METROLOGICAL EVALUATION OF MULTIMODAL 3D CAMERA SYSTEM FOR RELIABLE DYNAMIC FACIAL MOTION ANALYSIS

Wang Liao [a, *], *Chen Zhang* [a], *Shiyao Gao* [a], *Hongyu Chen* [a], *Hao Chen* [a], *Maik Rosenberger* [a], *Gunther Notni* [a, b]

[a]Ilmenau University of Technology, Ilmenau, Germany
[b]Fraunhofer Institute for Applied Optics and Precision Engineering, Jena, Germany
*Corresponding author: wang.liao@tu-ilmenau.de

Abstract: Traditional 2D-based dynamic facial motion analysis methods often rely on texture features and lack depth information, raising the fundamental question: can facial motions be measured rather than inferred? This work proposes a multimodal 3D measurement-based concept for dynamic facial motion analysis, leveraging a camera system that combines a GOBO-based active stereo 3D unit with synchronized RGB camera. A pretrained neural network extracts 2D facial landmarks, which are mapped into 3D-metric space. As proof-of-concept, 19 geometric features (e.g., distances and angles) are defined in based on 3D metric facial landmarks to correspond to selected Action Units (AUs). The concept is evaluated from a metrological perspective. Under static condition, where facial expressions remain unchanged, 90 frames were captured per video across 30 recordings for each of six participants. The average standard deviations of the defined features measurements were 0.85° (angles) and 0.37 mm (distances), yielding expanded uncertainties of 2.6° and 1.1 mm (99.73% confidence). Feasible measurement capability of the camera system is further supported by low variation coefficients (1.33% and 1.61%) of the measured features values. To further evaluate dynamic capability, facial motions were induced, producing feature changes much greater than the uncertainty, confirming statistical significance. The proposed concept and camera system provide a robust and precise foundation for dynamic facial motion analysis, with potential applications in scenarios like micro-expression recognition, lie detection, and healthcare monitoring.

Keywords: dynamic facial motion analysis, measurement uncertainty, metrology-based evaluation, camera system, 3D measurement

1. INTRODUCTION

Facial Action Units (AUs) characterize the facial muscle movements and serve as a fundamental basis for tasks such as emotion recognition, dynamic facial motion analysis and human behaviour analysis, including lie detection [1]. Traditional approaches typically predict AUs from 2D RGB images using conventional machine learning [2] or deep convolutional neural networks [3]. Metric measurements cannot be derived from 2D data due to the absence of depth information, which results in scale distortion (e.g., objects appearing larger when closer) and limits the representation to 3D-to-2D projections, causing attenuation of motion components in the image plane during head rotation. 2D methods primarily rely on visual texture features rather than capturing the actual geometric deformations, leaving a fundamental question unresolved: can facial motions be truly measured rather than simply inferred? Compared to conventional 2D imaging, 3D-metric measurements enable the direct capture of actual muscle geometric deformations, which may be possible to address the inherent limitations of 2D-based dynamic facial motion analysis, including poor cross-subject generalization ability and sensitivity to head-pose variations [4]. In this work, we introduce a multimodal 3D measurement-based concept for dynamic facial motion analysis (e.g. AU recognition) and the metrological evaluation of the capability of this camera system for our proposed concept.

2. PROPOSED CONCEPT

In this work, we introduce a multimodal 3D measurement-based concept for dynamic facial motion analysis (e.g. AU recognition) and the metrological evaluation of the capability of this camera system for this task. For our concept, firstly, we integrate a multimodal camera system consisting of a 3D unit composed of a GOBO (**GO**es **B**efore **O**ptics) projector [5] operating at the 850 nm wavelength and two NIR cameras of the same wavelength for active stereo matching, along with a synchronized RGB camera. All intrinsic and extrinsic parameters of the cameras were well calibrated. Next, as shown in Figure 1, we employ a pretrained neural network model provided by MediaPipe [6] to extract 468 2D facial landmarks from the captured facial color image and then map these landmarks into the metric facial point cloud. The 3D information can be then project to the 2D facial landmarks extracted on color image to obtain 3D-metric facial landmarks. Based on these 3D-metric facial landmarks, further 19 facial geometric features (F#1 to F#19) are defined, such as distances between selected landmarks and angles formed by connecting specific landmarks, which are corresponding to some of Aus as proof-of-concept. By measuring and analysing the variation of these facial geometric features, we can quantify dynamic facial motions. To evaluate the capability of this multimodal 3D camera system from a metrological perspective, we assume that facial expression remains unchanged over a very short time interval, i.e., under static condition, facial motions can be treated as negligible. Under this static condition, the camera system captures a series of images (e.g. 90 frames) to repeatedly measure the aforementioned defined geometric features. In this way, the standard deviation σ of the feature measurements and the corresponding expanded uncertainty ($U = k\sigma$) can be calculated, where k can be determined according to ISO/IEC Guide 98-3 [7]. Next, we observe the dynamic ranges in the defined facial geometric features when facial muscles move. By comparing with the expanded uncertainty determined under static condition, we can evaluate then with what level of confidence the camera system can detect different facial motions.

