Flash–Non-Flash Fusion for Contactless Fingerprint Recognition and Spoof Detection

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
in
Computer Science and Engineering
by Research

by

Roja Sahoo
2021111014
roja.sahoo@research.iiit.ac.in

Advisor: Dr. Anoop Namboodiri

International Institute of Information Technology Hyderabad
500 032, India

June 2026

International Institute of Information Technology Hyderabad
Hyderabad, India

CERTIFICATE

This is to certify that work presented in this thesis proposal titled ***Flash–Non-Flash Fusion for Contactless Fingerprint Recognition and Spoof Detection*** by *Roja Sahoo* has been carried out under my supervision and is not submitted elsewhere for a degree.

_____ _____

Date Advisor: Dr. Anoop Namboodiri

Acknowledgements

Firstly, I would like to thank my advisor, Dr. Anoop Namboodiri, for his support and guidance that were foundational for the completion of this thesis and for giving me an opportunity to work at the Centre for Visual Information Technology (CVIT). I would also like to thank Saketh and Sravanthi Ma'am for their support and for allowing me to access the resources at the Biometrics and Secure Identity Lab (BaSIL), which were crucial for this work. The experience of working at these centers and being able to work with various stakeholders has been greatly beneficial. I have always been fascinated by forensics, biometrics, and security. Therefore, being given an opportunity to delve deeper into this field has been inspiring.

I also wish to take a moment to highlight my experiences that defined a major part of my time at IIITH. Being part of Apex gave me the platform to mentor students and engage with parents, which I tried my best to do with sincerity and responsibility. Being part of The Dance Crew, Cultural Council, and Felicity Team, especially during IIIT's Silver Jubilee year, gave me great exposure and experience that helped me gain confidence on stage, learn to organize large-scale institutional events, and manage external collaborations. Each role was like a unique lesson in leadership, execution, and adversity, and I'm thankful for all the opportunities they afforded me. Balancing all these roles wasn't easy, something I both cherish and struggle with, but all the mishaps and missteps were learning opportunities, and I wouldn't be where I am today without them.

A special shout-out to my Prithvi house family for the bonds made across batches and the sense of community that developed from these connections. And to the smiling faces of the SLC, SAC, SLO, and Admissions teams – staff and professors alike – your approachability made all the more difference than you might have realized. The experiences and platforms this institution has given me have challenged me to always step outside my comfort zone, and I can only hope I used them well and that what I put into this campus leaves something worthwhile behind.

To all my close friends, seniors, and juniors, I just want to thank all of you for the amazing memories shared, for letting me vent, for sticking by me, and for showing up through it all. All of you have taught me something, and I will definitely take these lessons of friendship well beyond this college.

And finally, to my parents and brother, thank you for believing in me through all the highs and lows of my life, without any question. All this wouldn't have been possible without your love and support!

Abstract

Contactless fingerprint recognition provides a number of benefits in terms of hygiene and minimizing sensor-induced artifacts, over traditional contact-based systems, but it is also associated with a number of challenges that are currently hindering its adoption. The 2D representation of 3D features, in combination with unconstrained illumination, affects the quality of ridges, resulting in inconsistent features. Also, the lack of interaction information in contactless recognition makes it more prone to presentation attacks. Current techniques for acquisition, enhancement, and detection of spoofing attacks are mostly independent and focused on single images, without exploiting any complementary information for improved recognition. As the trend towards contactless biometric recognition is becoming increasingly necessary for usability and cost-effectiveness in terms of deployment in mobile and security-critical applications, lower matching accuracy and increased susceptibility to spoofing attacks are significant challenges. This highlights the need for a method that jointly addresses enhancement, recognition, and spoof detection while improving cross-modal interoperability.

In our work, we present a unified framework for contactless fingerprint recognition by exploiting the flash–non-flash acquisition pair as a lightweight active sensing mechanism that extracts complementary illumination information. We thus present the Flash–Non-Flash Fingerphoto (FNF) Database for a thorough analysis of the illumination impact, and Fusion2Print (F2P), an end-to-end deep learning framework that learns the optimal enhancement in the spatial, frequency, and color domains to enhance the quality of the ridge-valley representations and cross-domain recognition accuracy. Furthermore, we investigate a method for the detection of presentation attacks by exploiting the illumination-dependent photometric differences between genuine and spoofed fingerprints, and a low-cost 3D pipeline for the reconstruction of the fingerprint surface geometry from just two images, facilitating the simulation of digital spoof attacks.

Experimental results show that all components benefit from improvements. F2P obtains AUC of 0.999 and EER of 1.12%. The illumination-aware technique separates real and spoofing samples effectively, and the 3D reconstruction pipeline rebuilds finger geometry and ridge-valley information with improved contrast. These components constitute a refined unified framework for contactless fingerprint recognition.

Contents

List of Figures

List of Tables

List of Related Publications

[P1] Roja Sahoo and Anoop Namboodiri, **"Fusion2Print: Deep Flash-Non-Flash Fusion for Contactless Fingerprint Matching"**, in proceedings of *28th International Conference on Pattern Recognition (ICPR), Lyon, France*, 2026. Preprint at arXiv:2601.02318, 2026.

[P2] Roja Sahoo and Anoop Namboodiri, **"Illumination-Aware Contactless Fingerprint Spoof Detection via Paired Flash-Non-Flash Imaging"**, in proceedings of *14th IEEE International Workshop in Biometrics and Forensics (IWBF), Côte d'Azur, EURECOM*, 2026. Preprint at arXiv:2603.17679, 2026.

Chapter 1

Introduction

Biometric systems are significant for modern-day identity verification. Considering this, the utilization of fingerprint recognition is seen to be one of the most commonly used biometric systems, owing to the degree of distinctiveness and accuracy associated with it. Traditionally, fingerprint acquisition is done by employing contact-based sensors such as optical and capacitive sensors [1], which is known to offer high accuracy but is associated with several limitations, including hygiene issues, sensor wear, latent fingerprint residue, and user interaction.

1.1 Background and Motivation

Contactless fingerprint capture [2] has emerged as a promising alternative, enabling touch-free acquisition using standard cameras. This method improves usability, eliminates physical contact, and reduces risks associated with contamination and sensor degradation–factors that became particularly significant during the COVID-19 pandemic. Additionally, contactless systems support flexible capture conditions and scalable deployment across mobile and embedded platforms.

However, this shift introduces fundamental challenges. Fingerprint ridge structures are inherently 3D, while contactless systems capture only 2D projections, resulting in reduced ridge-valley clarity and degraded feature representation. Moreover, image quality is highly sensitive to illumination, pose, and background conditions, leading to variability that negatively impacts matching performance. Examples of such degradation are illustrated in Figure 1.1.

(a) High quality (b) Specular reflection (c) Blurry ridges

Figure 1.1 Examples of degradation in contactless fingerprint images.

Another critical challenge is vulnerability to presentation attacks, where adversaries attempt to deceive the system using printed images, molded replicas, or digital displays. Unlike contact-based systems, contactless acquisition lacks implicit liveness cues such as pressure deformation and capacitive responses, making spoof detection more challenging and dependent on visual information alone.

1.1.1 Key Observations and Technical Insights

One of the observations that has been the motivation behind the development of the presented work is the complementary property of illumination in the process of contactless fingerprint imaging. While the flash imaging is beneficial in terms of the enhancement of the high-frequency information of the ridges in the images, it is associated with the presence of specular reflections and low-frequency illumination artifacts. In the case of non-flash imaging, the intensity distribution is smooth with less noise but with less contrast in the ridges.

The paired flash–non-flash capture process is beneficial in terms of the acquisition of more information about the structure and the material of the surface. In the frequency domain, the combination of the two modalities can reduce the noise while maintaining the detail of the ridges.

Apart from frequency characteristics, another important aspect of fingerprint representation is color channel response. From empirical observations, it has been found that the response of the blue and green channels is stronger for containing information related to ridges, whereas the red channel is found to be dominated by subsurface skin and blood information [3]. This helps in adaptive channel-wise enhancement beyond conventional grayscale processing.

Recent advances in deep learning techniques [4] allow for illumination-invariant representations and embeddings that are robust for matching across different domains. In addition to 2D enhancement, illumination information can be used for surface geometry, reflectance information and subsurface scattering information. All these can be used to create a representation that can be used for approximating depth and surface characteristics, making it possible to achieve 3D reconstruction using just two images. Such representations can further enhance recognition and provide additional signals for spoof detection.

1.2 Research Challenges and Gaps

Despite these advancements, there are several critical issues that need to be addressed. These are listed below:

- **Limited datasets:** Lack of publicly available paired flash and non-flash fingerprint datasets under various conditions.
- **Illumination sensitivity:** Considerable sensitivity of ridges under varying illumination conditions.
- **Inadequate fusion strategies:** Dominance of single-image information in existing pipelines [5].

- **Color channel underutilization:** Inadequate utilization of RGB channel information in existing pipelines.
- **System incompatibility:** Commercial matchers are designed for contact-based inputs, and deep learning architectures lack interoperability.
- **Spoof detection limitations:** Current detection mechanisms are limited to static single-image features and lack generalization capability for various types of spoofs.
- **Absence of lightweight 3D cues:** Current 3D reconstruction techniques are either hardware-intensive or require multi-view configurations.

Moreover, practical issues such as alignment inconsistencies between paired images, sensitivity of segmentation to background variations, and difficulties in evaluating preprocessing using commercial matchers make this problem even more complicated.

These gaps highlight the need for a unified framework that leverages paired illumination for representation learning, enhances compatibility across systems, and introduces additional cues for both recognition and spoof detection.

1.3 Research Questions and Objectives

The primary research questions are:

1. *How can paired flash and non-flash images be effectively fused to enhance ridge-valley representation?*
2. *Can frequency-domain and color-aware processing, combined with deep learning, improve performance across traditional and learning-based systems?*
3. *Can illumination-induced differences be used as reliable signals for spoof detection?*
4. *Can meaningful 3D cues be extracted from only two images, and can they aid recognition and spoof modeling?*

The objective of this thesis is to develop a unified and practical framework for improving contactless fingerprint recognition and spoof detection by systematically leveraging illumination variation through paired flash–non-flash imaging. The goal is to transform illumination from a source of variability into a discriminative signal for enhancing fingerprint representation, improving matching robustness, and strengthening resistance to presentation attacks.

More specifically, this work investigates how complementary information from multiple illumination conditions can be utilized to enhance ridge structures, improve cross-domain matching, enable reliable spoof detection using photometric cues, and extract additional structural information, including approximate 3D characteristics, from minimal input data.

1.4 Key Contributions

This thesis introduces a unified perspective on contactless fingerprint recognition by treating flash–non-flash acquisition as a lightweight active sensing mechanism. The key contributions are as follows:

- **Flash–Non-Flash Fingerphoto (FNF) Database:** A paired dataset of flash and non-flash fingerprint images enabling a controlled study of illumination effects in contactless fingerprint imaging.
- **Fusion2Print Framework:** The first end-to-end pipeline integrating flash–non-flash fusion, adaptive enhancement, and discriminative embedding generation for robust fingerprint matching with cross-domain compatibility that improves performance across contact-based, contactless, and mixed-domain fingerprint matching scenarios.
- **Illumination-Aware Spoof Detection:** A novel perspective on spoof detection using paired acquisition leveraging illumination-induced material properties such as reflectance, texture realism, and subsurface scattering to distinguish genuine fingerprints from spoof artifacts.
- **Low-Cost 3D Reconstruction:** A lightweight approach for inferring surface geometry from two images captured under different illumination conditions, enabling approximate 3D modeling without specialized hardware.

Experimental findings demonstrate that flash–non-flash fusion improves ridge clarity and separation between mated and non-mated samples. The integration of learned enhancement and discriminative embedding generation further improves matching performance, while illumination-aware analysis strengthens robustness against presentation attacks.

1.5 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 reviews related work in contactless fingerprint recognition, enhancement, and spoof detection. Chapter 3 presents the acquisition pipeline, dataset construction, and preprocessing methods. Chapter 4 introduces the Fusion2Print (F2P) framework for fingerprint matching. Chapter 5 presents illumination-aware spoof detection. Chapter 6 explores low-cost 3D reconstruction and its applications. Chapter 7 concludes the thesis and outlines future research directions.

Chapter 2

Literature Review

Traditionally, fingerprint recognition is dominated by contact-based recognition systems that utilize capacitive or optical sensors to capture detailed information about ridges and valleys under controlled conditions [6, 7]. These systems deliver high accuracy due to the high quality of images and robustness of the overall pipeline. However, they are limited by hygiene issues, sensor wear, latent impressions, and deformation resulting from contact.

On the other hand, contactless fingerprint recognition has emerged as a more flexible solution that facilitates unconstrained recognition using cameras. Although this is accompanied by various issues, it has also led to tremendous research activities in various aspects of contactless fingerprint recognition, including acquisition, preprocessing, feature extraction, deep learning-based matching, spoof detection, and 3D reconstruction. The following sections review these developments in detail.

2.1 Contactless Fingerprint Recognition

Contactless fingerprint recognition systems are generally regarded as a new alternative to contact-based systems, addressing hygiene issues, user convenience, and reduced sensor-induced distortions [8–10]. An operational system was available in 2004 [11], spurring interest in solving problems related to pressure distortion, latent prints, and inter-device variation [12, 13]. Moreover, its relevance has been further increased through the activities of standardization and evaluation agencies such as NIST [14].

Contactless fingerprint recognition systems use high-resolution cameras to image fingerprint ridge-valley structures and skin texture [15, 16]. While this overcomes the problems related to contact fingerprint sensors, it presents challenges in terms of fingerprint segmentation from the background, reduction in the effective resolution of fingerprint images, pose and perspective variations, and sensitivity to illumination changes [17, 18]. All such variations during the image acquisition process degrade ridge contrast and introduce noise into the images, making it very challenging to extract features robustly.

2.1.1 Datasets and Benchmarks

Several datasets supporting contactless fingerprint research have been released to date, including ISPFDv2 [19], UNFIT [20], RidgeBase [21], and PolyU Contactless-to-Contact [22]. In spoof de-

tection, a few datasets have been proposed, such as COLFISPOOF [23], IIITD Spoofed Fingerphoto Database [24], and CLARKSON [25]. However, most datasets rely on single-modality acquisition and lack controlled illumination variation.

2.1.2 Preprocessing and Classical Feature Extraction

Classical contactless fingerprint recognition systems thus relied on preprocessing pipelines to increase ridge visibility and reduce the variability in acquisition conditions, including segmentation, normalization, and contrast enhancement [26–28].

Traditional feature extraction methods were central to these systems. Gabor filtering was commonly used for ridge enhancement since it is a frequency-selective filter [29] and Scale-Invariant Feature Transform (SIFT) descriptors are used because they are invariant to scale and rotation [30]. Beyond image-based descriptors, Topological Data Analysis (TDA) has also been explored for conventional contact fingerprint analysis, where persistent homology features extracted from minutiae point clouds and ridge patterns have been used for fingerprint classification and similarity assessment [31, 32], with limited application in contactless fingerprint recognition so far. In addition, techniques for core-point detection and spatial alignment have been proposed for pose normalization to improve matching performance [33, 34]. However, these methods remain sensitive to noise, illumination changes, and low ridge contrast, especially in unconstrained settings.

2.1.3 Mobile Fingerphoto Acquisition

With the mass adoption of smartphones, the idea of fingerphoto-based biometrics is becoming more relevant as a low-cost, widely available alternative to dedicated sensors. Systems such as [2, 35] have explored usability, acquisition protocols, and end-to-end mobile capture pipelines.

This unconstrained capture presents challenges because varying backgrounds, lighting, focus, and finger poses affect image quality and performance metrics [36, 37]. Although several solutions have been proposed to improve segmentation, scaling normalization, and illumination compensation, the resulting images are still far from the quality of controlled contact-based images, thus emphasizing the need to develop more robust and generalizable representations.

2.1.4 Illumination and Multi-Capture Fingerprint Imaging

Illumination plays a critical role in contactless fingerprint acquisition, affecting ridge-valley contrast and image quality. Flash-based captures enhance high-frequency ridge details but introduce specular reflections and noise, while non-flash images provide smoother intensity distributions with reduced ridge clarity. Existing approaches largely rely on single-modality inputs and mitigate illumination effects using preprocessing techniques such as histogram equalization and filtering [29], which are limited in recovering lost structural information.

Multispectral and illumination-aware approaches [38] show that varying lighting conditions reveal complementary material and structural cues, enabling improved recognition and spoof detection. However, paired multi-illumination datasets remain scarce, and combining complementary captures is still underexplored though with some preliminary work [39,40] in contactless palmprints using dual-instance (left–right) fusion via discrete cosine transform (DCT)-based feature extraction and discrimination power analysis (DPA) in a normalized joint feature space for improved recognition performance.

2.2 Contact-to-Contactless Interoperability

Interoperability with legacy contact-based databases is critical for practical deployment of contactless systems. Contact-to-contactless (C2CL) matching addresses this by bridging the modality gap between contact and contactless acquisitions.

Early work focused on geometric alignment and deformation modeling, with Robust Thin-Plate Splines (RTPS) [22] used to compensate for nonlinear distortions. Other approaches explored SIFT-based matching [28, 30], phase correlation [26], and scaling-invariant methods [36].

More recent pipelines [41] integrate preprocessing, enhancement, and matching to address low ridge-valley contrast and pose variation. Despite promising results, interoperability remains an open challenge, particularly under low-resolution and unconstrained conditions [15].

