Improved Likelihood Ratios for Surveillance Video Face Recognition with Multimodal Feature Pairing

Andrea Macarulla Rodríguez
Digital and Biometric Traces
Netherlands Forensic Institute
The Hague, The Netherlands
a.macarulla.rodriguez@nfi.nl

Zeno Geradts
Digital and Biometric Traces
Netherlands Forensic Institute
The Hague, The Netherlands
z.geradts@nfi.nl

Marcel Worring
Multimedia Analytics Lab Amsterdam
University of Amsterdam
Amsterdam, The Netherlands
m.worring@uva.nl

Luis Unzueta
Fundación Vicomtech
Basque Research and Technology Alliance (BRTA)
Donostia-San Sebastian, Spain
lunzueta@vicomtech.org

***Abstract*—The accuracy of face recognition in real-world surveillance videos plays a crucial role in forensic investigation and security monitoring systems. Despite advancements in technology, face recognition methods can be influenced by variations in pose, illumination, and facial expression that often occur in these videos. To address this issue, we propose a new method for images-to-video face recognition that pairs face images with multiple attributes (soft labels) and face image quality (FIQ). This is followed by the application of three calibration methods to estimate the likelihood ratio, which is a statistical measure commonly used in forensic investigations. To validate the results, we test our method on the ENFSI proficiency test 2015 dataset, using SCFace and ForenFace as calibration datasets and three embedding models: ArcFace, FaceNet, and QMagFace. Our results indicate that using only high quality frames can improve face recognition performance for forensic purposes compared to using all frames. The best results were achieved when using the highest number of common attributes between the reference image and selected frames, or by creating a single common embedding from the selected frames, weighted by the quality of each frame's face image.**

***Index Terms*—Face Recognition, Video processing, Face Image Quality, Likelihood Ratio, Multi Modal Analysis**

I. Introduction

Face recognition (FR) is an identification method that has become increasingly important in recent years, particularly in the field of forensic investigation [3]. With the proliferation of surveillance cameras and the capture of images of criminal events, the comparison of faces has become a key tool for gathering intelligence, guiding investigations, and providing evidence in court [3] [4]. While deep-learning based FR methods have demonstrated strong recognition performance for still images [5], such as those in the Labeled Faces in the Wild (LFW) dataset [6], video-based FR has not been as widely developed by the research community [7]. However, video FR offers additional information, such as temporal and multi-view details which can be used in conjunction with frame based face recognition techniques to quickly identify subjects of interest in CCTV footage [8].

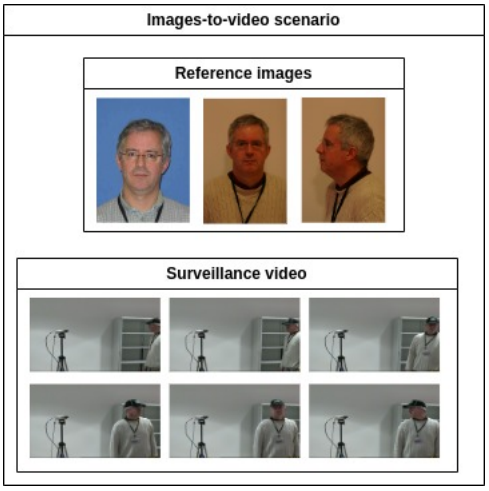

Fig. 1: Example of images-to-video scenario. Images taken from [1].

Despite the potential benefits of video-based FR, the process of analyzing such a large amount of data for each video is challenging due to the time needed to deal with all frames. Faces in these frames can be useless for recognition due to low video quality, motion blur, occlusions, and frequent changes in the scene [9], [10] (see figure 1. Some works focusing on face image quality (FIQ) [11]–[13], however, have indicated that using human-based attributes (such as resolution or illumination) for face image quality assessment may not represent the best characteristics for the face recognition

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101021797 as part of STARLIGHT.

