - Call for contribution for revision of ISO/IEC 29794-5
  - CNNs could be a solution
  - Training data could be synthetically generated

Only one Face Present

- Typical requirement for reference images
  - ISO/IEC 19794-5:2011 and 39794-5:2019
- SOTA face detectors can detect all faces in an image
- Impact on face recognition limited
- Quite easy to verify, if second face can be detected

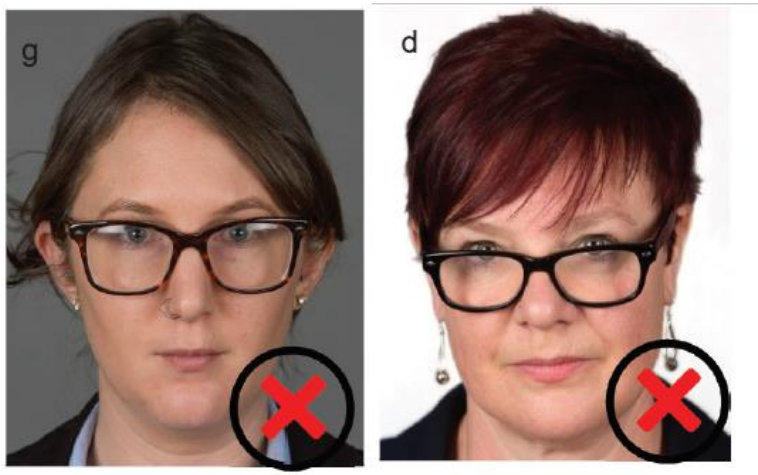
Detection of Occlusions of the Face

- Some de-occlusion CNNs explicitly segment occluded regions (occlusion mask)
  - If not, this mask can be obtained by taking difference between input and output
- Some face segmentation CNNs also output mask of un-occluded face (skin)

- Combination of occlusion mask with face landmarks allows to identify occluded face parts
- However, these CNNs are not lightweight

Detection of Occlusions by Eyeglasses

- Frames of eyeglasses should not occlude periocular region
  - ISO/IEC 39794-5:2019 defines Eyes Visibility Zone (EVZ) that should be un-occluded
- Detection of such occlusions is challenging
- No algorithms found



g) Frames partially covering the EVZ

d) Frame crossing the eyes

Eyes Open and Mouth Closed

- Requirement for reference images in ISO 19794-5 and ISO 39794-5
  - Eyes open and mouth closed
- SOTA face recognition CNNs are presumed to be quite robust
  - Other biometric algorithms (e.g. PAD, MAD) and manual inspection may benefit from open eyes and closed mouth

a) Compliant portrait c) Eyes not fully opened

- Detection of closed eyes or open mouth
  - Using facial landmarks (WD5 of ISO 29794-5)
  - CNNs for drowsiness detection (closed eyes, sometimes also yawning)

Location, Size and Completeness of the Face

- Face should be fully visible in image
  - Detailed requirements on location and size defined in ISO/IEC 19794-5:2011 and 39794-5:2019
- Impact on face recognition accuracy
  - Only if essential parts are missing
  - SOTA face recognition CNNs take small images (e.g. 112x112) as input
  - However, other biometric algorithms (e.g. PAD, MAD) can benefit from larger faces

a) Compliant face portrait | c) Invisible chin | d) Invisible crown and chin

a) True location and allowed region for the centre point M | b) Minimal size (inner rectangle) and maximal size (outer rectangle) of W and L depending on A and B

Reflections on eyeglasses

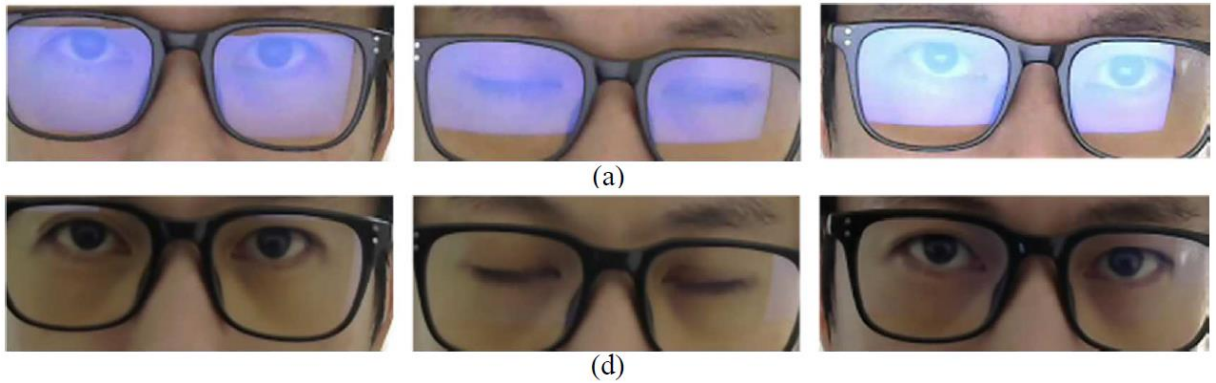
- Requirement in ISO/IEC 19794-5:2011 and 39794-5:2019
- Reflections on eyeglasses may result in occluded or spurious features
- No evaluations found of the impact on the accuracy of face recognition
  - Nevertheless, impact is plausible

a) Compliance depends on issuer policy

b) Strong reflections on glasses

Detection of reflections on eyeglasses

- No algorithms dedicated to the detection of reflections
- However, reflection removal algorithms could be used
  - Few CNNs specialized on eyeglasses
  - Many CNNs for general images



- Alternative: Detect hot spots in eye region
  - Works only for reflections of high luminance or characteristic colors

Background Uniformity

- Required for reference images by ISO/IEC 19794-5:2011 and 39794-5:2019
  - Uniform color or gradual changes from light to dark luminosity in a single direction
  - Also required by many other regulations
- No impact on accuracy of biometric recognition algorithms

- Algorithm in WD5 of ISO/IEC 29794-5
  - Segmentation by landmarks: areas in upper left and upper right of subject
    - Alternative: Identify background by subject segmentation (face parsing CNNs)
  - Entropy of luminance histogram
    - Can fail in case of gradual changes of luminosity → gradient magnitude may be better

Shoulder Pose

- Required for reference images by ISO/IEC 19794-5:2011 and 39794-5:2019
  - Shoulders frontal (parallel to the camera imaging plane)
- No impact on accuracy of biometric algorithms

- Detection by CNN for human pose estimation
  - E.g. Google mediapipe

Head Coverings

- ISO/IEC 39794-5:2019 prohibits head coverings if not specifically approved by the issuing state
  - E.g. for religious reasons
- No impact on accuracy of biometric algorithms
  - Unless face is occluded
- Detection possible by subject segmentation (face parsing CNNs)