Advancing Contextual Face Image Quality Assessment with the U3FQ: A Unified Tri-Feature Metric

Anonymous CVPR submission

Paper ID *****

Abstract

In the arena of face image quality assessment (FIQA), the incorporation of contextual nuances is critical for accuracy. Our novel approach, the Unified Tri-Feature Quality Metric (U3FQ), uniquely amalgamates three pivotal elements: age variance, facial affect, and congruence scores. Exploiting the extensive AGEDB dataset, U3FQ employs a semi-reference paradigm, augmented by profound learning algorithms, to achieve meticulous calibration of the model. U3FQ distinguishes itself by meticulously adjusting congruence scores using quantitative modifiers predicated on age discrepancies and expression intensities, guaranteeing that the resultant quality metric authentically encapsulates the true likelihood of a match. Rigorously appraised against the exacting NIST MEDII dataset, our methodology also undergoes comparative evaluations with established FIQA frameworks. Our metric heralds a substantial progression in the domain of FIQA, proffering an all-encompassing quality assessment that is deeply rooted in theoretical rigor while remaining acutely pertinent to a multitude of facial recognition contexts.

1. Introduction

2. Related Works

3. Methodology

Our methodology is anchored in the creation of a novel Unified Tri-Feature Quality Metric (U3FQ) for Contextual Facial Image Quality Assessment (FIQA), which is pivotal for enhancing the accuracy of biometric systems. This section elucidates the integrated approach we have employed to devise, enhance, and configure the U3FQ metric, encapsulating the synergetic impact of match scores, age differences, and facial expressions on image quality. It further expounds on the architecture of both conventional machine learning models and advanced deep learning networks used in our study. In Subsection 3.1, we detail the theoretical underpinnings of U3FQ, highlighting its relevance in FIQA and providing a historical context for the selected features. Subsection 3.2 delves into our method's practical application, describing the interplay between age-modulated match scores and expression-weighted adjustments within the model's fine-tuning process, using the AGEDB dataset. The architectural framework, comprising Random Forest and ResNet models, is discussed in Subsection 3.3, which delineates the feature integration for quality score prediction. Finally, Subsection 3.4 presents an analytical study on various computational strategies for match score calculation and the treatment of non-mated pairs, with insights drawn from the NIST MEDII dataset for robust recognition performance

3.1. Background

In the pursuit of advancing FIQA, the U3FQ metric incorporates a novel perspective on biometric system utility, converging on the quintessential attributes of match score, age difference, and facial expression. The conceptual foundation of U3FQ is predicated on the premise that the quality of a facial image in biometric systems is inherently linked to these three factors. A succinct overview of age difference and facial expression, within the context of their historical application in FIQA, sets the stage for our innovative metric.

Figure 1. The impact of age difference on face match score integration.

Figure 2. Influence of facial expression variations on match score adjustments.

3.2. Formulations and Optimizations

The efficacy of face matching systems is significantly influenced by the age difference between the anchor image and the comparison image. This influence varies notably with the anchor's age, necessitating a nuanced approach to modeling age difference penalties.

For anchors aged between 20 and 30 years, negative age differences typically correlate with child images, which present a considerable challenge due to the substantial facial feature changes that occur during maturation. As such, the penalty for negative age differences should be higher within this age range. Conversely, for anchors over 35 years of age, negative age differences represent younger adult images, where changes in facial features are less pronounced. Therefore, the penalty for negative age differences should be attenuated.

The mathematical formulation of the age difference penalty function can be adapted to account for this behavior:

$$f(d,a) = \begin{cases} e^{\alpha(d+\beta)} & \text{if } d < 0 \text{ and } a \leq 30 \\ e^{\alpha(d+\beta)/\theta} & \text{if } d < 0 \text{ and } a > 30 \\ \gamma \cdot d & \text{if } d \geq 0 \end{cases} \tag{1}$$

Here, d represents the age difference between the anchor and the comparison image, a denotes the anchor's age, and α, β, and γ are parameters dictating the function's shape. The factor θ serves as a damping parameter that reduces the penalty for older anchors, aligning with the observed trends in facial aging.

This model underscores the importance of considering the anchor's age in face matching systems. By calibrating the age difference penalties accordingly, we can enhance the system's robustness and accuracy, particularly in scenarios involving significant age progression or regression.

To empirically underpin this approach, we present Detection Error Tradeoff (DET) plots that demonstrate the variance in performance with different age gaps. For anchors aged 20-25, there is a pronounced increase in the False Non-Match Rate (FNMR) as the age difference becomes more negative, indicative of matching with significantly older faces. This trend gradually inverts as the anchor age increases, reflecting the maturation and stabilization of facial features over time.

These empirical findings underscore the critical role that age plays in the performance of face-matching systems. The integration of such a nuanced understanding into the FIQA model promises a more robust framework that can dynamically adjust to the intricacies of human aging and expressions, thereby enhancing the overall system reliability.