979-8-3503-3607-8/23/$31.00 ©2023 European Union

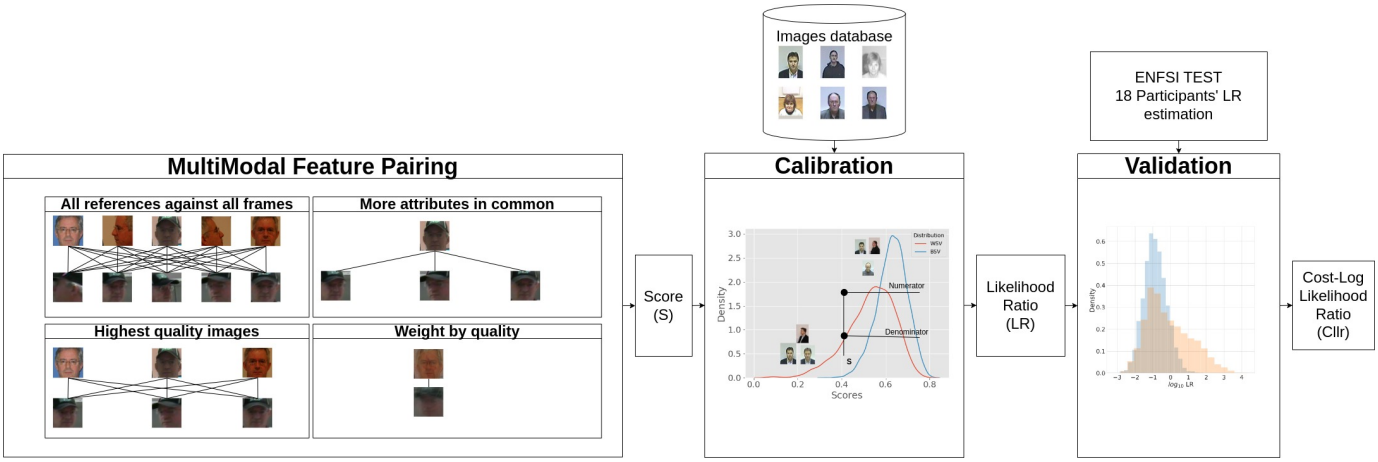

Fig. 2: LR computation workflow. First, faces are detected in reference images and the video frames. Then, pairing between the key reference images and keyframes is done either by all possible combinations, highest quality, coinciding attributes (soft labels), or quality weighted average. When the score is computed, the process of calibration can be either done using random images, the same attributes, or similar face image quality. Once calibrated, the Likelihood ratio is computed and validated against human performance with the ENFSI proficiency test 2015 using Cllr [2] as evaluation measure.

system being used. These authors use the SER-FIQ, MagFace, and SDD-FIQA face image quality assessment methods to test on IJB-C [14] videos, but only as 1:1 on individual frames, not using spatio-temporal information.

In automated facial recognition systems, the similarity between two samples is usually reported in terms of one or several score values which are intrinsic to each version of the facial recognition algorithm used [5]. In order to allow for comparisons between facial scores from different face recognition systems as well as for such an automated comparison to be useful in an evaluative forensic framework, there is a need to map the output scores to a Likelihood Ratio (LR) [15], which is defined as the probability of the evidence given hypothesis H_0 (i.e., the probability of the reference being the same person as in the video) divided by the probability of the evidence given the alternative hypothesis H_1 (i.e., the probability of the reference being a different person than the one appearing in the video). A possible approach to achieve this is to append such a score-to-LR mapping as a post-processing step in an existing score-producing facial recognition system [16]. Once a model for score-to-LR mapping has been set up, the forensic reporting can be presented using a level of conclusion where each grade on the scale is connected to an interval of LR values [4], [17].

In this paper, we propose a method for images-to-video face recognition in realistic forensic scenarios by utilizing a new model that pairs face images based on multi-modal face feature data like face attribute characteristics and FIQ. We aim to address the question of how to accurately estimate likelihood ratios (LRs) for face recognition systems in practice. In particular, we focus on a scenario where several reference images of a suspect or person of interest are available, and the goal is to determine whether this person is the same as the one appearing in a surveillance video. We aim to improve the accuracy of likelihood ratio (LR) estimation in automated face recognition using images-to-video comparisons. We then apply three calibration methods to estimate likelihood ratios and validate the results using the log-likelihood ratio cost (Cllr). Our contributions include the following:

1) **MultiModal Feature Pairing** using FIQ to select frames with highest quality and highest common attributes (soft labels), and combining them through weighted average.

2) **Calibration** involving selection of random pairs with the same attributes and same FIQ as the test pairs.