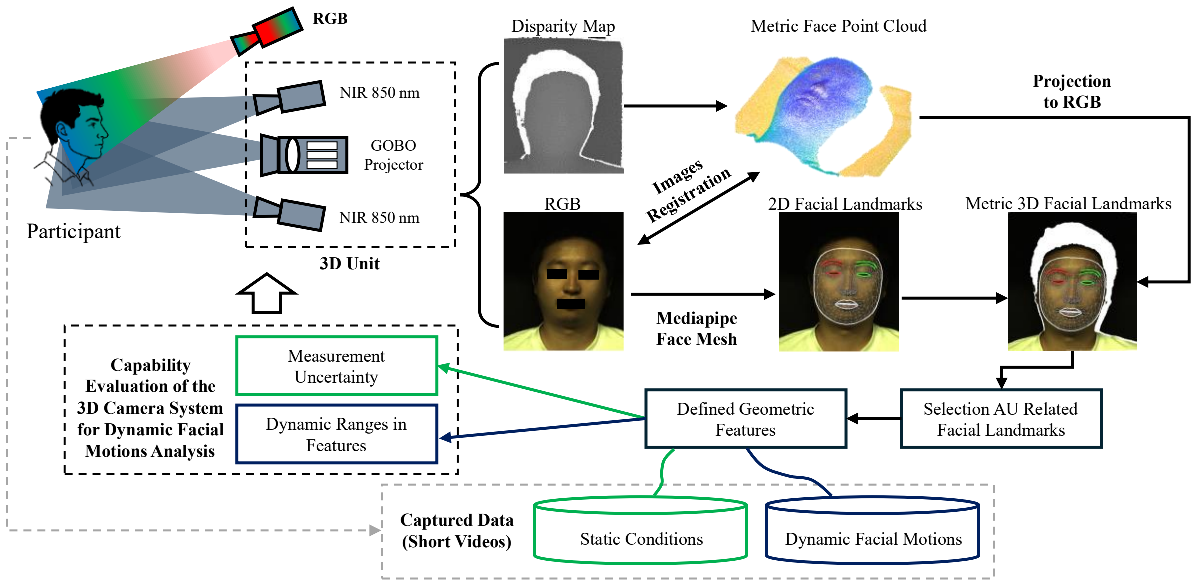

Figure 1. Flow chart of proposed metric 3D based concept for dynamic facial motions analysis and metrological evaluation approach

3. EXPERIMENT AND RESULTS

In the experiment, the camera system operates at a frame rate of 30 Hz. Six participants (P#1 to P#6) were invited, and each recorded 30 facial video clips under static condition, with each clip lasting 3 seconds (90 frames in total). During recording, participants were instructed to maintain a neutral facial expression and to avoid any facial or body movements. Since the facial muscles stayed relatively consistent throughout this short period, the 90 frames can be treated as 90 instances of repeated measurements. The primary variation among the 30 video clips per participant stemmed from changes in head orientation and spatial position. Additionally, to induce controlled facial motion, participants were instructed to perform several easily-expressible Action Units (e.g. AU 18). These data support analysis of the significant level of dynamic motion-induced feature changes compared with the system's measurement uncertainty.

For each captured video under static condition, we calculated the standard deviation for each defined facial geometric feature across the 90 frames of the video to assess the measurement fluctuations. Next, we averaged the standard deviations across 180 videos from the 6 participants, considering this as the "overall" standard deviation for each feature under static condition. For angle-related features, the overall standard deviation ranged from 0.81° to 0.90° (avg. 0.85°); for distance-related ones, from 0.10 mm to 0.79 mm (avg. 0.37 mm). This suggests that, according to ISO/IEC Guide 98-3 [7], our camera system can detect changes as small as 2.6° in angular features and 1.1 mm in distance-based features with 99.73 % confidence. We also calculated the "overall" coefficient of variation (CV) to assess relative dispersion. For angles, average CV was 1.33 %; for distances, 1.61%. The results provide a solid measurement foundation for dynamic facial motion analysis and even micro-expression recognition.

The significance of the changes in specific defined facial geometric features induced by AUs, relative to the system's measurement extended uncertainty can be illustrated through the example of participant P#1 and AU 18. AU 18 represents Lip Puckerer, which is related to the defined facial geometric features F#7, F#8, F#10, and F#11. The measurement expanded uncertainties for these four features of P#1 are 0.39 mm, 0.30 mm, 0.23 mm, and 1.91 mm, respectively. The changes induced by AU 18 in these features are 31.13 mm, 32.40 mm, 33.78 mm, and 47.78 mm, respectively. These changes are all more than 50 times, or even 100 times, much greater than the expanded uncertainty, making them statistically significant.

4. CONCLUSIONS

In this work, we present a concept to analyse dynamic facial motion using multimodal 3D camera system and assess the metrological performance of the system. By assessing the system's measurement uncertainty under static condition and its sensitivity to facial motions, we demonstrate that the camera system is reliable for dynamic facial motion analysis, while also confirming the feasibility of our proposed concept.

REFERENCES

[1] Ekman, P., & Friesen, W. V., **Facial action coding system**, Environmental Psychology & Nonverbal Behavior, 1978.

[2] Zhi, R., Liu, M., & Zhang, D., **A comprehensive survey on automatic facial action unit analysis**, The Visual Computer, 36(5), 1067-1093, 2020.

[3] Jacob, G. M., & Stenger, B., **Facial action unit detection with transformers**, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021.

[4] Li, S., & Deng, W., **Deep facial expression recognition: A survey**, IEEE transactions on affective computing, 13(3), 1195-1215, 2020.

[5] Heist, S., Dietrich, P., …, & Tünnermann, A., **GOBO projection for 3D measurements at highest frame rates: a performance analysis**, Light: Science & Applications, 7(1), 71, 2018.

[6] Bazarevsky, V., Kartynnik, Y., …, & Grundmann, M., **Blazeface: Sub-millisecond neural face detection on mobile gpus**, arXiv preprint arXiv:1907.05047, 2019.

[7] ISO/IEC Guide 98–3: 2008-09, **Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement** (GUM 1995).