2.3 Deep Learning for Fingerprint Recognition

Deep learning has significantly advanced fingerprint recognition by enabling data-driven feature learning and robust representations. Specifically, convolutional neural networks (CNNs) have been successfully applied in classification, enhancement, and matching tasks and have outperformed all earlier approaches based on handcrafted features.

Vision Transformers (ViTs) are inherently more robust to global distortion and illumination variations [42], achieving state-of-the-art results on datasets such as ISPFD [37] and UNFIT [20]. At the same time, minutiae-aware frameworks incorporate domain knowledge, improving alignment and matching [43, 44].

Siamese and triplet architectures [45, 46] learn a similarity metric from pairs or triplets of inputs to solve verification problems. Graph-based models [47] have been proposed to encode spatial relationships between minutiae points, which increase the robustness of the representations with respect to geometric deformations. Hybrid approaches combining classical preprocessing with deep learning [27, 44] further enhance performance. All these representations, however, suffer from the lack of large-scale contactless fingerprint datasets with sufficient diversity [4].

2.4 Presentation Attack Detection

Presentation attack detection (PAD) is essential for securing contactless fingerprint systems, especially in mobile and unconstrained environments. Early methods relied on handcrafted features such as second-order Gaussian derivatives [48], texture descriptors, and color-space analysis to distinguish real samples from spoof artifacts.

With deep learning, CNN-based approaches have become dominant. Multi-channel fusion methods [49] exploit complementary color spaces for improved robustness, while semi-supervised Generative Adversarial Network(GAN)-based approaches [50] demonstrate strong generalization across attack types.

Unsupervised methods using autoencoders and attention mechanisms [51] show promise for detecting unseen spoofs. Domain adaptation frameworks such as GRU-AUNet [52] further enhance cross-dataset generalization. Hardware-assisted systems like RaspiReader [53] combine imaging modalities but remain vulnerable to advanced spoofing artifacts.

2.5 3D Reconstruction and Fingerprint Modeling

3D fingerprint reconstruction addresses limitations of 2D contactless imaging, particularly for handling pose variations and recovering ridge geometry. Early work inferred 3D minutiae from single 2D images [54], demonstrating the utility of depth cues for recognition. Geometric and multi-view approaches [43] reconstruct structure using feature correspondences across captures. Structured light illumination (SLI) methods leverage SIFT, ridge patterns, and minutiae to estimate 3D point clouds, often incorporating prior finger shape models [55], but require controlled setups and multiple views. Photometric stereo-based methods provide efficient alternatives. For example, [56] uses a single-camera system with multi-colored LED illumination to recover 3D minutiae via a tetrahedron-based matching scheme, combining 2D and 3D features. Learning-based approaches address scalability and data limitations. Print2Volume [57] generates synthetic Optical Coherence Tomography(OCT)-like 3D volumes from 2D images using style transfer, 3D expansion, and GAN refinement, enabling large-scale pretraining. Similarly, deep models [54, 58] learn volumetric representations for robust feature extraction. Monocular reconstruction and view synthesis methods infer 3D structure from a single image using learned priors [59], reducing hardware requirements but remaining limited by depth ambiguity.

Despite these advances, most methods rely on specialized hardware, multiple captures, or large-scale 3D data, limiting deployment and motivating low-cost reconstruction under minimal capture constraints.

2.6 Summary

In spite of this progress, contactless systems remain sensitive to illumination, presentation attacks, depend on single-modality processing, and interoperability with contact-based systems is an issue. Current 3D reconstruction techniques also have limitations related to hardware and multiple captures. These

limitations have given rise to multi-capture techniques that take advantage of complementary information. Specifically, paired flash–non-flash images capture both high-frequency ridges and illumination-invariant structure. This is used to create a unified framework for enhancement, recognition, spoof detection, and cost-effective 3D reconstruction under unconstrained capture.

Chapter 3

Flash–Non-Flash Fingerprint Database and Differential Preprocessing

In fingerprint recognition, distinctive features are derived based on the ridges present in the fingerprint image. A fingerprint consists of ridges (raised regions) and valleys (recessed regions), whose clarity directly affects recognition performance. The most critical local features are minutiae–ridge endings and bifurcations–corresponding to ridge termination and splitting [60]. In addition, global structures such as the core (central region) and delta (points of ridge divergence) provide important spatial and orientation context for matching.

In contrast to traditional contact-based systems that operate under controlled acquisition conditions with consistent ridge impressions, contactless imaging poses problems such as illumination variation, perspective distortion, motion blur, and and background noise, making feature extraction more difficult.

The clarity of these extracted features also depends to a great extent on the quality of the image. Noise, illumination changes, and other image artifacts, especially in the context of contactless or flash/non-flash image capture, can make ridge-valley patterns less distinct or even lead to false minutiae detection. Hence, in the context of image preprocessing in the present work, the first aim is to make the ridge patterns clearer and reduce noise levels to ensure the accurate detection of both minutiae and core-delta structures. Earlier research [61] has indicated better fingerprint recognition performance with increase in the clarity of such features.

3.1 Paired Flash–Non-Flash Capture Dataset

To enable illumination-aware fingerprint processing, we construct a paired flash–non-flash dataset, where complementary information from both captures can be leveraged for noise suppression and enhanced ridge clarity.

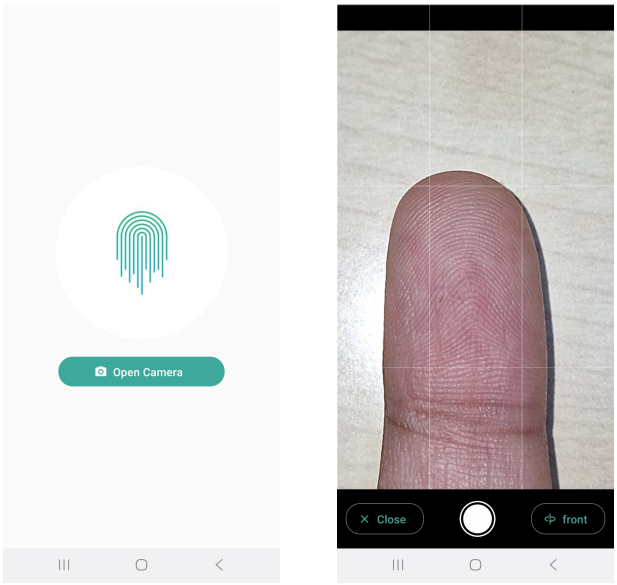
3.1.1 Data Acquisition Setup and Protocol

We developed a custom cross-platform smartphone application[1] using the open-source Expo Camera framework [62], implemented in React Native, to enable controlled data acquisition. The application

[1] Camera App Source Code

allows users to capture paired fingerprint images–one with flash and one without–while providing features such as camera switching, direct image storage, and a grid overlay for improved alignment, as shown in Figure 3.1.

To minimize inter-frame displacement, the application captures a flash-illuminated image followed by a non-flash image after a fixed delay of 600 ms, avoiding the latency associated with manual flash toggling. All images were acquired in macro mode with a fixed zoom factor of 0.45 to ensure consistent scale and ridge visibility. Data collection was performed using a Samsung A54 smartphone under unconstrained conditions.



(a) App home page (b) Camera interface

Figure 3.1 Screenshots from the contactless fingerprint capture application used with a Samsung A54 device.

3.1.2 Dataset Overview and Structure

Using this setup, we introduce the Flash–Non-Flash Fingerphoto (FNF) Database[2], a two-session contactless fingerprint dataset consisting of paired flash and non-flash images. The captured dataset contains 3,140 valid fingerprint images from 79 individuals, with 20 impressions per subject across sessions, capturing natural intra-subject variability.

The final released dataset consists of 12,560 processed images and is organized by session and illumination type as follows:

```
FNFDatabase/
|-- Session1/
```

[2]FNF Database

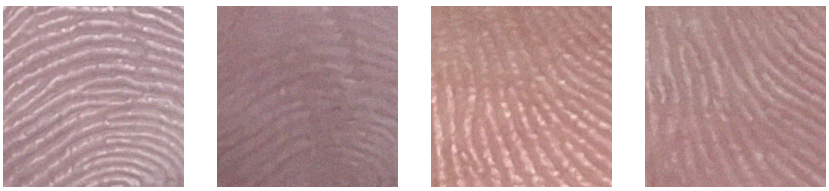
```
|   |-- Flash/
|   '-- Non-Flash/
'-- Session2/
    |-- Flash/
    '-- Non-Flash/
```

All images are provided in `.tif` format with a resolution of 256×256 pixels, RGB color space, and a spatial resolution of 500 dpi. The average file size is approximately 197 KB, with a total dataset size of 2.52 GB.

3.1.3 Privacy-Preserving Release and Usage Policy

Due to the sensitive nature of biometric data, only a processed and privacy-preserving version of the dataset is publicly released. The release includes cropped fingerprint patches, noise-added, non-reversible transformations, and augmented variants designed to prevent identity reconstruction. Both flash and non-flash processed images from multiple sessions are included, with representative samples shown in Figure 3.2.

The Flash–Non-Flash Fingerphoto (FNF) Database is released strictly for non-commercial academic research under institutional ethical approval by the IRB (Institute Review Board) and in compliance with established NIST biometric data handling guidelines [63]. Access is granted only upon signing a restricted license agreement[3], which prohibits redistribution, commercial use, or any attempt at identity reconstruction or re-identification. Approved users are provided with a secure, time-limited download link. Researchers are required to appropriately cite the dataset in any resulting publications.



(a) ID 1: Flash (b) ID 1: Non-Flash (c) ID 2: Flash (d) ID 2: Non-Flash

Figure 3.2 Example flash and non-flash processed fingerprint pairs from two subjects.

3.2 Flash–Non-Flash Fingerprint Analysis

This section provides a comprehensive analysis of flash (I_{flash}) and non-flash (I) fingerprint images across frequency and color domains, followed by a detailed preprocessing pipeline and extensive evaluation using both classical and deep learning-based matchers.

[3]FNF Database License Agreement

3.2.1 Frequency Domain Analysis

There is a significant difference in the spectral behavior between flash and non-flash images. The flash images have high intensity ridge features, while non-flash images have smooth intensity variations with less clear ridge features.

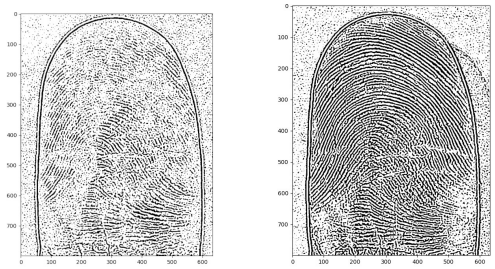
To analyze this, images are transformed into the frequency domain using:

$$F(u, v) = \mathcal{F}\{I(x, y)\} \tag{3.1}$$

The radial spectral energy profile is computed as $R(f) = \text{RadialAvg}(|F(u, v)|)$.

In addition to spectral analysis, we also analyze the ridge-enhanced images of flash and non-flash images as shown in Figure 3.3. From the images, it is clear that the flash images have more continuous ridges with sharper transitions than the non-flash images. This is because the flash images have higher-frequency components and well-defined ridge-valley structures. On the other hand, the non-flash images have broken ridges and blurred structures with more low-frequency components and less sharp ridge structures.

The ridge-enhanced representations further confirm that flash images preserve fine-grained texture details, while non-flash images suffer from loss of high-frequency information and dominance of smoother intensity variations. This distinction motivates the use of non-flash images for estimating low-frequency illumination, while leveraging flash images for high-frequency ridge information.



(a) Non-Flash (b) Flash

Figure 3.3 Comparison of ridge-enhanced non-flash and flash images.

3.2.2 Color Domain Analysis

To quantify ridge visibility across RGB channels, local contrast is computed using block-wise standard deviation as $C_{\text{local}} = \frac{1}{N} \sum_{k=1}^{N} \text{std}(B_k)$, where B_k denotes the k^{th} non-overlapping image block (patch) of size $p \times p$, and N is the total number of blocks covering the image.

We analyze the channel-wise contribution to ridge clarity for both I_{flash} and I images, as illustrated in Figure 3.4.

Empirical observations across the dataset show:

$$\text{Blue} > \text{Green} > \text{Red} \tag{3.2}$$

for both flash and non-flash images. Additionally, flash images consistently exhibit higher overall local contrast compared to non-flash images:

$$C_{\text{local}}(I_{\text{flash}}) > C_{\text{local}}(I) \tag{3.3}$$

Blue and Green channels provide higher ridge contrast, capturing finer ridge-valley structures, while the Red channel contributes less to ridge information and is more influenced by skin tone, blood content, and subsurface scattering [3]. Flash images exhibit stronger ridge contrast across all channels due to improved illumination, whereas non-flash images show reduced contrast and weaker ridge definition, particularly in higher frequency regions. These results imply that channel-aware processing targeting blue and green channels will further improve the visibility of ridges and representation of fingerprints.

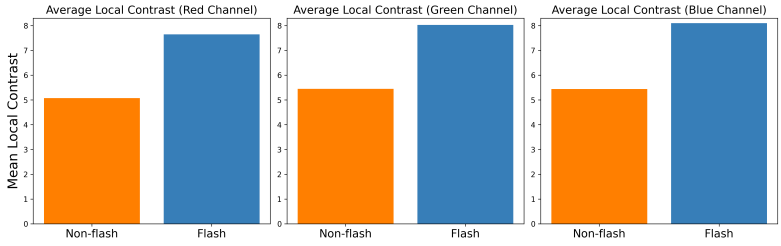


Figure 3.4 Channel-wise RGB local contrast of flash and non-flash images.

3.3 Spectral-Domain Preprocessing for Noise Attenuation

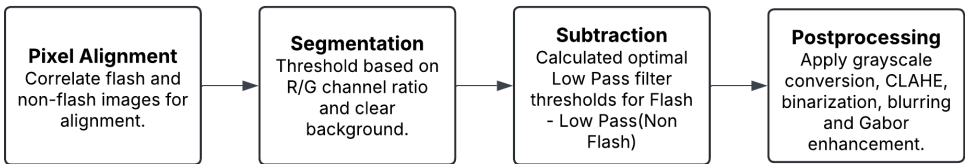
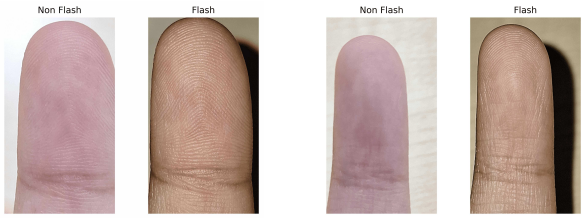


Figure 3.5 Preprocessing pipeline for flash and non-flash fingerprint images.

We present a step-wise preprocessing pipeline, summarized in Figure 3.5, that leverages paired flash and non-flash images to suppress illumination artifacts while preserving high-frequency ridge structures. The process combines spatial alignment, frequency-domain filtering, and differential enhancement.

3.3.1 Original Captured Images

We begin with paired I_{flash} and I fingerprint images captured consecutively to minimize displacement, as shown in Figure 3.6.

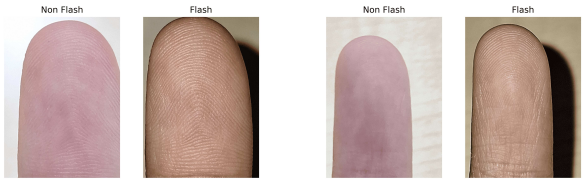


(a) Example 1 (b) Example 2

Figure 3.6 Captured flash and non-flash fingerprint image pairs.

3.3.2 ROI Selection

The images are cropped to a fixed region of interest (ROI) focusing on the upper fingerprint region where ridge structures are most stable, as seen in Figure 3.7.



(a) Example 1 (b) Example 2

Figure 3.7 Selected region of the top part of the finger used for identification.

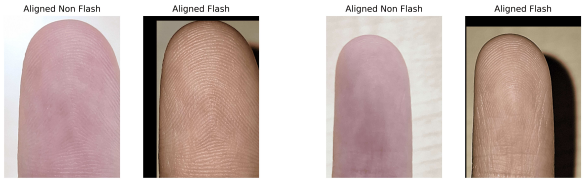
3.3.3 Pixel Alignment

To compensate for residual displacement, phase correlation is used to estimate translation between paired images:

$$(\Delta x, \Delta y) = \arg\max \ \mathcal{F}^{-1}\left\{\frac{F_1(u,v)\cdot F_2^*(u,v)}{|F_1(u,v)\cdot F_2^*(u,v)|}\right\}, \tag{3.4}$$

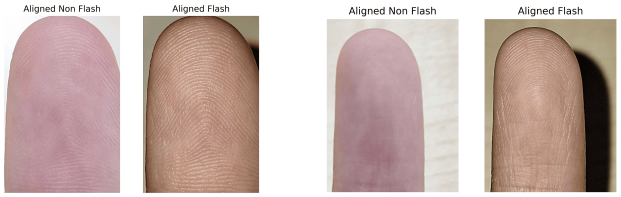
where F_1 and F_2 are the Fourier transforms of the I_{flash} and I images.

Initial attempts using edge-based alignment relied on Canny edge detection and boundary points; however, this approach failed under partial fingertip visibility. The flash image is then shifted using affine transformation, and black boundary regions are cropped to ensure spatial consistency (Figures 3.8 and 3.9).



(a) Example 1 (b) Example 2

Figure 3.8 Pixel-wise aligned flash and non-flash pairs.