3) **Validation** Evaluation of the LR estimation system against a forensic test performed with human experts.

The current study begins by providing an overview of the relevant literature pertaining to the estimation of likelihood ratios, face recognition in video, and the incorporation of FIQ in face recognition in images-to-video scenarios. Following this, the methodology for pairing and calibration is presented. The experiments and associated results are then discussed. Finally, the paper concludes with a discussion of the findings and implications. The workflow for the computation of the Likelihood Ratio is depicted in Figure 2.

II. Related Work

Likelihood ratios (LRs) have been extensively studied in the field of face recognition, with Molder et al. [16] testing score-to-LR models in forensic data and finding that the performance of the models depends on the available training data. Rodriguez et al. [18] improved LR estimation by using facial attributes and quality scores, and found that commercial software outperforms open-source software. Jacquet et al. [4] explored the importance of LR in face recognition, assessing the performance of the model with respect to its discriminating power and calibration state. Despite these efforts, the challenge of accurately estimating LRs for face recognition systems in practice remains an open question.

Spatio-temporal face recognition in videos has been explored by several researchers. Zheng et al. [10] present a system for image-to-video face recognition that performs well on video datasets. Huo et al. [19] examine n-shot face recognition in videos using metric learning and find that triplet loss outperforms contrastive loss. Rivero et al. [20] suggest an adaptive aggregation scheme for image-to-video face recognition and test its suitability. Despite these advances, the problem of face recognition in videos remains open and the results are not generalizable to real-world scenarios.

Research on keyframe extraction for face recognition in videos has also been conducted. Abed et al. [7] proposed a method based on face quality and deep learning which involves two steps: generation of face quality scores using Gabor, LBP, and HoG feature extractors, and training of a Convolutional Neural Network to select frames with the best face quality. Bahroun et al. [9] also proposed a keyframe extraction method based on face quality, reducing data by rejecting frames without faces and clustering face images by identity and then selecting a frame with the best face quality based on four metrics: pose, sharpness, brightness, and resolution. However, the effectiveness of this quality assessment has been questioned by [11] and [12] who suggest other methods should be considered.

Face image quality assessment to improve face recognition in videos has been studied by Terhorst et al. [11] who propose SER-FIQ (Subjective and Objective Quality Factors of Images), which they tested on IJB-C [14] videos, showing good performance in face recognition tasks. Meng et al. [12] propose MagFace using a multi-attention guided network, outperforming state-of-the-art methods when tested on IJB-C videos. Ou et al. [13] propose SDD-FIQA (Single Shot Detector based Face Image Quality Assessment), tested on IJB-C videos, performing well in face recognition tasks. Although they propose interesting methodologies, these works only evaluate 1:1 (face verification) image-to-video scenarios, not considering the use of video frame information in video sequences for recognition.

III. Methodology

The question we aim to answer is: How likely is this person the same as the one appearing in the surveillance video? To that end, we propose a workflow as seen in figure 2. Our focus is on the comparison of several reference images of the same person to a video in order to determine if the person appears in the video.

To estimate the likelihood ratio, the biometric score obtained from the comparison between the images and the video has to go through a process of calibration in which two distributions are computed: the within source variability (WSV) and the Between source variability (BSV). In this paper, we focus on two specific aspects of this process as it pertains to images-to-video comparisons: (1) methods for pairing reference images with videos, and (2) the use of different types of images to create the WSV and BSV distributions during the calibration step. To estimate the likelihood ratio, the biometric score obtained from the comparison between the images and the video must be calibrated using these distributions. LR based on biometric similarity scores is referred as Score based Likelihood Ratio (SLR) and defined as:

$$SLR = \frac{P(s|H_p, I)}{P(s|H_d, I)}, \tag{1}$$

where H_p is the null hypothesis or the prosecution hypothesis (evidence originates from the same source) and H_d is the alternative hypothesis or defence hypothesis (evidence originates from a different source). The value s is the score returned by the biometric system and I is the background information available in the case apart from the evidence. Although it can be used for any type of forensic evidence (such as DNA or fingerprints), in our work it corresponds to face evidence. In our case, we used the Logistic Regression [21] to fit the $P(s|H_p, I)$ and $P(s|H_d, I)$ functions.