Our methodology concentrates on the integration of key factors into the facial match score, specifically age differences and facial expressions. The approach accounts for the subtle yet significant influence of facial expressions, distinguishing between 'weak' emotions—such as smile, sadness, or a neutral expression—which are used directly in the

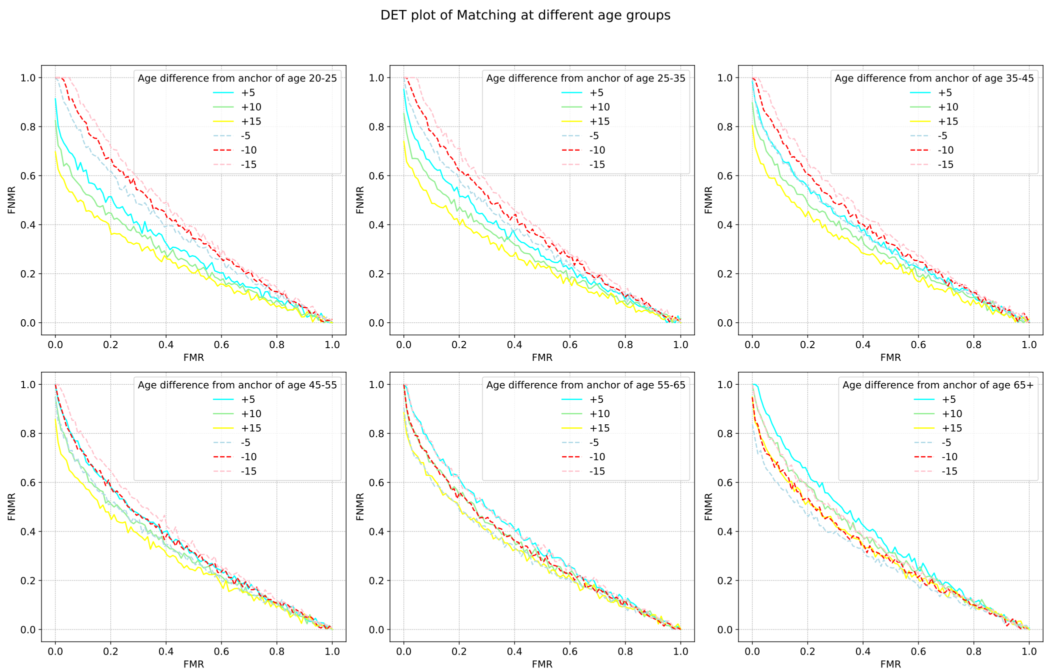

Figure 3. DET subplots for different age differences in face matching systems, illustrating the varying impact of age difference across different anchor ages.

match score due to their minimal impact on facial recognition accuracy, and 'strong' emotions—like surprise, anger, or happiness—which alter the match score to reflect their pronounced effect on recognition. This differentiation is quantified using a facial expression impact function $g(e)$:

$$g(e) = \begin{cases} c & \text{if } e \text{ is a weak emotion} \\ d \cdot \text{EXPR\_SCORE}(e) & \text{if } e \text{ is a strong emotion} \end{cases} \tag{2}$$

In Equation 2, c is a constant factor that applies when the facial expression is weak, and d is a multiplier that scales the expression score EXPR_SCORE(e), which is determined by the intensity of strong emotions. This function $g(e)$ is then utilized to adjust the match score accordingly, as part of the comprehensive quality assessment.

Together with the age difference penalty function previously defined in Equation ?? and depicted in Figure ??, these formulations collectively enhance the fidelity of the FIQA model's predictions.

3.3. Architecture

Initially, a Random Forest model was utilized to assess the impact of incorporating match scores, age difference, and facial expression data on image quality predictions. Subsequently, we transitioned to a deep learning framework, employing ResNet50 and ResNet18 architectures to

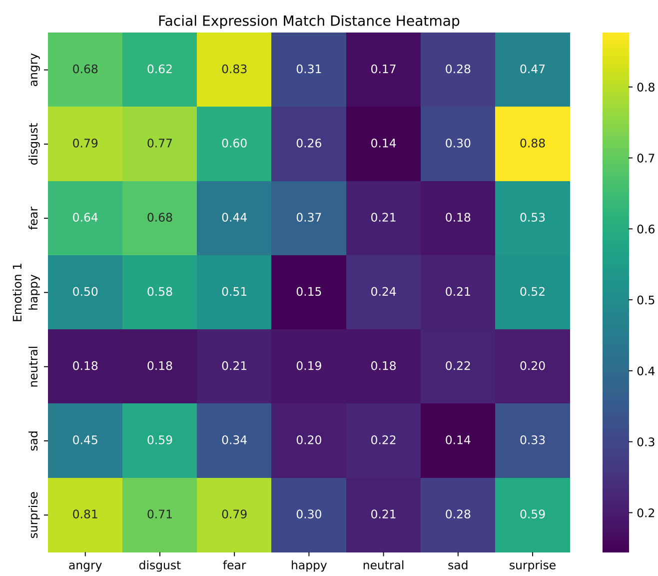

Figure 4. The differential impact of facial expressions on the match score, with weak emotions having a constant effect and strong emotions modifying the score proportionally to their intensity.

further refine the predictive capability of our system.

The following diagram delineates the workflow pipeline,

Figure 5. The architecture of the proposed U3FQ metric incorporating deep learning models.

highlighting the integration of the deep learning network with the Random Forest model to compute the face quality score.

Figure 6. Workflow pipeline utilizing ResNet models and Random Forest for quality score prediction.

3.4. Analytical Study

This subsection embarks on a discourse surrounding alternative methodologies that hold potential for future exploration. Considerations include the grouping of facial images by expression, the differential impacts of varying age differences, and other conceivable permutations that could influence the FIQA landscape.

Figure 7. Visualization of alternative approaches and potential experiments for advancing FIQA.

4. Experiments

CVPR
#*****

5. Results

6. Conclusion

References