(a) Example 1 (b) Example 2

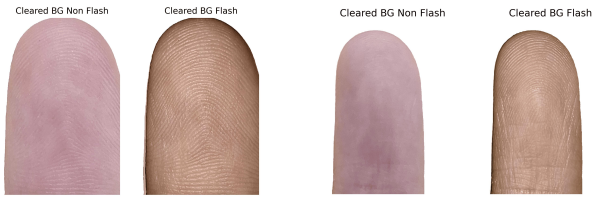
Figure 3.9 Aligned images with cropped boundary artifacts.

3.3.4 Finger Segmentation and Background Removal

Segmentation is performed to isolate the fingerprint region. While adaptive thresholding was initially explored, it proved unstable under varying illumination. Therefore, an R/G ratio-based method is used:

$$\frac{R}{G} > T \tag{3.5}$$

This exploits the chromatic property of skin where red intensity dominates green. Pixels outside the segmented region are set to white to suppress background interference, with results shown in Figure 3.10.



(a) Example 1 (b) Example 2

Figure 3.10 Background removal using segmentation mask.

3.3.5 Low-Frequency Component Extraction

The non-flash image is used to estimate illumination noise. It is transformed into the frequency domain as $F(u, v) = \mathcal{F}\{I(x, y)\}$. The spectral energy profile is computed as $R(f) = \text{RadialAvg}(|F(u, v)|)$. The cutoff frequency is defined as $f_c = \min\{f \mid R(f) < 0.5 \cdot \max(R(f))\}$.

Unlike previously used methods that employed predefined low-pass filter characteristics that are often inferior in terms of generalization and sometimes produce poor color artifacts in subtraction, this formulation allows for an optimal cutoff frequency to be determined dynamically based on the spectrum of the image. This allows for a more effective filtering that takes into consideration illumination characteristics unique to each image, resulting in improved subtraction quality and more consistent ridge preservation across varying capture conditions.

A low-pass filter is constructed:

$$H(u, v) = \begin{cases} 1, & \sqrt{u^2 + v^2} \leq min(f_c, f_N), [f_N = \text{Nyquist frequency}] \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

The filtered image is reconstructed as:

$$I_{\text{low}}(x, y) = \mathcal{F}^{-1}\{F(u, v) \cdot H(u, v)\} \tag{3.7}$$

This retains illumination and smooth variations while removing ridge-related high-frequency components.

3.3.6 Differential Enhancement via Subtraction

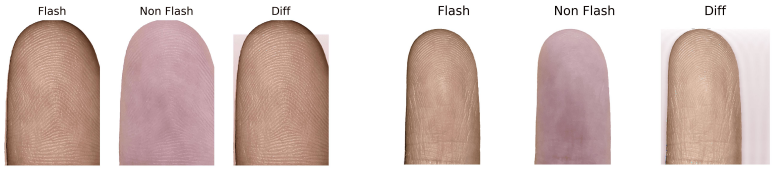
The difference image is obtained by subtracting the low-frequency non-flash image(I_{low}) from the aligned flash image(I_{flash}):

$$I^c_{\text{diff}}(x, y) = I^c_{\text{flash}}(x, y) - I^c_{\text{low}}(x, y), \quad c \in \{R, G, B\} \tag{3.8}$$

This subtraction is performed independently for each color channel (Red, Green, and Blue), capturing channel-specific differences in ridge and illumination characteristics. The final difference image, I_{diff}, is then formed by combining the resulting channel-wise difference maps.

$$I_{\text{diff}}(x, y) = \left[I^R_{\text{diff}}, I^G_{\text{diff}}, I^B_{\text{diff}}\right] \tag{3.9}$$

This operation removes low-frequency illumination noise while preserving high-frequency ridge structures, with results shown in Figure 3.11.



(a) Example 1 (b) Example 2

Figure 3.11 Flash, non-flash and resulting difference images.

3.3.6.1 Comparison with Alternative Subtraction Strategies

To evaluate the effectiveness of the proposed method, multiple input variations were considered, including standalone images (I, I_{flash}), direct subtraction ($I_{\text{flash}} - I$), symmetric low-pass subtraction ($I_{\text{flash}} - \mathcal{L}(I_{\text{flash}})$) alobng with the proposed method ($I_{\text{flash}} - \mathcal{L}(I)$), where $\mathcal{L}(\cdot)$ denotes the low-pass filtering operation.

Empirical analysis shows that standalone non-flash images (I) and direct subtraction methods ($I_{\text{flash}} - I$, $I_{\text{flash}} - \mathcal{L}(I_{\text{flash}})$) introduce distortions in ridge structures, adversely affecting minutiae detection and feature reliability. Although $I_{\text{flash}} - \mathcal{L}(I_{\text{flash}})$ produces a higher number of detected features due to amplification effects, this does not translate to improved matching accuracy. In contrast, the original flash image I_{flash} already provides strong ridge information, yielding performance comparable to the proposed method $I_{\text{flash}} - \mathcal{L}(I)$.

These observations indicate that naïve or symmetric subtraction strategies may introduce artifacts or redundant features without improving discriminative power. The proposed approach, which leverages non-flash images specifically for low-frequency estimation via $\mathcal{L}(I)$, provides a more controlled enhancement, effectively suppressing illumination noise while preserving meaningful ridge details.

3.3.7 Post-Processing

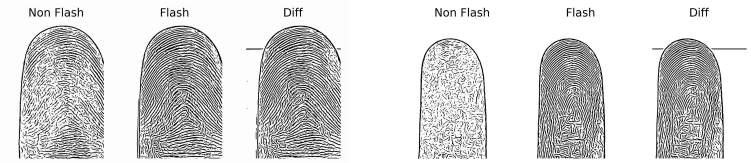
Finally, I_{diff} is refined through a sequence of post-processing steps to improve ridge clarity and structural continuity. The enhanced image is first converted to grayscale to simplify intensity representation, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) for local contrast enhancement [64]. Adaptive binarization is then applied to separate ridge and valley regions, after which Gaussian blurring is used to reduce residual noise and smooth spurious variations.

To further reinforce ridge structures, Gabor filtering is employed [65]. This step enhances ridge flow by selectively responding to specific orientations and spatial frequencies that align with fingerprint ridge patterns. The filtering operation is defined as:

$$I_{\text{gabor}}(x, y) = I_{\text{gray}}(x, y) * g_{\theta,\lambda,\sigma,\gamma}(x, y) \tag{3.10}$$

where θ controls the orientation, λ the wavelength corresponding to ridge spacing, σ the scale of the filter, and γ its aspect ratio. This operation strengthens coherent ridge structures while suppressing residual noise and discontinuities.

The processed outputs are shown in Figure 3.12, showing improvements in ridge clarity across I, I_{flash}, and I_{diff} images. We also performed comparisons against grayscale-only and Gabor-enhanced settings for the same fingerprint image to evaluate the contribution of each processing stage.



(a) Example 1 (b) Example 2

Figure 3.12 Final processed outputs for I, I_{flash}, and I_{diff} images.

3.4 Experimental Results and Analysis

We evaluate the visual quality and feature extraction capability of I, I_{flash}, and the proposed I_{diff} images in grayscale and gabor-enhanced settings for qualitative analysis, minutiae detection, and matching performance.

3.4.1 Qualitative Assessment and Minutiae Analysis

I images exhibit poor ridge contrast due to insufficient illumination, resulting in weak ridge-valley separation and loss of fine details. In contrast, I_{flash} images provide stronger ridge visibility and capture high-frequency ridge structures with improved continuity, but are often affected by specular highlights and illumination noise variations. The proposed preprocessing to get I_{diff} images suppresses these low-frequency artifacts while preserving high-frequency ridge information, leading to cleaner and more consistent ridge patterns, as shown in Figure 3.13.

(a) I (b) I_{flash} (c) I_{diff}

Figure 3.13 Gabor enhanced images of I, I_{flash} and I_{diff}, demonstrating quality of ridges captured.

These observations are consistent with earlier channel-wise analysis, where I_{flash} exhibits stronger high-frequency content across RGB channels compared to I. The subtraction-based formulation in I_{diff} retains these high-frequency ridge components while attenuating low-frequency illumination variations, improving perceptual ridge clarity.

From a minutiae detection perspective, I_{flash} and I_{diff} yield comparable feature counts, indicating that I_{flash} already contains strong ridge information. While I_{diff} may reveal additional structures due to improved contrast and reduced illumination artifacts, these do not consistently improve matching performance. This is largely due to the robustness of commercial matchers, which apply internal preprocessing, limiting gains from enhanced ridge clarity. Moreover, slight misalignment during subtraction and amplification of residual noise can reduce the benefits of enhancement. Failure cases highlight that direct subtraction ($I_{\text{flash}} - I$) introduces artifacts, while binarized subtraction is sensitive to misalignment, producing distorted ridges. Segmentation inconsistencies can further introduce background interference. These findings show that minutiae count or visual quality alone is insufficient, and that matching performance is a more reliable measure of improvement.

3.4.2 Grayscale vs Gabor Enhancement

The impact of the post-processing step is also studied by comparing the grayscale and Gabor filter-based images. Grayscale images capture overall intensity variations but are not effective in capturing the structural information since the orientation of the ridges has not been considered in the images. Gabor filtering improves the quality of the ridges significantly by acting as an orientation-selective band-pass filter that enhances the quality of the ridges by reinforcing ridge flow along dominant directions and filtering out the noise in the images, resulting in a more continuous set of ridges and valleys, as shown in Figure 3.14. This improves the quality of the images, resulting in $I_{\text{gabor}} > I_{\text{gray}}$, making Gabor-enhanced images more suitable for reliable minutiae extraction and matching.

(a) Grayscale (b) Gabor

Figure 3.14 Comparison of grayscale and Gabor-enhanced images, showing improved ridge continuity and orientation preservation with Gabor filtering.

3.5 Baseline Matching Performance

Fingerprint verification typically follows two paradigms: *minutiae-based* and *pattern/texture-based* approaches. Minutiae, defined as ridge endings and bifurcations, are highly discriminative features

whose reliable extraction depends on ridge clarity and continuity. In contrast, texture-based methods capture global ridge flow and appearance patterns, offering robustness to local distortions.

To evaluate the proposed preprocessing pipeline, we perform fingerprint verification by enrolling images from Session 1 and matching them with Session 2. Two complementary matchers are used: a commercial minutiae-based system, *VeriFinger* [66], and a deep learning-based embedding model, *DeepPrint* [67], representing explicit feature-based and learned representation-based matching paradigms, respectively.

Evaluation Metrics

Verification performance is assessed using similarity score distributions for mated and non-mated pairs, where mated pairs correspond to the same finger across sessions and non-mated pairs to different fingers. To address class imbalance due to the large number of non-mated comparisons, a subset of non-mated scores is randomly sampled.

- **ROC Curve and AUC:** The Receiver Operating Characteristic (ROC) curve plots the True Accept Rate (TAR) against the False Accept Rate (FAR), capturing the trade-off between correctly accepting matches and rejecting non-matches. The Area Under the Curve (AUC) summarizes separability between mated and non-mated distributions, with higher values indicating better discrimination.
- **FAR vs FRR and EER:** The False Accept Rate (FAR) measures the proportion of non-mated pairs incorrectly accepted, while the False Reject Rate (FRR) measures the proportion of mated pairs incorrectly rejected. The Equal Error Rate (EER), defined where FAR equals FRR, provides a concise measure of system accuracy, with lower values indicating better performance.

3.5.1 VeriFinger (Minutiae-Based Matching)

VeriFinger is a commercial fingerprint recognition system that extracts minutiae maps and performs matching based on corresponding minutiae between two fingerprints. Typical mated matches contain 20-40 consistent correspondences, and the system is designed to operate at very low false accept rates (e.g., FAR@0.01%). While it incorporates deep learning for auxiliary tasks such as classification and spoof detection, its core verification pipeline remains minutiae-based.

Table 3.1 Verification performance using VeriFinger SDK.

Image	**Gray**		**Gabor**	
	AUC	EER(%)	AUC	EER(%)
I	0.64	34.87	0.69	30.90
I_{flash}	0.80	20.45	0.81	19.11
I_{diff}	**0.82**	**18.13**	**0.86**	**14.23**

The results in Table 3.1 show that the processed I_{diff} consistently outperforms I and I_{flash} inputs. This improvement stems from suppressing low-frequency illumination artifacts while preserving high-frequency ridge structures essential for accurate minutiae extraction.

Additionally, Gabor-enhanced representations outperform grayscale across all inputs, indicating that orientation-selective filtering improves ridge continuity and supports more reliable minutiae detection.

3.5.2 DeepPrint (Embedding-Based Matching)

DeepPrint is a deep learning-based fingerprint recognition framework that learns representations using a dual-branch architecture: a *texture branch* capturing global ridge patterns and appearance, and a *minutiae branch* encoding spatial and geometric information of ridge endings and bifurcations. In our experiments, we use a pre-trained DeepPrint model [68], trained on approximately 8,000 fingerprint images, to generate 512-dimensional embeddings (256 for minutiae and 256 for texture). Similarity between two fingerprints is computed using cosine similarity:

$$S(f_1, f_2) = \frac{f_1 \cdot f_2}{\|f_1\| \|f_2\|} \tag{3.11}$$

Table 3.2 Verification performance using DeepPrint embeddings.

Embedding	**Image**	**Gray**		**Gabor**	
		AUC	EER(%)	AUC	EER(%)
minu	I	0.55	50.41	0.60	45.07
	I_{flash}	0.58	48.91	0.66	37.71
	I_{diff}	0.63	39.87	0.77	30.68
tex	I	0.59	46.70	0.65	40.88
	I_{flash}	0.61	43.14	0.69	36.42
	I_{diff}	0.65	38.09	0.82	23.21
minutex	I	0.60	44.11	0.70	36.38
	I_{flash}	0.63	42.65	0.74	33.83
	I_{diff}	**0.69**	**35.79**	**0.86**	**21.53**

In addition to verifying that our processing enhancement resulting in I_{diff} is helpful, through Table 3.2 we observe evaluation across three representation types; minutiae-only (*minu*), texture-only (*tex*), and combined embeddings (*minutex*). The performance consistently follows *minutex* $>$ *tex* $>$ *minu*, indicating that texture features are more robust to illumination variations while their combination with minutiae provides complementary information for improved verification with the DeepPrint model.

3.5.3 Performance Trend

Across both matchers and representations, a consistent trend is observed:

$$I_{\text{diff}} > I_{\text{flash}} > I \tag{3.12}$$

indicating that spectral-domain subtraction effectively enhances ridge visibility by removing illumination noise while preserving discriminative ridge patterns. The improvement is more pronounced in DeepPrint than in VeriFinger, as the latter already incorporates internal preprocessing and is relatively sturdy to image quality variations. In contrast, DeepPrint benefits more directly from improved input representations. Although I_{diff} images show clearer ridge continuity and stronger high-frequency detail, these visual enhancements don't always translate proportionally to matching performance due to factors like spurious minutiae, residual misalignment, and matcher sensitivity. This highlights that preprocessing should ultimately be judged based on end-to-end verification performance rather than intermediate visual or feature-level improvements.

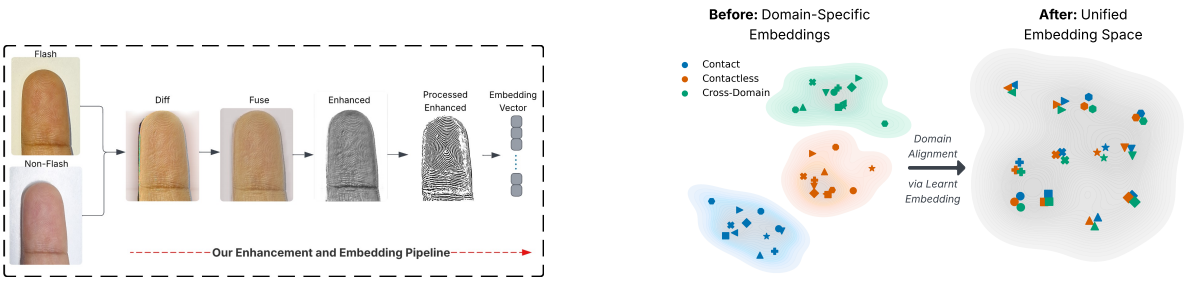
3.6 Conclusion

This chapter describes an illumination-aware framework for the preprocessing of contactless fingerprint images based on paired flash and non-flash images. A specific pipeline was designed to develop the **Flash–Non-Flash Fingerphoto (FNF) Database** to facilitate the capture of images with complementary illumination conditions while maintaining variability. The spectral and spatial differences between flash and non-flash images are utilized to enhance the quality of ridge structures in the images by separating high-frequency structures from low-frequency illumination components. Further improvements are made through orientation-aware filtering and learning-based representations.

Although the quality improvements, such as better ridge continuity and contrast, are visible, the impact on verification performance also depends on the matcher used. For example, classical minutiae-based matchers do not experience significant improvements due to their inherent robustness and training domain bias, whereas learning-based matchers experience a more direct benefit from quality improvements in the input images. Some of the takeaways include the efficacy of frequency domain subtraction, the dominance of the blue and green channels for ridge information, the benefit of Gabor filtering, and the refinement obtained by the use of deep learning, especially in learning-based matchers in comparison to commercial matchers. Additionally, the use of a combination of deep learning-based matchers with embedding analysis, such as clustering or visualization of the distributions, may offer a more precise assessment of the improvements beyond traditional verification metrics which are inherently contact-based.