A. MultiModal Feature Pairing

In this work, we aim to improve the accuracy of likelihood ratio (LR) estimation in automated face recognition using images-to-video comparisons.

One approach involves using score pairs that are based on the common attributes (i.e. soft labels) between the reference image and the video frame. These attributes include: gender, age, emotion, race, yaw, pitch, roll, headgear, glasses, beard, and other occlusions. Let there be m reference images and k video frames. We start by extracting and computing the attributes of all the reference images and video frames. Let R_i be the set of attributes for reference image i, where $i \in [1, m]$, and let F_j be the set of attributes for video frame j, where $j \in [1, k]$. We then compare the attributes of each reference image with the attributes of each video frame, and select those pairs that have the highest number of attributes in common. Let the number of common attributes between the i-th reference image and the j-th video frame be denoted by n_{ij}. We then select all score pairs that have exactly n attributes in common, and perform likelihood ratio estimation on these selected pairs. We start with n being the highest number of common attributes. From there, we iteratively repeat the process with pairs that have one attribute less in common until there are no common attributes. When making pairs with the same number of common attributes, the order within those sets is not taken into account. By considering the attributes and iterating through different number of shared attributes, the algorithm can make more informed decisions and potentially improve the overall accuracy of the face recognition system.

An alternative approach for pairing involves matching all m reference images with all k video frames, and then sorting the pairs according to their quality. The LR is calculated using all pairs, and subsequently, a process of pruning is applied. The pruning process starts with the removal of 10% of the pairs with the lowest quality, followed by the removal of an additional 10% of the pairs, and so on. The aim of this method is to determine whether the information lost by discarding pairs is valuable, i.e., the LR improves as this would indicate

that the discarded images were noisy and, thus, detrimental to the face recognition system.

In addition, we propose to process all the frames in which a face is detected in the video, compute the FIQ of each frame, and create a combined embedding vector for the video using a weighting scheme based on the FIQ scores. Similarly, we process all the available reference images. This method is based on the following equation:

$$\mathbf{e_{face}} = \sum_{i=1}^{n} q_i * \mathbf{e_i}, \tag{2}$$

where $\mathbf{e_{face}}$ is the embedding vector of the face image, q_i is the FIQ of frame i and $\mathbf{e_i}$ is the embedding of each face image. This process is applied to both the video frames and the reference images.

B. Calibration

To improve the accuracy of the LR estimation for automated face recognition in video, we will consider three different approaches for selecting images from the calibration database to use in the estimation process:

- **Random selection:** Using random images from the calibration database as a baseline.
- **Same attributes:** Using images with the same attributes as the reference and video, such as pose or facial expression.
- **Quality pairs:** Using pairs that have the same FIQ for the reference face and the combined face image qualities of the video frames.

By implementing these approaches, we aim to improve the accuracy of the LR estimation for automated face recognition in video.

C. Validation

To assess the performance of our proposed methods, we use the log cost likelihood ratio (Cllr) as a measure due to its capacity to represent both discrimination and calibration [15]. Cllr is defined as:

$$C_{llr} = \frac{1}{2N_p} \sum_{i_p} \log_2(1 + \frac{1}{SLR_{i_p}}) \frac{1}{2N_d} \sum_{j_d} \log_2(1 + SLR_{j_d}), \tag{3}$$

where the indices i_p and j_d respectively denote summing over the computed SLR scores using equation 1 for each face pair comparison where the respective proposition for prosecutor or defense is true, with N referring to the number of samples. Minimizing the value of Cllr implies an improvement of both discrimination and calibration performance of the automated system [15]. The value ranges from zero (perfect decision making), to infinity (completely wrong). A value of one indicates the system makes a random selection. A value larger than one indicates that the system is making a decision worse than random, i.e. supporting the prosecution hypothesis when it should be supporting the defence hypothesis or vice versa. In addition, we also use boxplots to assess the impact of discarding pairs on the variability of our results, for both i_p (the summands corresponding to the prosecutor's proposition) and j_d (the summands corresponding to the defense's proposition). Specifically, we plot boxplots on the Cllr metric for each quality drop, indicating the percentiles of 25, median, and 75. The use of boxplots allows us to visualize the distribution of the Cllr metric and better understand how discarding pairs impacts the variability of the results, measure our approach for validating the performance of our proposed methods, and assessing the impact of discarding pairs on the variability of the results.

