ABSTRAK

Keamanan sistem *pengenalan wajah* semakin rentan terhadap *presentation attack* seperti foto, video, dan topeng tiga dimensi, sehingga memunculkan kebutuhan mendesak terhadap mekanisme *face anti-spoofing (*FAS*)* yang andal. Pendekatan *unimodal* yang hanya memanfaatkan citra *RGB* terbukti kurang efektif dalam mendeteksi serangan canggih, khususnya pada kondisi pencahayaan yang bervariasi dan serangan berkualitas tinggi. Untuk mengatasi keterbatasan ini, penelitian ini mengusulkan metode FAS berbasis fusi *multimodal* dengan memanfaatkan citra RGB, inframerah (IR), dan *depth* yang dikombinasikan melalui mekanisme *bi-directional cross-attention*. Metode ini diharapkan mampu meningkatkan pemahaman kontekstual antar-modalitas sehingga sistem lebih *robust* terhadap variasi serangan dan kondisi lingkungan.

Penelitian ini dilaksanakan dengan pendekatan *eksperimental* dan *prototyping* yang mencakup tahapan studi literatur, akuisisi dan pra-pemrosesan data, perancangan arsitektur model, implementasi, pelatihan, dan evaluasi. *Dataset* yang digunakan meliputi CASIA-SURF yang merupakan *benchmark* publik untuk penelitian FAS *multimodal*. Setiap sampel diproses melalui tahap *cropping* wajah, *resizing*, normalisasi, serta *data augmentation* untuk meningkatkan keragaman data. Ekstraksi fitur dilakukan secara independen untuk masing-masing modalitas menggunakan *Convolutional Neural Network* (CNN) sebelum digabungkan melalui modul *bi-directional cross-attention* yang memungkinkan interaksi dua arah antar-modalitas. Proses pelatihan dilakukan secara *end-to-end* dengan optimisasi *AdamW*, *loss function* berbasis *binary cross-entropy*, serta mekanisme *early stopping* untuk mencegah *overfitting*.

Hasil pengujian menunjukkan bahwa metode yang diusulkan mampu mencapai performa yang kompetitif, dengan nilai APCER sebesar 0,2291%, BPCER sebesar 0,2186%, ACER sebesar 0,2239%, dan AUC mencapai 99,99%. Evaluasi dilakukan pada skenario *intra-dataset* untuk mengukur kemampuan generalisasi model, sementara analisis *ablation study* membuktikan efektivitas mekanisme *bi-directional cross-attention* dalam meningkatkan kinerja deteksi serangan. Kontribusi utama penelitian ini adalah pengembangan strategi fusi yang dapat memperkuat sistem FAS *multimodal* terhadap berbagai jenis serangan, sekaligus memberikan fondasi untuk integrasi ke aplikasi keamanan biometrik di dunia nyata.

Kata Kunci: *face anti-spoofing, multimodal fusion, bi-directional cross-attention, deep learning, biometric*

ABSTRACT

The security of face recognition systems is increasingly vulnerable to presentation attacks such as photos, videos, and three-dimensional masks, creating an urgent need for reliable face anti-spoofing (FAS) mechanisms. Unimodal approaches that rely solely on RGB images have proven to be less effective in detecting advanced attacks, particularly under varying illumination conditions and high-quality spoofing attempts. To address these limitations, this study proposes an FAS method based on multimodal fusion by leveraging RGB, infrared (IR), and depth images combined through a bi-directional cross-attention mechanism. This approach is expected to enhance contextual understanding across modalities, making the system more robust against attack variations and environmental conditions.

This research was conducted using an experimental and prototyping approach, which includes stages of literature review, data acquisition and preprocessing, model architecture design, implementation, training, and evaluation. The datasets used include CASIA-SURF which are publicly available benchmarks for multimodal FAS research. Each sample undergoes face cropping, resizing, normalization, and data augmentation to increase data diversity. Feature extraction is performed independently for each modality using Convolutional Neural Networks (CNN), followed by integration through a bi-directional cross-attention module that enables bidirectional interaction between modalities. The training process is conducted end-to-end using AdamW optimization, a binary cross-entropy loss function, and an early stopping mechanism to prevent overfitting.

The experimental results demonstrate that the proposed method achieves competitive performance, with APCER of 0.2291%, BPCER of 0.2186%, ACER of 0.2239%, and an AUC of 99.99%. Evaluations were carried out on intra-dataset scenario to measure the generalization capability of the model, while ablation study analysis confirms the effectiveness of the bi-directional cross-attention mechanism in enhancing attack detection performance. The main contribution of this research is the development of a fusion strategy that strengthens multimodal FAS systems against various attack types, providing a solid foundation for integration into real-world biometric security applications.

Keywords: *face anti-spoofing, multimodal fusion, bi-directional cross-attention, deep learning, biometrics*