Chapter 4

Fusion2Print: Deep Flash–Non-Flash Fusion for Contactless Fingerprint Matching



(a) Flash–non-flash enhancement-embedding pipeline (Fusion2Print)

(b) Embedding space before and after domain alignment

Figure 4.1 Overview of proposed fingerprint acquisition and enhancement framework: (a) The pipeline progressively enhances ridge-valley structure to produce an identity-discriminative embedding vector; (b) Embedding fine-tuning aligns domain-specific representations into a unified embedding space for contact and contactless fingerprints.

From the analysis of the flash and non-flash fingerprint images, it is clear that there is complementary information present across illumination conditions. In our first work, we took advantage of this fact by using the approach of manual subtraction between the flash(I_{flash}) and non-flash(I) images, in order to obtain an intermediate representation, I_{diff}. The results showed improved ridge-valley contrast, visibility of minutiae, and clarity of structure. Nevertheless, the approach we took was based on heuristics and the use of certain parameters, which might not be universally applicable. In addition, although commercial systems such as Verifinger and deep learning-based models like DeepPrint perform well for contact-based images, they might not perform well for contactless images due to variations in domain, illumination, perspective, and skin deformation. Therefore, there is a need for a deep learning approach, which can learn the optimal method for enhancing the images in the spatial, frequency, and color domains, ultimately producing a robust and discriminative fingerprint representation.

4.1 Introduction

Thus to address these challenges, we propose a unified learning-based enhancement–embedding framework, illustrated in Figure 4.1, that jointly handles contact, contactless, and mixed-domain fingerprint matching. The proposed approach also incorporates spatial transformations to align core fingerprint regions, thereby improving robustness to variations in pose and acquisition conditions.

Building on our previously collected paired flash–non-flash dataset, FNF Database, we introduce Fusion2Print (F2P), an end-to-end deep learning pipeline designed to learn illumination-aware enhancement and generate domain-invariant embeddings. The framework implicitly captures complementary information across frequency and color domains, replacing manual preprocessing with learned feature fusion. Additionally, we propose a unified embedding generator that aligns representations from different capture modalities into a common feature space, enabling consistent and reliable matching. Experimental results demonstrate that the proposed approach significantly improves recognition performance, particularly in cross-domain scenarios, compared to traditional single-capture and minutiae-based baselines.

4.2 Proposed Method

This section presents our full contactless fingerprint enhancement and embedding pipeline shown in Figure 4.2.

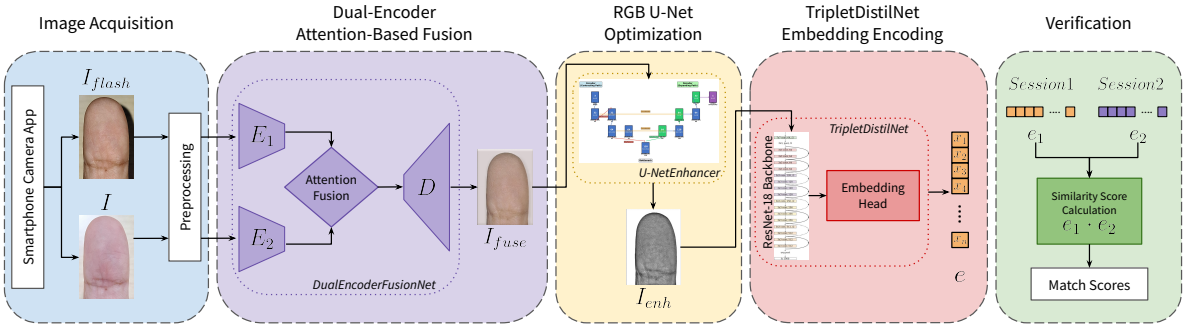


Figure 4.2 Overview of the proposed Fusion2Print framework. Aligned flash (I_{flash}) and non-flash (I) images are processed by a dual-encoder attention fusion network, where E_1 and E_2 denote flash and non-flash encoders and D produces the fused image I_{fuse}. A U-Net–based enhancer refines I_{fuse} to obtain I_{enh}, which is encoded by a ResNet-18-based feature extractor to produce discriminative embeddings. Verification is performed by comparing embeddings e_1 and e_2 from two capture sessions.

4.2.1 Spatial Transformation for Core Region Alignment

Along with the preprocessing and alignment described in Section 3.3, a spatial transformation is performed on both I_{flash} and I fingerprint images to enforce consistent geometric alignment. This step

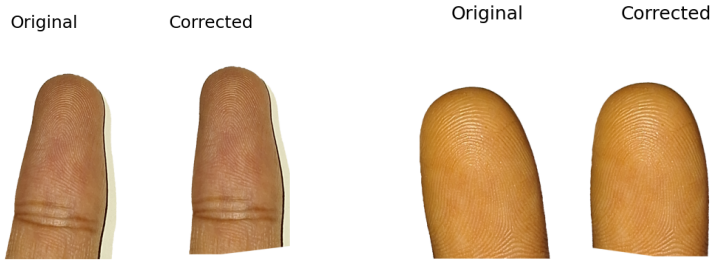
reduces variations arising from pose, rotation, and translation, enabling the downstream network to focus on identity-specific features rather than nuisance factors. To achieve geometric consistency across capture sessions, all input images are first upright-aligned by estimating a global rotation angle from the slopes of the left and right fingerprint boundaries, given by

$$\theta = \tfrac{1}{2}\big(\arctan(m_{\text{left}}) + \arctan(m_{\text{right}})\big), \tag{4.1}$$

where m_{left} and m_{right} denote the slopes of the respective boundary edges. Each image is then rotated by $-\theta$ to obtain a canonical upright orientation.

Following orientation correction, the fingerprint core point (x_c, y_c) is detected from the orientation field using coherence and the Poincaré index [34], where coherence is computed as $C_{ij} = \frac{\sqrt{V_x^2+V_y^2}}{\sum(G_x^2+G_y^2)+\epsilon}$ from aggregated orientation vectors (V_x, V_y) and image gradients (G_x, G_y) with a small stabilizing constant ϵ, and locations with a Poincaré index $P_{ij} \approx +0.5$ correspond to core points. The image is subsequently translated such that the detected core aligns with the image center, resulting in a stable and consistent ROI. This spatial normalization ensures that corresponding anatomical structures are consistently positioned across samples, allowing the network to learn identity-discriminative patterns rather than being affected by misalignment.

Example results of upright rotation and core-region centering are shown in Figure 4.3. This normalization step ensures that both I_{flash} and I inputs are geometrically aligned prior to enhancement and embedding generation.



(a) Example 1 (b) Example 2

Figure 4.3 Example of spatial alignment (upright rotation and core-region centering) applied to few sample flash contactless fingerprints.

4.2.2 Dual-Encoder Attention-Based Fusion Network

In order to fully utilize the complementary information from the paired I_{flash} and I images, we propose *DualEncoderFusionNet*, a unified enhancement network capable of learning to synthesize a refined and discriminative fused representation. Unlike flash-only processing methods or subtraction-based methods, this method learns modality-specific cues and adaptively combines them for noise reduction and ridge-valley feature enhancement.

4.2.2.1 Network Architecture

The network takes a six-channel input formed by concatenating RGB flash and non-flash images and produces a three-channel fused output I_{fuse} and consists of the following key components:

Dual Encoders. Two parallel convolutional encoders extract hierarchical features from I_{flash} and I, capturing illumination-specific and structural information. At each layer, feature extraction is defined as:

$$F_x = \text{ReLU}(\text{BN}(\text{Conv}(x))). \tag{4.2}$$

Attention-Based Feature Fusion. The extracted features are combined using an pixel-wise soft attention-based fusion mechanism that assigns spatially adaptive weights:

$$F_{\text{fuse}} = W_1 \odot F_{\text{flash}} + W_2 \odot F, \quad \text{where } W_1 + W_2 = 1, \tag{4.3}$$

allowing the network to dynamically emphasize the more informative modality at each spatial location.

Frequency and Edge-Aware Decoder. The fused representation is passed through a decoder D that reconstructs the enhanced fingerprint image. To explicitly improve ridge clarity, high-frequency components are selectively amplified in the Fourier domain:

$$I' = D(F_{\text{fuse}}) \cdot \Big(1 + \lambda \cdot \frac{\log(1 + |\mathcal{F}\{D(F_{\text{fuse}})\}|)}{\max(|\mathcal{F}\{D(F_{\text{fuse}})\}|) + \epsilon}\Big), \tag{4.4}$$

where $\mathcal{F}$ denotes the Fourier transform and λ controls the degree of amplification. Edge refinement is further applied using a residual convolution, $I_{\text{fuse}} = I' + \beta \cdot \text{Conv}_{\text{edge}}(I')$.

A segmentation mask M (see Section 3.3.4 for mask generation) is used to suppress background regions and restrict learning to fingerprint foreground areas.

4.2.2.2 Training Strategy

In the initial experiment, a simple encoder-decoder architecture, which was only trained using the $L1$ and $SSIM$ loss, produced blurred images with poor discriminative details. The use of the frequency-based difference image I_{diff} as a supervisory signal provided a solid starting point. Nevertheless, the aim is not just to mimic this, but to outperform it with sharper and more informative fused outputs.

All images were padded and resized to 512×512 to eliminate geometric distortions. The network was optimized using the Adam optimizer with a learning rate of 10^{-4}, weight decay of 10^{-5}, and ran for 30 epochs with a 70-15-15 train–val–test split. A learning rate scheduler was used for stable convergence.

4.2.2.3 Loss Function Design

To guide the network toward producing structurally accurate and perceptually sharper outputs, we employ a composite loss integrating spatial, structural, frequency, edge, and perceptual constraints. The

objective is not only to reconstruct the reference image I_{diff}, but to surpass it in terms of ridge clarity, texture richness, and discriminative quality.

L1 Loss (Spatial Fidelity). The $L1$ loss provides a stable optimization signal by enforcing pixel-wise similarity between the fused output and the reference image:

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i} \left| I_{\text{fuse},i} - I_{\text{diff},i} \right| . \tag{4.5}$$

This term preserves global intensity consistency and acts as a coarse alignment constraint.

SSIM Loss (Structural Consistency). To ensure that the overall ridge flow and structural layout are preserved, we incorporate the Structural Similarity Index:

$$\mathcal{L}_{SSIM} = 1 - \text{SSIM}(I_{\text{fuse}}, I_{\text{diff}}). \tag{4.6}$$

This discourages structural distortions and maintains the integrity of fingerprint patterns.

Fourier Loss (High-Frequency Enhancement). Fingerprint discriminability is strongly dependent on high-frequency ridge details. To explicitly promote sharper textures, we define a frequency-domain hinge loss:

$$\mathcal{L}_{Fourier} = \max\left(0,\, E_{hf}(I_{\text{diff}}) - E_{hf}(I_{\text{fuse}})\right), \tag{4.7}$$

where $E_{hf}(\cdot)$ denotes high-frequency energy. This formulation penalizes the model only when the fused output has lower-frequency content than the reference, encouraging the network to generate sharper ridge-valley transitions.

Edge Loss (Ridge Sharpness). To further enhance ridge boundaries, we enforce consistency in image gradients:

$$\mathcal{L}_{Edge} = \|\nabla_x(I_{\text{fuse}}) - \nabla_x(I_{\text{diff}})\|_1 + \|\nabla_y(I_{\text{fuse}}) - \nabla_y(I_{\text{diff}})\|_1. \tag{4.8}$$

By aligning horizontal and vertical gradients, this loss promotes clearer and more pronounced ridge edges.

Perceptual Loss (Feature-Level Consistency). Low-level similarity alone is insufficient for capturing discriminative fingerprint characteristics. Therefore, we incorporate a perceptual loss based on a pretrained VGG16 network [69]:

$$\mathcal{L}_{Perceptual} = \|\phi(I_{\text{fuse}}) - \phi(I_{\text{diff}})\|_1, \tag{4.9}$$

where $\phi(\cdot)$ represents feature extraction. This term enforces similarity in a higher-level feature space, implicitly encouraging richer and more meaningful representations.

Total Loss. The final objective combines all components as:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{L1} + \beta\mathcal{L}_{SSIM} + \gamma\mathcal{L}_{Fourier} + \delta\mathcal{L}_{Edge} + \epsilon\mathcal{L}_{Perceptual}, \tag{4.10}$$

where $\alpha, \beta, \gamma, \delta, \epsilon$ are empirical weighting coefficients(heuristically determined) controlling the contribution of each term with $\alpha = 0.6$, $\beta = 0.1$, $\gamma = 0.15$, $\delta = 0.1$, and $\epsilon = 0.05$.

This composite formulation not only reconstructs the guidance signal but actively drives the network to produce I_{fuse} that is sharper, more structured, and more discriminative for downstream recognition tasks.

Figure 4.4 shows the training and validation loss curves, while Figure 4.5 presents qualitative results including inputs, target, and fused output.

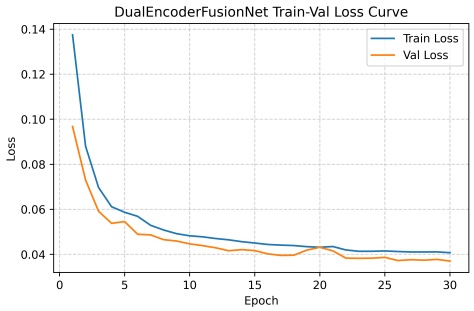


Figure 4.4 Train vs validation loss for *DualEncoderFusionNet* with custom loss.

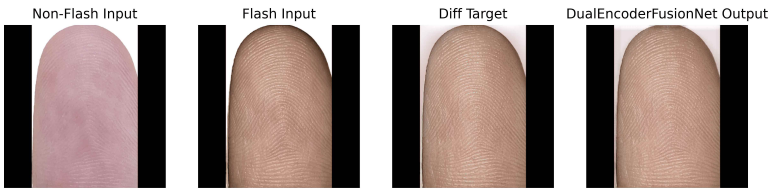


Figure 4.5 Padded visual inputs and outputs. **Inputs:** RGB I and I_{flash} images, **Target:** RGB difference image, I_{diff}. **Output:** RGB fused frequency- and edge-aware attention-based reconstruction, I_{fuse}.

4.2.3 Color-Space Ridge Enhancement

To further enhance ridge clarity and suppress illumination and skin reflectance artifacts, we design a color-space enhancement pipeline that computes an ideally weighted combination of the RGB channels with a learning-based U-Net refinement model.

Channel-Wise RGB Analysis. We analyze the contribution of individual RGB channels to ridge visibility and observe that the blue and green channels consistently exhibit stronger ridge-valley contrast compared to the red channel, computed as described in Section 3.2.2.

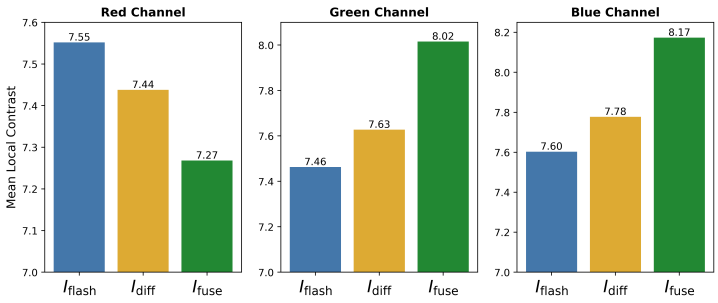


Figure 4.6 Channel-wise RGB local contrast of I_{flash}, I_{diff}, and I_{fuse}.

As shown in Fig. 4.6, ridge information is more prominent in the blue and green channels across I_{flash}, I_{diff}, and I_{fuse} representations. This motivates assigning higher importance to these channels during enhancement, while reducing the influence of the red channel to suppress illumination and subsurface scattering artifacts.

4.2.3.1 U-Net Based Enhancement Model

Based on these observations, we design *U-NetEnhancer* ($\mathcal{U}_\theta$), which maps the fused RGB image $I_{\text{fuse}} \in \mathbb{R}^{H\times W\times 3}$ to a weighted grayscale ridge-enhanced output $I_{\text{enh}} \in [0,1]^{H\times W}$:

$$I_{\text{enh}} = \mathcal{U}_\theta(I_{\text{fuse}}). \tag{4.11}$$

The network follows a standard encoder–decoder U-Net architecture [70], where each block consists of two Conv–BN–ReLU layers with skip connections to preserve fine spatial details. A sigmoid activation at the output ensures normalized intensity predictions.

Unsupervised variants were observed to produce overly smooth outputs; therefore, we adopt supervised training using aligned grayscale I_{flash} images as targets, along with additional ridge-aware constraints. All inputs are padded and resized to 512×512, and a binary mask M is used to restrict learning to valid fingerprint regions.

We further explored multiple configurations, including supervised vs. unsupervised training, U-Net variants (small, medium, ResU-Net [71]), different loss functions, training schedules, varying target representations(grayscale vs gabor enhanced images) and robustness to noise augmentations. The final model is trained using the Adam optimizer ($lr = 10^{-4}$) for 30 epochs with a 70–15–15 train-val-test split.

4.2.3.2 Custom Loss Formulation for Enhancement

The objective of the enhancement network is not merely to reproduce the grayscale I_{flash} image, but to surpass it by generating sharper ridges, improved local contrast, and better ridge-valley separation.

Masked L1 Loss (Spatial Fidelity).

$$\mathcal{L}_{L1} = \frac{\sum |I_{\text{enh}} - I_{\text{flash}}| M}{\sum M + \epsilon} \tag{4.12}$$

This term enforces pixel-wise reconstruction within valid regions and stabilizes training.