IV. Experiments

We will explore the workflow explained in section III doing experiments in the two parts of the method: pairing and calibration.

- **Datasets**. Our study uses three datasets: ENFSI proficiency test [2], ForenFace [1] and SCFace [22]. ENFSI proficiency test 2015 focuses on mugshot images to CCTV video and includes 18 individual participants in 17 CCTV to mugshot facial image comparisons. ForenFace contains video sequences and extracted images of 98 subjects recorded with six different surveillance cameras, and also includes a training/testing protocol. SCFace is designed for experiments in person identification/verification under real-world surveillance conditions, with 130 subjects in uncontrolled indoor environments using five video surveillance cameras of various qualities. All datasets consist of video sequences and face images with variations in illumination, pose, and sharpness. The goal of the study is to train and test the performance of the proposed method on these datasets and to find the best method to improve the accuracy of the LR estimation for automated face recognition in video.

- **Face recognition models and Face quality models**. The experiment uses three face recognition models: ArcFace [23], Facenet [24], and QMagFace [25], which are recent methods with state-of-the-art performance and different characteristics. We adopted ArcFace and Facenet from the library serengil2020lightface [26]. Two quality models, SER-FIQ [11] and SDD-FIQA [13], are used, as they are both unsupervised and have shown to outperform state-of-the-art quality assessment methods with good generalization across different recognition systems.

V. Results

Results of four experiments on the effect of using different pairing methods on LR estimation in face recognition in videos are presented in figure 3.

- **Experiment 1: Highest number of common attributes**. We aim to assess if using pairs that share attributes (multi-attribute, i.e. pairs that have the same pose or the same facial expression) outperforms using pairs that have nothing in common. We do the LR estimation with 10000 random images from the calibration set and 0 to 6 attributes in common (pitch, yaw, roll, facial expression, age and gender). The results show that the higher number of attributes in common, the lower the Cllr.

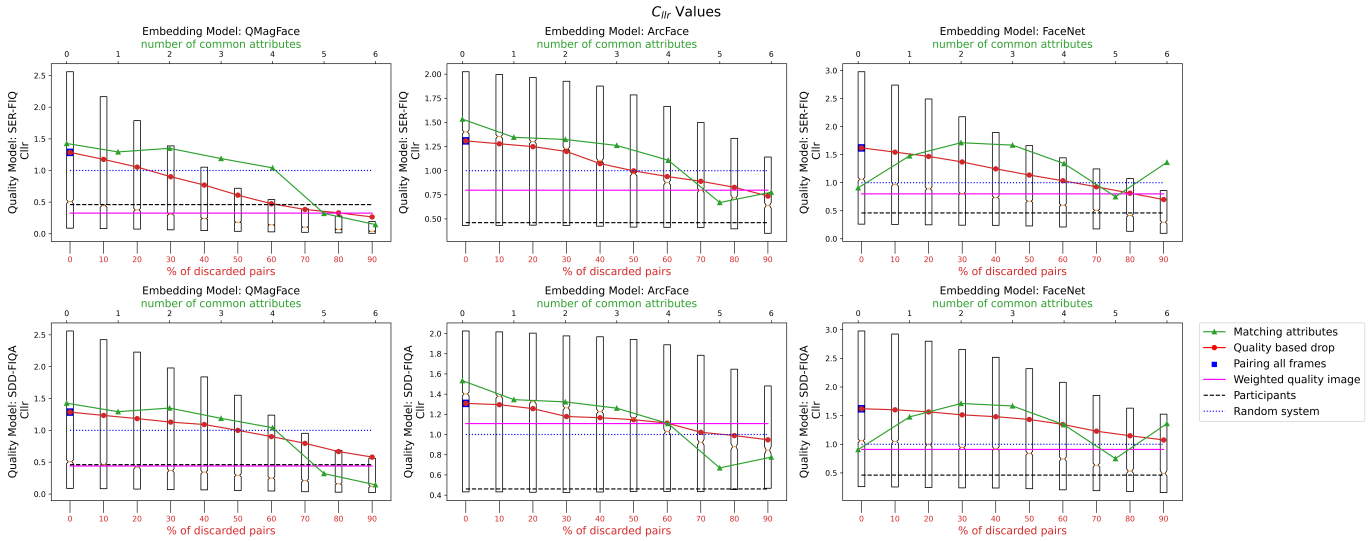

Fig. 3: Cllr computed for the ENFSI year 2015, in which the scenarios of 4 pairing methods (all frames, highest quality frames, weighted quality face image and highest number of attributes is presented). The boxplots help to interpret the variability in the Cllr when discarding pairs with the quality drop method.