Local Contrast Reward. To explicitly enhance ridge visibility, we define a contrast reward based on patch-wise standard deviation, where $\sigma_p(I) = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(I_i - \mu_p)^2}$.

$$\mathcal{R}_{\text{contrast}} = \frac{1}{N}\sum_{p=1}^{N} \sigma_p(I_{\text{enh}})\, w_p, \tag{4.13}$$

where K is the number of pixels in a patch, μ_p is the patch mean, and w_p are adaptive weights derived from RGB analysis ($w_B > w_G > w_R$). Since this term is subtracted, higher contrast directly reduces the loss, encouraging sharper ridge structures.

SSIM Loss (Structural Preservation).

$$\mathcal{L}_{SSIM} = 1 - \text{SSIM}(I_{\text{enh}}, I_{\text{flash}}) \tag{4.14}$$

This preserves global ridge flow and prevents structural distortions.

Gabor Loss (Ridge Consistency).

$$\mathcal{L}_{Gabor} = \frac{\sum |I_{\text{enh}} - G_{\text{ref}}|\, W_{\text{ridge/valley}}\, M}{\sum M + \epsilon}, \tag{4.15}$$

where $W_{\text{ridge/valley}}(x) = \begin{cases} w_{\text{ridge}}, & G_{\text{ref}}(x) < 0.5, \\ w_{\text{valley}}, & \text{otherwise}, \end{cases}$ G_{ref} is a Gabor-filtered ridge map, and $w_{\text{ridge}} > w_{\text{valley}}$. This enforces stronger emphasis on ridge regions and promotes cleaner ridge-valley separation.

Edge Loss (Sharpness Constraint). The Sobel gradient is defined as $\nabla_{\text{Sobel}} I = (\nabla_x I, \nabla_y I)$.

$$\mathcal{L}_{edge} = \left\| \nabla_{\text{Sobel}} I_{\text{enh}} - \nabla_{\text{Sobel}} I_{\text{flash}} \right\|_1, \tag{4.16}$$

which promotes sharper ridge boundaries by aligning gradient responses.

Total Loss. The final loss function is defined as:

$$\mathcal{L}_{\text{final}} = \alpha \mathcal{L}_{L1} - \beta \mathcal{R}_{\text{contrast}} + \gamma \mathcal{L}_{SSIM} + \delta \mathcal{L}_{Gabor} + \eta \mathcal{L}_{edge}, \tag{4.17}$$

where $\alpha, \beta, \gamma, \delta, \eta$ are empirical weighting coefficients(heuristically determined) with $\alpha = 0.4$, $\beta = 0.2$, $\gamma = 0.3$, $\delta = 0.1$, and $\eta = 0.1$.

Together, these loss components guide the network to not only reconstruct the baseline signal but also enhance ridge clarity, contrast, and discriminability beyond the original grayscale I_{flash} image. As shown

in Figures 4.7 and 4.8, the model converges stably and produces visually improved ridge structures with higher local contrast. The final enhanced output I_{enh} can be further refined using Gabor filtering (as described Section 3.3.7) to improve ridge continuity and matching performance.

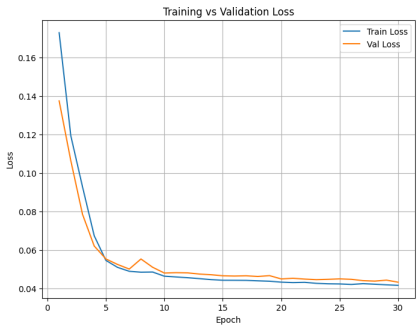


Figure 4.7 Train vs validation Loss for *U-NetEnhancer* with custom loss.

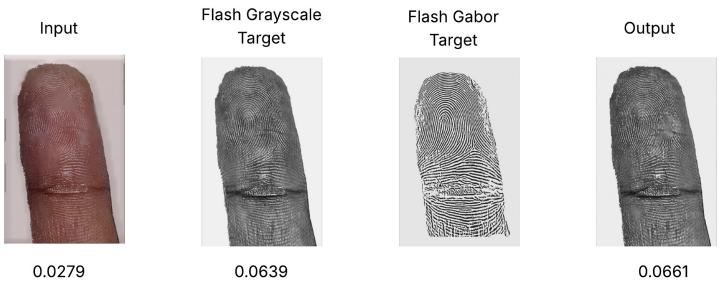


Figure 4.8 Visual outputs and corresponding local contrast scores *(shown below images)*. **Input:** noisy fused image, I_{fuse}, **Targets:** grayscale flash image (baseline), I_{flash}, and gabor flash image, G_{ref}, for gabor based ridge loss, **Output:** weighted grayscale reconstruction, I_{enh}.

4.2.4 Deep Embedding Learning for Verification

To enable robust fingerprint verification across contactless, contact-based, and mixed-domain scenarios, we propose a unified embedding framework, *TripletDistilNet*, which combines metric learning with knowledge distillation. This model learns discriminative representations from enhanced fingerprints, I_{enh}, while maintaining consistency with a strong contact-based teacher model, DeepPrint.

4.2.4.1 Architecture and Embedding Formulation

The proposed *TripletDistilNet* builds upon a pretrained ResNet-18 [72] backbone, followed by a lightweight embedding head that projects features into a compact latent space. The network $f_\theta(.)$ maps an input fingerprint image x to an ℓ_2-normalized embedding:

$$\mathbf{e} = \frac{f_\theta(x)}{\|f_\theta(x)\|_2} \in \mathbb{R}^{128}\,(256{\times}256),\ \mathbb{R}^{256}\,(512{\times}512). \tag{4.18}$$

Normalization ensures that similarity comparisons done using cosine similarity are based purely on angular relationships, which is particularly effective for biometric verification. Training is performed on I_{enh} fingerprint images with geometric transformations (rotation, scaling, slight perspective distortions) and Gaussian noise augmentation to improve invariance to real-world capture variations.

Triplet Dataset. Each identity provides paired samples $(I_i^{(1)}, I_i^{(2)})$. Triplets (I_a, I_p, I_n) are constructed such that I_a and I_p share identity, while I_n belongs to a different identity. To ensure meaningful gradients during optimization, semi-hard mining is employed: $d(\mathbf{e}_a, \mathbf{e}_p) < d(\mathbf{e}_a, \mathbf{e}_n) < d(\mathbf{e}_a, \mathbf{e}_p) + m(= 0.3)$. This avoids trivial or overly difficult triplets, leading to stable convergence and improved embedding separation.

Verification Mated/non-mated pairs are classified using a threshold on cosine similarity between the final embeddings, with identity-wise performance reported.

4.2.4.2 Training and Loss Formulation

The embedding network is optimized using a combination of triplet loss and knowledge distillation, enabling both discriminative learning and structural alignment with a strong teacher model.

Triplet Loss.

$$\mathcal{L}_{\text{triplet}} = \max\big(0,\ d(\mathbf{e}_a, \mathbf{e}_p) - d(\mathbf{e}_a, \mathbf{e}_n) + m\big), \tag{4.19}$$

which enforces intra-class compactness and inter-class separation by introducing a margin between positive and negative pairs.

Distillation Loss. Teacher embeddings $(\mathbf{t}_a, \mathbf{t}_p, \mathbf{t}_n)$ from DeepPrint are aligned with student embeddings:

$$\mathcal{L}_{\text{distill}} = 1 - \tfrac{1}{3}[\cos(\mathbf{e}_a, \mathbf{t}_a) + \cos(\mathbf{e}_p, \mathbf{t}_p) + \cos(\mathbf{e}_n, \mathbf{t}_n)]\,. \tag{4.20}$$

This transfers high-level discriminative priors and preserves global embedding structure.

Total Loss.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{triplet}} + \alpha\,\mathcal{L}_{\text{distill}}, \quad \alpha = 0.7. \tag{4.21}$$

Training is performed using the AdamW optimizer ($lr = 10^{-4}$, batch size = 16) with per-epoch triplet re-sampling and learning rate scheduling to avoid convergence plateaus.

4.2.4.3 Architecture Variants and Training Strategies

Prior to arriving at the final architecture, a number of different embedding learning strategies were explored to understand the impact of different objectives and feature representations.

Siamese Network with Contrastive Loss. A baseline Siamese CNN model with contrastive loss was able to attain a moderate level of performance ($\sim$0.93 AUC). However, the contrastive formulation tends to compress intra-class variance excessively, limiting its ability to model fine-grained ridge variations critical for fingerprint recognition.

LBP Feature Augmentation. Incorporating Local Binary Pattern (LBP) features helped gain invariance to illumination changes but could not represent global ridge flow and subtle structural variations. This made the separability very poor for contactless images, which typically have large variabilities.

Triplet Learning Variants. Replacing the contrastive loss with the triplet loss yielded better embedding discrimination ($\sim$0.97 AUC). Among different mining strategies, semi-hard mining provided the best trade-off by avoiding noisy gradients from hard negatives while still enforcing a meaningful margin.

Scaling Triplets and Data Diversity. Scaling the number of triplets per identity and including several variants of preprocessing, such as enhancement and fusion of the inputs, allowed the network to generalize better across different illumination, pose, and capture conditions. This points to the critical importance of data diversity in metric learning and balanced sampling.

Knowledge Distillation. Introducing distillation loss from DeepPrint into the embedding space further regularized the learning of embedding and stabilized the training by aligning it with a well-structured contact-domain representation. This is very useful for generalizing to the cross-domain scenario, especially when the triplet sampling is sparse or noisy.

Architecture Ablation Study and Embedding Analysis

The results in Table 4.1 demonstrate that combining triplet learning, semi-hard mining, and knowledge distillation yields the most discriminative embedding space, with improved inter-class separation and stable intra-class compactness.

Table 4.1 Comparison of verification modules trained and tested on I_{enh} using 128-D embeddings. Lower Intra, higher Inter, and higher SepRatio indicate better separability.

Model	**Euclidean**			**Cosine**		
	$\text{Intra}_{\downarrow}$	$\text{Inter}_{\uparrow}$	$\text{SepRatio}_{\uparrow}$	$\text{Intra}_{\downarrow}$	$\text{Inter}_{\uparrow}$	$\text{SepRatio}_{\uparrow}$
DeepPrint	**0.24**	0.84	3.44	**0.03**	0.36	10.46
TripletNet	0.52	1.31	2.52	0.15	0.99	6.39
TripletDistilNet-NoMining	0.40	1.35	3.44	0.09	0.99	11.37
TripletDistilNet	0.38	**1.35**	**3.53**	0.08	**0.99**	**11.92**

The baseline model *TripletNet* uses margin-based separation and outperforms other traditional approaches. However, it also has a larger intra-class variation due to the lack of additional regularization. The addition of knowledge distillation in *TripletDistilNet-NoMining* reduces the intra-class variation and aligns representations with a stronger teacher model, but still has a relatively low inter-class discrimination without effective triplet selection. The combination of both distillation and semi-hard mining in the proposed *TripletDistilNet* model strikes the best balance by tightening the mated clusters and separating non-mated samples. This results in the highest separation ratios being obtained by the proposed model under both Euclidean and cosine metrics. Also, increasing the dimensionality of the embedding space

from 128 to 256 dimensions results in improved performance, indicating a more expressive and robust embedding space for cross-domain fingerprint verification.

4.2.5 End-to-End Enhancement–Embedding Pipeline

Building on the frequency-aware fusion, color-space enhancement, and deep embedding modules, we propose a unified framework by integrating the above components into Fusion2Print (F2P). Our complete framework processes the paired flash and non-flash fingerprint images to obtain robust and discriminative embeddings for verification across varying capture conditions.

4.2.5.1 Architecture Overview

The proposed F2P pipeline composes three pretrained modules in sequence:

- *DualEncoderFusionNet*: performs frequency- and edge-aware fusion of flash and non-flash inputs to enhance ridge continuity while suppressing illumination artifacts,
- *U-NetEnhancer*: refines the fused output using channel-aware contrast enhancement, emphasizing ridge-rich blue and green channels,
- *TripletDistilNet*: generates discriminative embeddings using metric learning and knowledge distillation.

Given paired inputs I and I_{flash} along with a fingerprint mask M, the pipeline produces an ℓ_2-normalized embedding $\mathbf{e}$ through sequential transformations:

$$I, I_{\text{flash}} \xrightarrow{\mathcal{F}_{\text{fusion}}} I_{\text{fuse}} \xrightarrow{\mathcal{U}_{\text{enh}}} I_{\text{enh}} \xrightarrow{\mathcal{E}_{\text{triplet}}} \mathbf{e} \in \mathbb{R}^d,\ d \in \{128, 256\}. \tag{4.22}$$

The full mapping can also be expressed as: $\mathbf{e} = \mathcal{E}_{\text{triplet}}(\mathcal{U}_{\text{enh}}(\mathcal{F}_{\text{fusion}}(I, I_{\text{flash}}, M)))$.

This formulation enables end-to-end propagation of enhancement cues into the embedding space, ensuring that ridge-level improvements directly contribute to identity discrimination.

4.2.5.2 End-to-End Fine-Tuning

To further improve the cross-session consistency and quality of embeddings, we perform end-to-end fine-tuning of the entire F2P pipeline. The enhancement modules (*DualEncoderFusionNet* and *U-NetEnhancer*) are partially frozen to preserve learned enhancement priors, while the embedding network *TripletDistilNet* is optimized using paired flash–non-flash samples (80/20 train-test split).

Training is performed using the Adam optimizer ($lr = 10^{-6}$, weight decay = 10^{-7}) for 10 epochs, with gradient clipping to stabilize updates. Unlike earlier stages that rely on explicit reconstruction targets, this stage directly optimizes embedding quality.

4.2.5.3 Fine-Tuning Loss Formulation

To guarantee discriminative separation and intra-class compactness simultaneously, a two-term objective that combines soft-margin triplet learning and cosine consistency is adopted.

Soft-Margin Triplet Separation. For a batch of embeddings, each anchor–positive pair is compared against non-matching samples. The loss encourages the anchor–positive distance to be smaller than the anchor–negative distance by a margin δ:

$$\mathcal{L}_{\text{trip}} = \log\big(1 + \exp(d_{ap} - d_{an} + \delta)\big), \tag{4.23}$$

where d_{ap} and d_{an} denote anchor–positive and anchor–negative distances, respectively, and $\delta = 0.08$. This soft-margin formulation avoids unstable hard-negative mining while still enforcing meaningful separation.

Positive-Pair Cosine Consistency. To enforce tight clustering of embeddings belonging to the same identity, we minimize their cosine distance:

$$\mathcal{L}_{\text{intra}} = 1 - \cos(\mathbf{e}_a, \mathbf{e}_p). \tag{4.24}$$

This term ensures identity coherence and reduces intra-class variance.

Final Objective. The overall fine-tuning loss is defined as a weighted combination:

$$\mathcal{L}_{\text{fine}} = w_t \mathcal{L}_{\text{trip}} + w_i \mathcal{L}_{\text{intra}}, \tag{4.25}$$

where $(w_t, w_i) = (0.75, 0.25)$ control the trade-off between inter-class separation and intra-class compactness. This joint optimization ensures that the final embeddings are not only well-separated across identities but also tightly clustered within each identity, resulting in improved verification performance across contactless fingerprint scenarios.

4.3 Results and Analysis

Since no public paired flash–non-flash fingerprint dataset exists, we evaluate the proposed F2P pipeline on our curated dataset. We compare performance across multiple input configurations: non-flash (I), flash (I_{flash}), manually computed difference (I_{diff}), fused output (I_{fuse}), enhanced output (I_{enh}), and spatially transformed enhanced images ($I_{\text{enh\_ST}}$).

4.3.1 Ridge Structure Quality Analysis

We first evaluate ridge quality using NFIQ2 [73] along with three complementary metrics, as NFIQ2 is not specifically designed for contactless fingerprints: *Local Contrast*, which captures ridge-valley separation through patch-wise intensity variation; *Sharpness*, which measures high-frequency ridge de-

tail using Laplacian variance; and *Edge Clarity*, which quantifies ridge boundary visibility using Canny edge detection.

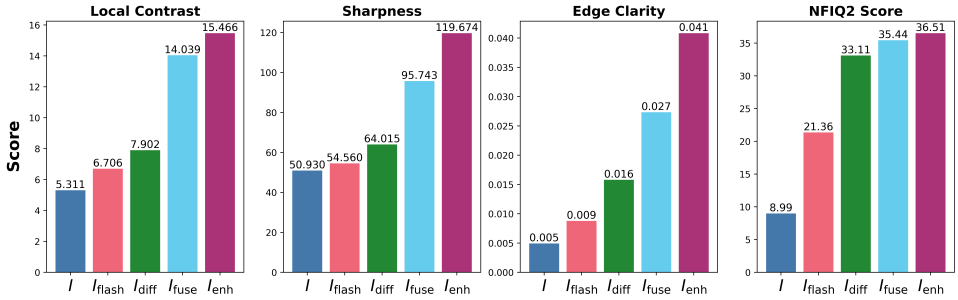


Figure 4.9 Comparison of ridge-quality scores across different input types.

As shown in Fig. 4.9, the proposed enhancement pipeline consistently improves ridge quality across all metrics. The final enhanced output I_{enh} outperforms all intermediate representations, demonstrating that learned fusion and enhancement effectively recover sharper and more discriminative ridge structures compared to both raw inputs and manual subtraction.