- **Experiment 2: Quality based drop**. We aim to assess how much using the highest quality frames influences the LR estimation. We compare the performance of the LR estimation with 10000 random images from the calibration dataset and compute the LR estimation by dropping 10% of the poorest quality face images in each iteration. We use the ENFSI test 2015 dataset for this experiment. The results show that the higher the quality of the frame, the lower the Cllr and therefore the better the LR estimation. However, increasing the number of frames used does not necessarily improve the Cllr. In fact, using more frames with lower quality can result in a worse Cllr. These results suggest that selecting high quality frames is important for improving the accuracy of LR estimation in automated face recognition in videos.

- **Experiment 3: Weighted face quality images**. We aim to assess if using all the frames, but giving those with lower quality less impact on the final result (i.e. making an averaged face embedding by quality), affects the LR estimation. We compare the performance of the LR estimation with 10000 random images from the calibration dataset in two different scenarios: using all the combinations of possible pairs or using one single distance in which the two embedding vectors are a weighted average by quality of all the embedding vectors available for the reference and the frames. The results show that using all the frames taking the quality into account lowers the Cllr and therefore gives a better LR estimation.

- **Experiment 4: Calibration**. Once the comparison of images-to-video is computed, the difference in calibration can be appreciated by using random images or images that have the same FIQ as the test pair. For that, experiments 1-3 are repeated using three different settings: with calibration of 10000 random images, 10000 images with the same attributes, or 10000 images that have the same quality as the test pair.

VI. Discussion & Conclusion

In our experiments, we found that using higher quality frames improves the performance of face recognition in video compared to using all frames. We explored different methods for pairing reference images with video frames and found that using images with the same attributes as the reference and video, or similar FIQ score forthe reference face and the combined face image qualities of the video frames, can improve the accuracy of the likelihood ratio estimation. Furthermore, we found that using a weighted quality average of all available reference and video frames improved results even more. On the other hand, slightly poorer results were obtained when pairing facial images based on the maximum number of common attributes. Although SDD-FIQA [13] outperforms SERFIQ in the LFW [6] and IJB-C [14] benchmarks, SERFIQ [11] seems to be more robust in our experiments. The Cllr obtained in the best case is close to 1, which is worse than the 0.45 of the expert participants in the ENFSI proficiency test [2]. This could be due to the difficulty and low face quality of the video frames used. Even discarding those with the poorest quality, the remaining ones are not suitable for the face recognition system in this experiment (ArcFace). However, using Facenet as the face recognition system in our experiments, we were able to achieve a Cllr of 0.8, which is a better result. With QMagFace, we achieved even better results, with a Cllr of 0.26 using the method of the weighted quality image, surpassing the human participants in the ENFSI 2015 test, who scored a Cllr of 0.46. The best result was obtained by QMagFace and SER-FIQ with the method of pairing the highest number of attributes in common, with a Cllr of 0.13. This demonstrates the effectiveness of using FIQ as a metric to improve the performance of automated face recognition in video surveillance. The boxplots suggest that there is less

variability when a greater number of pairs are excluded. It is worth noting that in surveillance settings, errors in attribute estimation can occur, which may affect the accuracy of face recognition systems that rely on shared attributes to select reference images and video frames. It is, therefore, crucial to investigate how errors in attribute estimation impact the performance of the proposed method that pairs the highest number of attributes in common. It is also important to explore alternative approaches for selecting reference images and video frames that do not solely rely on shared attributes, such as deep metric learning, which can learn discriminative features for face recognition directly from the data. Future work should also consider examining the proposed methods on more diverse datasets, including those that present greater variability in facial attributes, to ensure the generalizability of the findings. Our results show the potential for using FIQ, spatio-temporal information and additional information such as gait, clothes, or hair to improve the performance of automated face recognition in video surveillance. Further research could explore the use of additional metrics for keyframe selection, and examine the performance of the proposed methods on a wider range of datasets and face recognition algorithms.