4.3.2 Baseline Matcher-Based Verification Results

To evaluate the impact of each processing stage on recognition performance, we perform verification using two strong baseline matchers: VeriFinger (minutiae-based) and DeepPrint (deep minutiae- and texture-based embeddings). Results across different input types are summarized in Table 4.2.

Table 4.2 Verification performance of DeepPrint and VeriFinger across our dataset variations.

Matcher	**Metric**	I	I_{flash}	I_{diff}	I_{fuse}	I_{enh}	$I_{\textbf{enh\_ST}}$
VeriFinger	AUC	0.69	0.81	0.86	0.90	0.93	**0.94**
	EER (%)	30.90	19.11	14.23	11.30	8.05	**7.60**
DeepPrint	AUC	0.64	0.74	0.87	0.88	0.94	**0.97**
	EER (%)	37.45	33.06	21.67	19.75	12.68	**8.92**

The results show a consistent improvement in verification performance as we move from raw inputs to enhanced representations. While manual subtraction (I_{diff}) provides a strong baseline, the learned fusion (I_{fuse}) and enhancement (I_{enh}) further improve both AUC and EER. Applying spatial transformation ($I_{\text{enh\_ST}}$) yields the best performance, highlighting the importance of geometric alignment for robust matching.

4.3.3 Ablation Study on TripletDistilNet

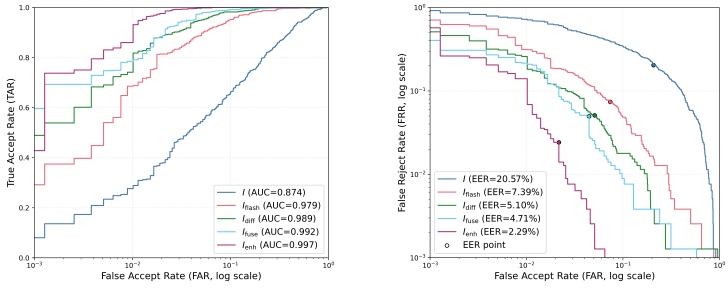
We evaluate the verification performance of *TripletDistilNet* across all enhancement variants, namely I, I_{flash}, I_{diff}, I_{fuse}, and I_{enh}, as summarized in Tab. 4.3. All models use 128-dimensional embeddings to

ensure computational efficiency. This study analyzes how training on a specific representation generalizes across other input domains, thereby capturing cross-modality robustness.

Table 4.3 Verification performance across different train–test configurations for the *TripletDistilNet* embedding module.

Train ↓ / Test →	I		I_{flash}		I_{diff}		I_{fuse}		I_{enh}	
	AUC	EER (%)	AUC	EER (%)	AUC	EER (%)	AUC	EER (%)	AUC	EER (%)
I	**0.91**	**16.31**	0.80	26.31	0.82	24.71	0.86	21.53	0.91	15.97
I_{flash}	0.87	20.64	0.969	9.49	0.973	7.77	0.989	4.71	0.994	3.44
I_{diff}	0.86	21.21	0.95	10.96	0.975	8.66	0.984	5.99	0.996	2.74
I_{fuse}	0.88	20.83	0.96	9.43	0.98	7.20	0.986	5.16	0.995	2.68
I_{enh}	0.874	20.57	**0.979**	**7.39**	**0.989**	**5.10**	**0.992**	**4.71**	**0.997**	**2.29**

Performance improves consistently with stronger enhancement. Training on I_{enh} yields the best overall performance (AUC = 0.997), significantly outperforming both non-flash (0.91) and flash (0.969) baselines. This highlights the importance of enhancement in improving ridge clarity and downstream discriminability. The corresponding ROC and FRRvsFAR curves for the best-performing configuration are shown in Fig. 4.10. A mixed-domain setup was also evaluated, but its averaged cross-dataset performance gave limited clarity on dataset-specific performance.



(a) ROC Curve (b) FRR vs FAR

Figure 4.10 ROC and FRR vs FAR curves for the best *TripletDistilNet* model trained on I_{enh}.

4.3.3.1 Cross-Dataset Generalization

To evaluate robustness, we tested *TripletDistilNet* across multiple datasets spanning both contactless and contact-based domains. These include our Flash–Non-Flash (FNF) Database (paired smartphone-captured contactless images), HKPolyU-CL and HKPolyU-C [22] (containing corresponding contactless and contact-based fingerprints acquired using CMOS cameras and optical sensors), the UNFIT dataset [20] (a highly unconstrained mobile fingerphoto dataset captured using 45 different smartphones), and the SP dataset [74] (a small contact-based benchmark with multiple impressions per finger).

Table 4.4 Cross-dataset verification performance of DeepPrint and *TripletDistilNet* (with and without contact fine-tuning).

Method	Metric	FNF	HKPolyU-CL	UNFIT	HKPolyU-C	HKPolyU-CPhoto	SP
DeepPrint	AUC	0.94	0.92	0.90	0.97	0.96	0.97
	EER (%)	12.68	15.87	20.38	8.63	9.08	10.6
TripletDistilNet	AUC	**0.997**	0.96	0.95	0.91	0.83	0.89
	EER (%)	**2.29**	10.77	11.14	15.98	24.20	19.35
TripletDistilNet w/ Contact FT	AUC	0.994	**0.962**	**0.97**	**0.988**	**0.975**	**0.982**
	EER (%)	3.31	**10.42**	**9.86**	**4.74**	**9.25**	**4.86**

We observe in Table 4.4 that DeepPrint performs strongly on contact-based datasets but struggles with contactless images due to domain bias, while *TripletDistilNet*, trained on contactless data, generalizes well to unconstrained scenarios but requires adaptation for contact-based inputs. These performance variations are primarily driven by differences in ridge distortion, sensor characteristics, and illumination conditions across datasets.

4.3.3.2 Contact-Based Fine-Tuning

To bridge the domain gap between contactless and contact-based fingerprints, we fine-tune *TripletDistilNet* on a combined contact dataset of 493 identities curating larger triplet sets (up to $\sim$2.5k triplets), exposing the model to contact-specific ridge distortions and sensor artifacts.

A key challenge is preventing the model from drifting away from the contactless embedding space learned during initial training. To address this, we use a joint optimization strategy combining triplet learning on contact images with knowledge distillation from the original model.

Fine-tuning Loss.

$$\mathcal{L} = \underbrace{\max(0,\ d(e_a, e_p) - d(e_a, e_n) + m)}_{\text{Triplet loss}} + \lambda_{\text{distill}} \underbrace{\|f_s(x_{\text{cl}}) - f_t(x_{\text{cl}})\|_2^2}_{\text{Distillation loss}}. \tag{4.26}$$

Here, e_a, e_p, and e_n are embeddings of anchor, positive, and negative *contact* samples, $d(\cdot, \cdot)$ is the Euclidean distance, and $m = 0.4$ is the margin. The distillation term aligns student embeddings $f_s(\cdot)$ with teacher embeddings $f_t(\cdot)$ on *contactless* inputs x_{cl}, preserving the learned embedding structure. The weight $\lambda_{\text{distill}} = 0.9$ balances adaptation and retention.

4.3.3.3 Cross-Device Evaluation

We also perform cross-device testing across sample datasets of different mobile devices, including Samsung A54 (Device 1), Redmi Note 7 Pro (Device 2), and iPhone 14 (Device 3), which vary in camera sensors, resolution, and flash characteristics. The results in Table 4.5 indicate strong generalization. Despite these variations, *TripletDistilNet* maintains consistent performance, demonstrating robustness to device-specific capture conditions such as illumination differences, noise levels, and perspective dis-

tortions. Furthermore, improved performance on inked fingerprints after fine-tuning indicates successful adaptation to contact-based ridge characteristics.

To better bridge the domain gap, we additionally incorporate ink-style transfer on a subset of contact training images using a pretrained VGG19-based Adaptive Instance Normalization (AdaIN) framework [69, 75], which helps simulate ink-like texture variations and improves generalization to inked impressions. While performance on inked data improves significantly after fine-tuning, it remains influenced by inherent ink artifacts and the absence of dedicated preprocessing tailored to ink-based fingerprints. The model is trained for 10 epochs using Adam ($lr = 10^{-4}$, weight decay $= 10^{-5}$), enabling effective adaptation to contact-domain features while maintaining a unified embedding space for robust cross-domain verification.

Table 4.5 Cross-device verification performance of DeepPrint and *TripletDistilNet* (with and without contact fine-tuning) across multiple contactless devices and inked fingerprints.

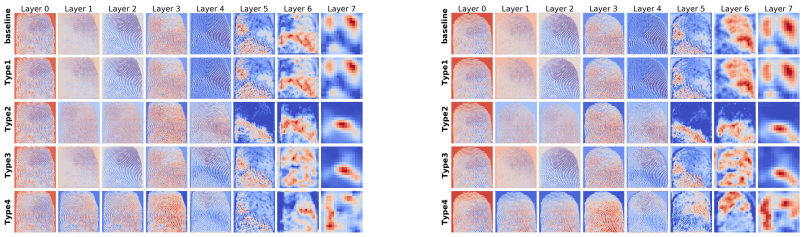
Method	**Metric**	**Device 1**	**Device 2**	**Device 3**	**Inked**
DeepPrint	AUC	0.94	0.91	0.93	0.963
	EER (%)	11.23	16.21	13.94	10.02
TripletDistilNet	AUC	**0.996**	0.95	0.96	0.84
	EER (%)	**2.98**	11.86	9.34	22.42
TripletDistilNet w/ Contact FT	AUC	0.995	**0.96**	**0.97**	**0.97**
	EER (%)	3.26	**10.98**	**8.91**	**9.01**

4.3.3.4 Fine-Tuning Configurations and Analysis

To study the effect of adapting different parts of *TripletDistilNet* for cross-domain generalization, we evaluated four fine-tuning strategies:

1. *Type 1*: Head-only fine-tuning
2. *Type 2*: Backbone-only fine-tuning
3. *Type 3*: Backbone Late-layer + head fine-tuning
4. *Type 4*: Backbone Early-layer + head fine-tuning (best)

These configurations enable us to explore how different levels of feature abstraction contribute to domain adaptation. The embedding head mainly learns high-level identity representations, and the backbone encodes hierarchical features from low-level ridge texture to global-level fingerprint structure. Fine-tuning selective parts of the network strikes a balance in adapting to the contact domain and preserving the previously learned contactless domain representations.



(a) Contact example 1

(b) Contact example 2

(c) Contactless example 1

(d) Contactless example 2

Figure 4.11 Representative activation maps illustrating ridge-focused attention after fine-tuning. Top row shows contact samples; bottom row shows contactless samples. Red regions indicate stronger feature emphasis due to higher activation values and blue lower values.

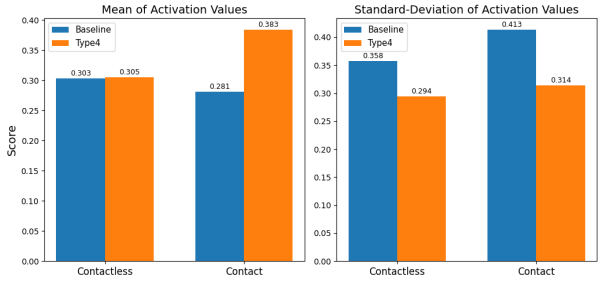


Figure 4.12 Plots for mean and standard-deviation of activation values for contact and contactless inputs across all layers for baseline and Type 4 (best) fine-tuned *TripletDistilNet* models.

Observations

- Fine-tuning both early backbone layers and the embedding head (*Type 4*) provides the best balance between adaptation and stability, enabling the model to learn contact-specific ridge statistics while preserving higher-level discriminative features. This leads to improved modeling of ridge flow and texture in contact images through early-layer adaptation, while maintaining global fingerprint structure and identity-specific features via stable deeper layers, and enhancing focus on ridge patterns over background or boundaries in contactless inputs.

- Head-only fine-tuning (*Type 1*) results in limited improvements, as the fixed backbone cannot adapt to low-level variations such as ridge compression and smudging, restricting learning to minor embedding-level adjustments.

- Backbone-only fine-tuning (*Type 2*) risks catastrophic forgetting, where early convolutional filters overwrite previously learned contactless features, improving contact performance but degrading generalization to contactless inputs.

- Late-layer fine-tuning (*Type 3*) provides partial adaptation by refining higher-level features while keeping early layers fixed, but fails to adequately capture low-level ridge variations, making it less effective than *Type 4*.

- Activation map analysis in Figures 4.11 and 4.12 further supports these findings, showing stronger emphasis on core and ridge regions after fine-tuning. Compared to the baseline, the fine-tuned model exhibits increased mean activation values, indicating stronger propagation of ridge-related features, and reduced standard deviation, suggesting more consistent and task-focused feature extraction, while maintaining minimal drifting from baseline distributions, confirming that the overall embedding structure remains stable.

In conclusion, these results emphasize the importance of selectively adapting early feature extractors, while preserving deeper representations, for achieving a unified embedding space that generalizes effectively across both contact and contactless fingerprint domains.

4.3.4 Verification Performance of Fusion2Print

We further evaluate the complete F2P pipeline by training deep embeddings using both 128-dimensional and 256-dimensional representations. The average templating latency is approximately *0.28s* per identity for Apple M4 GPU and *0.34s* for a Samsung A54 device, demonstrating practical efficiency.

Table 4.6 compares DeepPrint with various F2P configurations using intra-class distance (Intra), inter-class distance (Inter), and separation ratio (SepRatio = Inter/Intra). Fusion2Print significantly improves embedding separability, achieving up to a 1.66$\times$ improvement in SepRatio over the DeepPrint baseline.

Table 4.7 reports AUC and EER for different embedding configurations. Fine-tuning consistently improves performance, and higher-dimensional embeddings (256-D) provide better discrimination. Incorporating spatial transformation further enhances robustness, achieving the best overall results. This is also visualized in Figure 4.13.

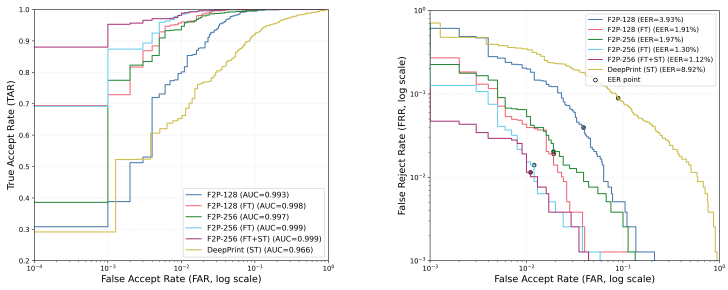
We also present a comparison between the DeepPrint baseline and F2P representations (Figure 4.14). In panels (a) and (b), the t-SNE embeddings illustrate that DeepPrint exhibits lower intra-subject consistency, as evidenced by the numerous red links, whereas F2P produces more stable features, resulting in higher verification accuracy. Panels (c) and (d) show pairwise embedding distance heatmaps, where

Table 4.6 Comparison of DeepPrint baseline and Fusion2Print variants with 128-D and 256-D embeddings, with and without spatial transform optimization.

Model Variant	Embed Dim	Euclidean			Cosine		
		Intra$_{\downarrow}$	Inter$_{\uparrow}$	SepRatio$_{\uparrow}$	Intra$_{\downarrow}$	Inter$_{\uparrow}$	SepRatio$_{\uparrow}$
DeepPrint (Baseline)	512	0.24	0.84	3.44	0.03	0.36	10.46
DeepPrint w/ ST	512	**0.20**	0.73	3.62	**0.024**	0.28	11.34
F2P (Base)	128	0.43	1.34	3.07	0.11	0.993	8.90
F2P (Fine-tuned)	128	0.33	1.36	4.02	0.06	0.99	15.15
F2P (Base)	256	0.42	1.34	3.27	0.09	0.97	10.08
F2P (Fine-tuned)	256	0.32	1.36	4.19	0.06	**1.01**	16.37
F2P (Fine-tuned w/ ST)	256	0.31	**1.37**	**4.30**	0.05	0.995	**17.38**

Table 4.7 Verification performance for I and I_{flash} inputs across different embedding dimensions and F2P pipeline variants.

Embedding Dim	Pipeline Type	AUC	EER (%)
128	Base	0.993	3.93
	Finetuned	0.998	1.91
256	Base	0.997	1.97
	Finetuned	0.999	1.30
	Finetuned w/ ST	**0.999**	**1.12**

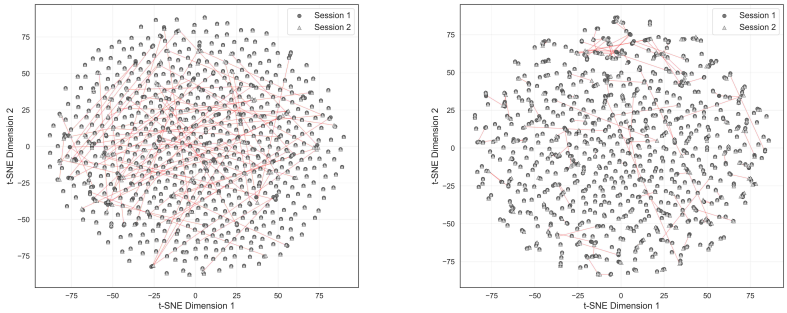


(a) ROC Curve (b) FRR vs FAR

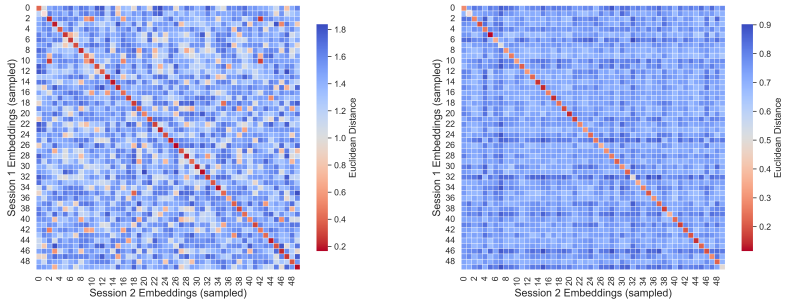
Figure 4.13 ROC and FRRvsFAR curves for DeepPrint baseline and Fusion2Print pipeline variants.