REFERENCES

[1] C. G. Zeinstra, R. N. Veldhuis, L. J. Spreeuwers, A. C. Ruifrok, and D. Meuwly, "Forenface: a unique annotated forensic facial image dataset and toolset," *IET biometrics*, vol. 6, no. 6, pp. 487–494, 2017.

[2] R. Moreton, F. Eklöf, and A. Ruifrok, "Facial image comparison test 2015," 2015.

[3] A. Ruifrok, P. Vergeer, and A. M. Rodrigues, "From facial images of different quality to score based lr," *FSI*, vol. 332, p. 111201, 2022.

[4] M. Jacquet and C. Champod, "Automated face recognition in forensic science: Review and perspectives," *Forensic science international*, vol. 307, p. 110124, 2020.

[5] S. P. K Wickrama Arachchilage and E. Izquierdo, "Deep-learned faces: a survey," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 1, pp. 1–33, 2020.

[6] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[7] R. Abed, S. Bahroun, and E. Zagrouba, "Keyframe extraction based on face quality measurement and convolutional neural network for efficient face recognition in videos," *MTA*, vol. 80, no. 15, pp. 23 157–23 179, 2021.

[8] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 1002–1014, 2017.

[9] S. Bahroun, R. Abed, and E. Zagrouba, "Ks-fqa: Keyframe selection based on face quality assessment for efficient face recognition in video," *IET Image Processing*, vol. 15, no. 1, pp. 77–90, 2021.

[10] J. Zheng, R. Ranjan, C.-H. Chen, J.-C. Chen, C. D. Castillo, and R. Chellappa, "An automatic system for unconstrained video-based face recognition," *IEEE TBBIS*, vol. 2, no. 3, pp. 194–209, 2020.

[11] P. Terhorst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "Serfiq: Unsupervised estimation of face image quality based on stochastic embedding robustness," in *Proceedings of the IEEE/CVF*, 2020, pp. 5651–5660.

[12] 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*, 2021, pp. 14 225–14 234.

[13] 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/CVF*, 2021, pp. 7670–7679.

[14] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney *et al.*, "Iarpa janus benchmark-c: Face dataset and protocol," in *2018 ICB*. IEEE, 2018, pp. 158–165.

[15] D. Ramos, R. P. Krish, J. Fierrez, and D. Meuwly, "From biometric scores to forensic likelihood ratios," in *Handbook of biometrics for forensic science*. Springer, 2017, pp. 305–327.

[16] A. L. Mölder, I. E. Åström, and E. Leitet, "Development of a score-to-likelihood ratio model for facial recognition using authentic criminalistic data," in *2020 8th IWBF*. IEEE, 2020, pp. 1–6.

[17] ENFSI, *Best Practice Manual for Facial Image Comparison*, European Network of Forensic Science Institutes (ENFSI), 2018.

[18] A. M. Rodriguez, Z. Geradts, and M. Worring, "Calibration of score based likelihood ratio estimation in automated forensic facial image comparison," *FSI*, vol. 334, p. 111239, 2022.

[19] J. Huo and T. L. van Zyl, "Unique faces recognition in videos," in *2020 IEEE 23rd FUSION*. IEEE, 2020, pp. 1–7.

[20] J. Rivero-Hernández, A. Morales-González, L. G. Denis, and H. Méndez-Vázquez, "Ordered weighted aggregation networks for video face recognition," *PRL*, vol. 146, pp. 237–243, 2021.

[21] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[22] M. Grgic, K. Delac, and S. Grgic, "Scface–surveillance cameras face database," *MTA*, vol. 51, pp. 863–879, 2011.

[23] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF*, 2019, pp. 4690–4699.

[24] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE CVPR*, 2015, pp. 815–823.

[25] P. Terhörst, M. Ihlefeld, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper, "Qmagface: Simple and accurate quality-aware face recognition," in *Proceedings of the IEEE/CVF*, 2023, pp. 3484–3494.

[26] S. I. Serengil and A. Ozpinar, "Lightface: A hybrid deep face recognition framework," in *2020 ASYU*. IEEE, 2020, pp. 23–27. [Online]. Available: https://doi.org/10.1109/ASYU50717.2020.9259802