F2P has fewer off-diagonal matches, indicating stronger identity discrimination and clearer separation between subjects.

Overall, the proposed F2P pipeline demonstrates consistent improvements in both ridge quality and verification accuracy. The combination of learned enhancement, discriminative embedding learning, and spatial alignment results in a robust system capable of handling contactless fingerprint variability while outperforming strong minutiae-based and deep-learning baselines.



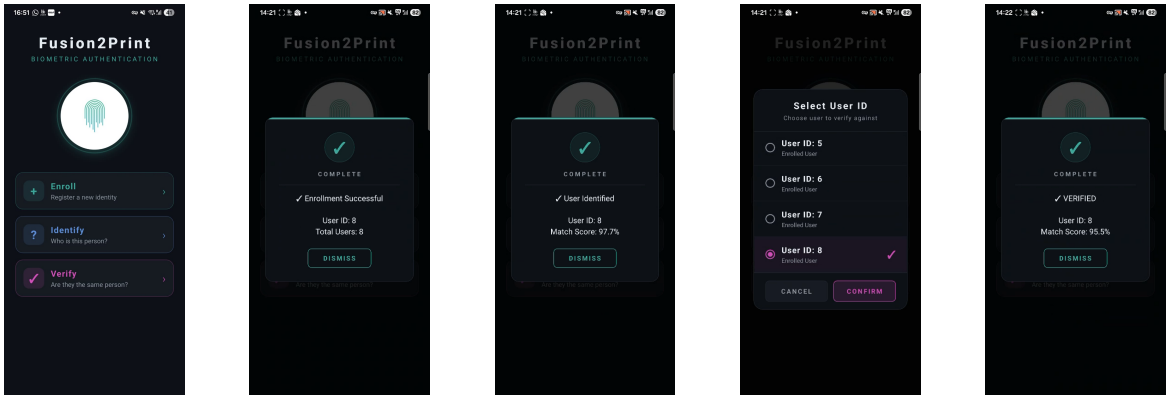
(a) DeepPrint (b) F2P



(c) DeepPrint (d) F2P

Figure 4.14 Embedding space visualization and distance analysis. F2P exhibits improved intra-class compactness and inter-class separation compared to DeepPrint.

4.4 Preliminary Development of End-to-End Application



(a) Home Page (b) Enrolled Users Page (c) Identification Page (d) Verify: Select User (e) Verified Results Page

Figure 4.15 The snapshots illustrate the main functionalities of the application, (a) including the home page, (b) enrolled users page, (c) identification page, (d) verification selection interface, and (e) verified results page.

To evaluate the performance of our framework on edge devices, we developed a smartphone application[1] that supports three key operations: *enrollment*, *identification* (1:N matching) and *verification* (1:1 matching). The app captures both flash and non-flash fingerprint images consecutively, then sends them to the backend where F2P processes the inputs to produce discriminative embeddings. Depending on the selected operation, the embeddings are used to enroll a new user, verify a claimed identity, or identify a user from the database.

Some UI screens are shown in Figure 4.15. This setup demonstrates the practical deployment of F2P in a real-world mobile environment and validates its usability for on-device contactless fingerprint recognition.

4.5 Conclusion

This work introduces **Fusion2Print (F2P)**, the first end-to-end contactless fingerprint verification system integrating flash–non-flash fusion, learned enhancement, and discriminative embedding generation. Additionally, this work uses our **Flash–Non-Flash Fingerphoto (FNF) Database**, a paired dataset for modeling the complementary characteristics of flash and non-flash images.

Aligned flash–non-flash subtraction revealed ridge-preserving cues for improving the localization of minutiae features. *DualEncoderFusionNet* fuses illumination-robust features, and *U-NetEnhancer* refines ridge details. *TripletDistilNet* generates compact embeddings for generalization across contact, contactless, and mixed-domain images, as cross-domain fine-tuning bridges modality gaps. The F2P pipeline finally integrates fusion, enhancement, and embedding generation in a combined framework.

End-to-end optimization provides excellent performance across sessions, where spatially normalized 256-D embeddings attain **AUC=0.999** and **EER=1.12%**, outperforming minutiae-based and single-capture approaches. F2P thus facilitates accurate contactless and **cross-domain fingerprint verification**. It is resilient to changes in devices, environments, and modalities but still lacks adversarial robustness. Code has been released for reproducibility[2].

4.5.1 Future Work

Future work can include adding 3D or multi-view information to enhance the robustness to pose and perspective distortion in contactless capture. The challenge of handling problematic artifacts such as motion blur, illumination imbalances, and occlusions is another area to explore for more advanced forms of augmentation and domain generalization.

The integration of spoof detection in the embedding pipeline is another promising avenue for improving security in real-world systems. Lastly, the development of more efficient architectures and deployment schemes is necessary to support real-time processing on edge devices such as smartphones, facilitating wider adoption in mobile biometric applications.

[1] E2E Fusion2Print Application Code
[2] Fusion2Print Source Code

Chapter 5

Illumination-Aware Contactless Fingerprint Spoof Detection via Paired Flash–Non-Flash Imaging

Contactless fingerprint recognition is being considered a hygienic and user-friendly substitute for traditional contact-based biometric systems. By eliminating physical interaction with sensors, it avoids issues such as latent prints, surface wear, and hygiene concerns. Nevertheless, this has also created a number of problems, particularly with regard to presentation attack detection (PAD). Unlike contact-based systems, which can use physical characteristics such as skin deformation, perspiration patterns, and conductivity for spoof detection, contactless systems rely solely on visual characteristics under unconstrained imaging conditions. Hence, by their very nature, contactless systems are more prone to spoof attacks by printed images, digital displays, and high-quality 3D replicas to name a few.

5.1 Introduction and Background

Existing contactless fingerprint-based PAD systems are mostly based on single-image acquisition and appearance-based features. Although such systems are effective under constrained conditions, there are limitations to their generalization capabilities under variations of devices, lighting conditions, and spoofing materials. This limitation has created a need for additional sensing dimensions that can highlight the intrinsic physical differences between genuine and spoof fingerprints.

In this work, we explore illumination variation as a lightweight and practical sensing mechanism for spoof detection. Specifically, we consider paired flash–non-flash image acquisition using smartphone cameras. This setup introduces controlled variation in lighting while keeping the fingerprint identity fixed, enabling the analysis of how different materials and structures respond to illumination changes. Unlike single-image approaches, paired acquisition captures both passive appearance (non-flash) and active illumination response (flash), providing complementary information within a single capture process.

5.2 Proposed Objective

The central question we explore is: *Can illumination-induced variations in contactless fingerprint images be exploited to aid spoof detection?*

To investigate this, we conduct a preliminary empirical analysis of illumination-induced variations in contactless fingerprint images. This analysis is aligned with our broader end-to-end contactless fingerprint pipeline, Fusion2Print (F2P), which performs enhancement and embedding generation. The insights derived here introduce physics-informed cues that can be integrated into the pipeline to improve robustness against presentation attacks.

5.3 Preliminary Analysis of Illumination-Dependent Cues

We perform a systematic study of paired flash–non-flash fingerprint images to characterize how illumination affects ridge visibility, texture, and material response.

5.3.1 Dataset

The dataset(unreleased due to privacy contraints) consists of 800 genuine fingerprint images and 1600 spoof samples–including digital replay (HD laptop screen) and colour print attacks–collected from 20 diverse individuals across two sessions. Images were captured using a Samsung Galaxy A54 smartphone with a custom app(as used in Section 3.1.1) that records consecutive flash and non-flash image pairs, following the same protocol used in collecting FNF Database[1].

5.3.2 Photometric and Image Quality Variations

We first analyze how flash illumination impacts image quality and ridge clarity. Metrics such as Orientation Certainty Level (OCL) [76], Local Clarity Score (LCS) [77], and NFIQ2 [73] are used to quantify ridge consistency and perceptual quality, along with additional sharpness, contrast, and edge-based measures.

As shown in Fig. 5.1, blockwise OCL maps indicate stronger ridge coherence under flash illumination. Fig. 5.2 further demonstrates improved ridge-valley contrast using LCS and intensity profiles. Quantitative comparisons in Fig. 5.3 show that flash images consistently outperform non-flash images across both standard (AIT Sharpness [78], NFIQ2) and custom patch-wise quality metrics.

These observations indicate that controlled illumination enhances discriminative ridge structures, which are critical for both recognition and spoof detection.

[1]FNF Database

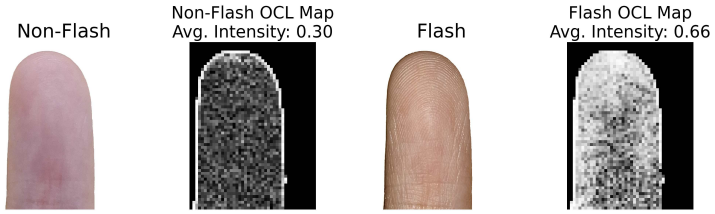


Figure 5.1 Blockwise OCL maps for sample flash and non-flash fingerprint images. Brighter regions(higher intensity) indicate higher ridge clarity; flash images show consistently stronger ridge coherence.

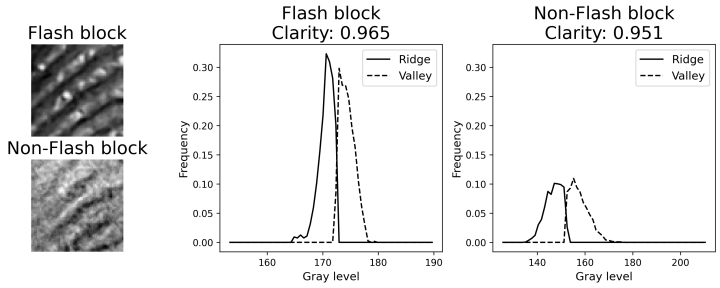


Figure 5.2 LCS and ridge-valley intensity profiles for sample flash and non-flash fingerprint blocks, showing improved ridge-valley contrast under flash illumination.

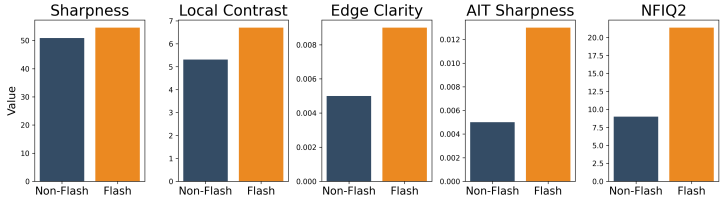


Figure 5.3 Comparison of standard (AIT Sharpness, NFIQ2) and custom patch-wise image quality metrics over the sample dataset, where flash images outperform non-flash images across all measures.

5.3.3 Channel-wise Photometric Analysis

Beyond overall quality, we analyze channel-specific behavior under varying illumination. *Local contrast*(average standard deviation of pixel intensities of non-overlapping patches) and *edge energy*(mean squared Sobel gradient magnitude) are computed independently for RGB channels to capture material-dependent reflectance properties.

As illustrated in Fig. 5.4, flash illumination introduces stronger highlights and channel imbalances, particularly in spoof artifacts. These effects are less pronounced in genuine fingerprints due to natural skin properties such as subsurface scattering and consistent pigmentation. This difference provides an important cue for distinguishing real and fake samples.

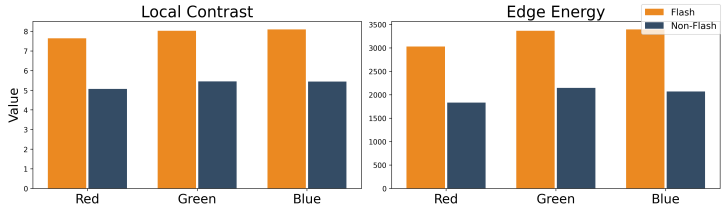
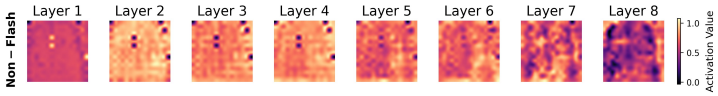
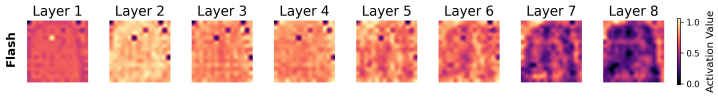


Figure 5.4 Photometric analysis of local contrast and edge energy for Red, Green, and Blue channels of the sample dataset under flash and non-flash illumination.

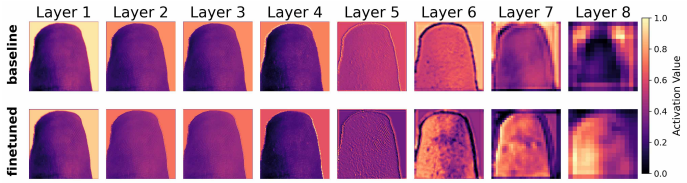
5.3.4 Model Interpretability Under Illumination Change



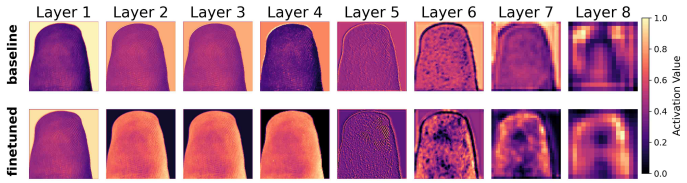
(a) Non-Flash with DINOv2



(b) Flash with DINOv2



(c) Non-flash with ResNet-18 based model



(d) Flash with ResNet-18 based model

Figure 5.5 Attention maps for sample flash and non-flash contactless fingerprint images. (a)–(b) Baseline DINOv2 attention maps with limited ridge awareness but improved spatial structure under flash lighting. (c)–(d) Fine-tuned ResNet-18–based model showing enhanced ridge-focused attention, especially for flash images.

To understand how learning-based models respond to illumination variations, we analyze attention maps from both generic and fingerprint-specific architectures.

As shown in Fig. 5.5, baseline DINOv2 [79] models exhibit coarse spatial attention with limited ridge awareness, while fine-tuned ResNet-18 [72] models focus more effectively on ridge structures. Flash images produce stronger and more localized attention responses, indicating enhanced feature saliency.

Table 5.1 Activation statistics for genuine (G) and spoof (S) fingerprints under flash and non-flash conditions for the ResNet-18–based model (mean $\pm$ standard deviation).

Model	**G (Non-Flash)**	**G (Flash)**	**S (Non-Flash)**	**S (Flash)**
Baseline	0.21 ± 0.19	$\mathbf{0.23 \pm 0.16}$	0.22 ± 0.24	0.23 ± 0.21
Fine-tuned	0.37 ± 0.16	$\mathbf{0.49 \pm 0.13}$	0.26 ± 0.22	0.27 ± 0.14

Further quantitative analysis in Table 5.1 demonstrates improved genuine–spoof class separation after fine-tuning. Specifically, Fisher Discriminant Ratios (FDR) increase from $5\mathrm{e}{-4}$ to 0.18 for non-flash images and from $7\mathrm{e}{-4}$ to 1.48 for flash images, indicating significantly stronger discriminative capability under illumination-aware learning. The FDR measures the separation between class means normalized by their variances, with higher values indicating better class separability.

5.4 Presentation Attack Scenarios

In contactless fingerprint recognition systems, the absence of contact between the finger and the sensor has a significant impact on the variety of possible presentation attack instruments (PAIs), which is not the case with contact-based systems. In contactless systems, there are many media that are possible PAIs, each having unique photometric and material characteristics. These characteristics are particularly noticeable under different lighting conditions.

Printed spoofs, such as inkjet or laser printouts, can maintain ridge layout and look realistic under non-flash lighting conditions. However, their planar nature introduces flat reflectance characteristics, ink-dependent color distortions, and strong specular reflections under flash lighting. The absence of subsurface scattering and surface uniformity introduces illumination responses that are quite different from real skin.

Digital display-based spoofs, presented on smartphones or laptops, have high visual realism under ambient lighting. Under flash, however, they show artifacts such as screen glare, local saturation, and even pixel-level patterns such as grids or moiré effects [80]. These displays also show unnatural RGB responses with strong channel imbalances, unlike the smoother spectral behavior of genuine skin.

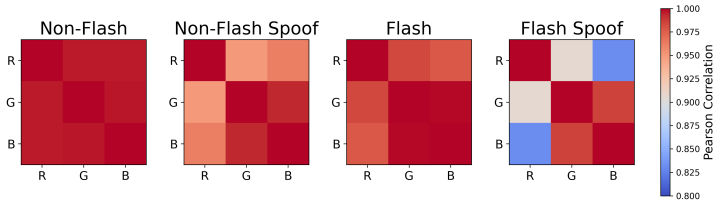
Physical or molded spoofs, made of materials like silicone, latex, resin, or even 3D models created from 2D contactless images [56], try to reproduce appearance and geometry. Although they might resemble ridge depth under non-flash conditions, under flash, the inconsistencies of the materials are revealed, such as the over uniformity of specular reflections, loss of micro-texture variation, and the absence of subsurface scattering effects. In addition, the absence of natural skin components such as oil distribution results in inconsistent ridge-valley reflectance.

Overall, while spoof artifacts might resemble genuine fingerprints in a single illumination setting, they do not demonstrate the consistency when varying the illumination conditions. These differences are effectively captured by the paired flash–non-flash acquisition, which provides discriminative cues for robust PAD in contactless fingerprint systems.

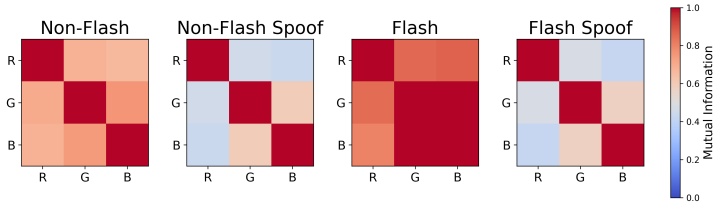
5.5 Illumination-Aware Cues for Spoof Detection

Based on the empirical analysis we identified illumination-dependent cues capturing material, structural, and photometric differences between genuine and spoof fingerprints. The following subsections describe these cues and their applications, supported by observations from our dataset.

5.5.1 Inter-channel Correlation



(a) Pearson correlation heatmaps



(b) Mutual information heatmaps

Figure 5.6 Inter-channel correlation analysis for the sample dataset of flash and non-flash fingerprint images and their spoof counterparts. (a) Pearson correlation and (b) mutual information (MI) heatmaps show stronger off-diagonal channel relationships for genuine images, with clearer distinction from spoof under flash illumination. Mann-Whitney U tests confirm the statistical significance of off-diagonal activations (Pearson: $p < 0.001$ for both conditions; MI: $p < 0.001$ for flash and $p = 0.035$ for non-flash).

In real skin, there is a high inter-channel correlation between the RGB channels due to uniform pigmentation and subsurface scattering while in spoofing artifacts, there are inconsistencies between the channels.

As shown in Fig. 5.6, both Pearson correlation and mutual information heatmaps reveal stronger off-diagonal relationships for genuine samples, particularly under flash illumination. Flash lighting am-

plifies composition-dependent reflection differences, resulting in more pronounced inter-channel decorrelations between genuine and spoof samples. Quantitatively, this is reflected in higher genuine–spoof separation (Δ) under flash conditions (Pearson: $\Delta_{\text{flash}} = 0.090$ vs. $\Delta_{\text{non-flash}} = 0.032$; Mutual Information: $\Delta_{\text{flash}} = 0.167$ vs. $\Delta_{\text{non-flash}} = 0.041$), where Δ denotes the separation in inter-channel correlation statistics.

5.5.2 Specular Reflection Characteristics

Flash illumination highlights the specular reflections, which vary significantly depending on the surface reflectance properties. Genuine fingerprints have well-controlled reflections depending on the ridges due to skin micro-geometry and subsurface scattering, while spoofed ones have over-emphasized or spatially uniform highlights.

We quantify this behavior using a specular highlight ratio (SHR), defined as the proportion of high-intensity pixels under flash illumination relative to non-flash conditions. Analysis over our dataset shows that spoof samples exhibit higher(0.043) and more variable average specular highlight ratio compared to genuine fingerprints(0.009). Printed attacks show elevated reflectance due to ink properties, while display-based attacks exhibit localized saturation caused by screen glare.

Overall, genuine fingerprints maintain lower and more stable specular responses, making this cue highly discriminative for spoof detection.

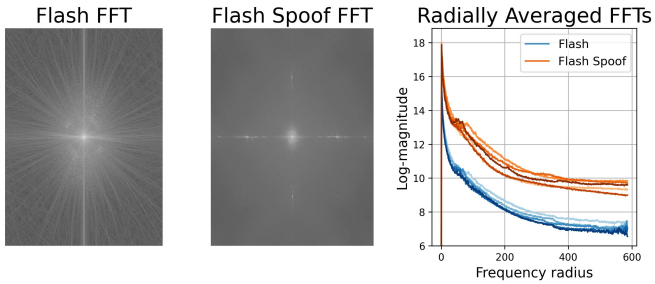
5.5.3 Texture Realism and Frequency Analysis

Texture descriptors such as Local Binary Patterns (LBP) [81], Gray Level Co-occurrence Matrices (GLCM) [82], and Fourier spectrum analysis reveal differences in ridge periodicity and structural realism. Genuine fingerprints exhibit natural variability in ridge structure, while spoof artifacts tend to show smoother, repetitive, or grid-like patterns due to printing or casting processes.

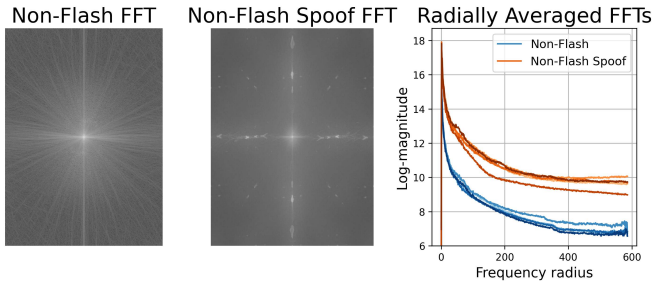
Frequency-domain analysis further highlights these differences, with spoof samples exhibiting dominant peaks and structured high-frequency components indicative of artificial periodicity. As shown in Fig. 5.7, this results in a higher texture realism ratio(fraction of blocks with dominant FFT peaks) of the dataset for spoof fingerprints (0.226) compared to genuine fingerprints (0.087). This is indicated by disrupted, grid-like textures and increased high-frequency artifacts enabling a clear distinction. Additionally, variations in texture descriptors across flash and non-flash images (e.g., ΔLBP, ΔGLCM, ΔFourier) provide a measure of illumination-dependent texture stability, where genuine fingerprints remain more consistent than spoof materials.

5.5.4 Differential Imaging

Differential imaging, computed as the difference between flash and non-flash images, suppresses ambient illumination and isolates material-dependent responses.



(a) Texture Realism analysis for Flash



(b) Texture Realism analysis for Non-Flash

Figure 5.7 Log-magnitude FFTs and radially averaged spectra are shown for (a) flash and (b) non-flash contactless fingerprint images of genuine and spoof samples.

Genuine fingerprints exhibit structured variations due to subsurface scattering and ridge geometry, whereas spoof artifacts show weaker or inconsistent changes. This effect is particularly evident in display-based attacks, where illumination changes do not correspond to physical surface interactions.

5.5.5 Physical and Material Properties

Apart from photometric and texture-based cues, physical and material properties are also important discriminative factors for spoof attack detection.

- **Subsurface scattering**: Human skin has a subsurface scattering effect [83], where light is absorbed within the skin and diffusely reflected back to the camera. This causes a soft, spatially varying illumination effect when flash is used. Spoofing attacks generally lack this effect and have a direct surface-level reflection of light.

- **Micro-geometry**: Genuine fingerprints contain natural ridge irregularities, including variations in ridge width, spacing, and curvature. In contrast, spoof artifacts–especially printed or molded ones–tend to exhibit more uniform and smoothed structures due to limitations in fabrication processes.

- **Surface oils and sweat**: Real skin contains oils and perspiration that create non-uniform, dynamic specular highlights. These irregular reflections vary spatially and temporally, whereas spoof materials typically produce more uniform and static highlight patterns.

These cues collectively form a *physics-informed feature space* for spoof detection.

5.6 Future Integration with End-to-End Pipeline

The insights from this analysis are not standalone; they directly complement our broader contactless fingerprint pipeline. Our system, F2P includes enhancement and embedding production which benefits from illumination-aware processing.

The proposed spoof detection cues can be integrated into this pipeline in multiple ways

- As auxiliary features for learning-based PAD models
- As regularization signals during training
- As interpretable metrics for decision validation

This unified framework enables a single paired acquisition to support enhancement, recognition, and spoof detection, improving both performance and interpretability.

5.7 Discussion and Limitations

Although illumination-aware spoof detection offers a feasible and effective sensing method, challenges exist. Changes in acquisition environment, such as pose, distance, background, and lighting, may have an impact on the consistency between flash and non-flash images. Smartphone acquisition may bring challenges, including motion blur, exposure, and temporal misalignment between images, which may affect the accuracy of the paired differential analysis methods.

Another important limitation stems from the ever-increasing levels of sophistication associated with presentation attacks. Advanced spoofing attacks involving high-quality 3D replicas created using skin-like materials and display attacks with anti-reflective coatings and brightness adaptation have been shown to partially mimic the illumination responses. This reduces the discrimination difference between genuine and spoof photometric signals, making discrimination more challenging, especially under uncontrolled conditions.

In addition, there is a minor overhead in terms of capture time and user interaction, which might have usability implications. In terms of data, the existing datasets are still limited in terms of scale, diversity, and the availability of high-fidelity spoof material. There are also issues related to the generalizability of the results due to device and sensor variations, as the illumination conditions and camera pipelines vary significantly.

In order to overcome these limitations, there is a need for large and diverse datasets, aligning and synchronizing data methods as well as effective models for learning invariant features across devices and environmental conditions.

5.8 Conclusion

In this work, we proposed an **illumination-aware framework for contactless fingerprint spoof detection** based on paired flash and non-flash imaging modalities. Through empirical analysis, we showed that illumination variations between the two modalities capture complementary cues related to material properties, inter-channel behavior, texture realism, and micro-geometry, as supported by quantitative and qualitative evaluations and statistically significant observations, highlighting their effectiveness for spoof discrimination.

From the results, we see that while spoof images may look similar under single illumination, they have different photometric and structural behaviors under flash variation. The flash images help in enhancing the ridges and reflection, while the non-flash images provide the reference appearance. The differential imaging and attention-based model interpretability help in highlighting the differences, enabling physics-informed feature extraction beyond traditional appearance-based methods.

Through these differences, paired acquisition becomes a simple active sensing approach that improves interpretability and provides strong quantitative and qualitative indicators for detecting a wide range of spoofing types, including printed, digital, and molded spoofing artifacts, without the need for special hardware.

These insights establish a foundation for integrating illumination-aware and physics-informed spoof detection into end-to-end contactless fingerprint pipelines, enabling improved robustness, interpretability, and generalization in real-world biometric systems, despite challenges in capture variability, dataset scale, and evolving spoof technologies. Code is available in the repository[2].

[2]Spoof Detection Source Code

Chapter 6

Low-cost 3D Reconstruction of Contactless Fingerprints using Flash–Non-Flash Pairs

Three-dimensional (3D) reconstruction of fingerprints from contactless images is a crucial step towards enhancing the robustness of fingerprint recognition systems, as well as spoof detection and sensor interoperability. Traditional methods for 3D reconstruction usually require specific equipment, for example, multi-camera systems, structured light systems, or even controlled illumination systems, which limit scalability and real-world deployment.

6.1 Introduction

In this chapter, we explore a **low-cost 3D reconstruction pipeline** that leverages only **paired flash and non-flash images** captured using smartphones. The key idea is that illumination variation between flash and non-flash captures implicitly encodes surface geometry. While a single image lacks sufficient depth cues, combining these two complementary observations provides valuable information about surface normals and ridge-valley structures.

This approach is inspired by photometric stereo principles but operates under highly fewer conditions (only two images, varying illumination). We investigate how effectively depth and surface structure can be recovered under these limitations.

A key challenge lies in the non-Lambertian nature of human skin [84], which introduces specularities and subsurface scattering effects. Despite these challenges, we demonstrate that meaningful 3D approximations of fingerprint geometry can be obtained using lightweight computational methods.

6.2 Background and Terminology

This section is an introduction to the basic concepts and tools that are used in the proposed low-cost 3D reconstruction method. The task of recovering 3D structure from two-dimensional (2D) images is an extensively studied computer vision problem, with solutions ranging from traditional physics-based approaches to recent learning-based techniques. For the task of contactless fingerprint reconstruction, it is necessary to consider how surface geometry, lighting, and reflectance combine to affect the intensity

of the observed images. The following definitions are an essential background on the task of recovering shape, modeling reflectance, and recent progress in neural and real-time 3D reconstruction, forming the basis for the methodology presented in this chapter.

- **Shape from Normals:** A 3D reconstruction method in which surface normals are first determined and then used to obtain the depth of the surfaces.
- **Shape from Shading (SfS):** This is a method for 3D surface reconstruction from a single grayscale image. It is done by analyzing the shading variations in the image and estimates surface orientation assuming known lighting and typically relies on Lambertian reflectance.
- **Lambertian Surface:** This is an ideal diffuse surface, meaning it reflects light in all directions. However, *human skin does not conform to the Lambertian assumption* because of the presence of specular reflections and subsurface scattering.
- **Bidirectional Reflectance Distribution Function (BRDF):** A function that models how light reflects off a surface, mapping incoming light directions to outgoing viewing directions. It generalizes reflectance beyond the Lambertian assumption.
- **Photometric Stereo:** This is a method for determining surface normals from multiple images captured under varying illumination. For classical photometric stereo, at least three images are required for Lambertian surfaces.
- **Structure from Motion (SfM):** This is another multi-view reconstruction method that estimates 3D scenes using images obtained at different viewpoints. However, this method requires viewpoint changes rather than illumination changes.
- **Frankot Chellappa Method:** This is a gradient integration method that produces a physically valid surface (depth map) using gradient estimations by enforcing integrability using Fourier transforms.
- **Gaussian Splatting (3DGS) [85]:** This is a new rendering method that employs Gaussian volume primitives for fast and realistic rendering.
- **Neural Radiance Fields (NeRF) [86]:** A neural representation of 3D scenes that models volumetric density and radiance. While highly realistic, it is computationally expensive.
- **Deep Learning for Photometric Stereo (PS-FCN):** This method employs learning techniques to obtain normal maps using multiple images and outperforms traditional methods in non-ideal scenarios.
- **SHARP (Sharp Monocular View Synthesis) [59]:** This is a new model that produces 3D representations using a single image in real-time and demonstrates the potential of learning-based monocular methods.

- **UV Mapping [87]:** This is a method that maps a 2D texture onto a 3D object in order to visualize its geometry in a realistic manner.

6.3 Proposed Methodology

The proposed pipeline shown in Figure 6.1 offers a unified framework for low-cost 3D fingerprint reconstruction with the integration of illumination cues and geometric priors. Beginning with flash and non-flash inputs, the method first computes a spectrally enhanced difference image FNF, then recovers the coarse global finger shape using monocular 3D representation, followed by the refinement of local ridge-valley structures using the SfS representation. These two representations are then aligned and fused to achieve an overall coherent 3D fingerprint model that captures both macroscopic curvature and fine-scale surface details.

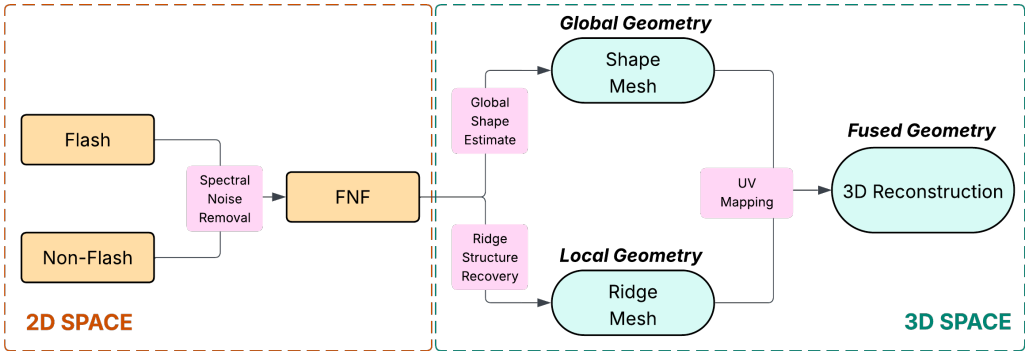


Figure 6.1 Overview of the proposed low-cost 3D reconstruction pipeline using 2D flash and non-flash fingerprint images.

6.3.1 Global Shape Estimation

We estimate the coarse 3D geometry (finger dome shape) from a single flash image using a monocular reconstruction model, SHARP [59]. The flash image is preferred over non-flash due to its enhanced edge definition and stronger shading cues, which improve geometric inference.

1. Monocular 3D Representation via Gaussian Splatting

Given an input image $I(x, y)$, the model predicts a 3D Gaussian Splat representation of the scene. Instead of an explicit mesh, the geometry is represented as a set of N volumetric primitives:

$$\mathcal{G} = \{(\boldsymbol{\mu}_i, \Sigma_i, \alpha_i, \mathbf{c}_i)\}_{i=1}^{N} \tag{6.1}$$

where $\boldsymbol{\mu}_i \in \mathbb{R}^3$ denotes the 3D mean corresponding to the center of each Gaussian, $\Sigma_i \in \mathbb{R}^{3\times3}$ represents the covariance matrix that defines its shape and orientation, α_i controls the opacity, and $\mathbf{c}_i$

encodes the color or radiance. This representation implicitly encodes the finger's global surface structure, as illustrated in Figure 6.2.

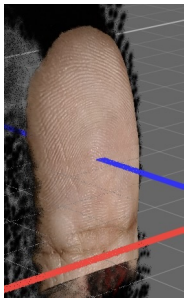


Figure 6.2 High-quality gaussian splat rendering of a flash-captured fingerprint.

2. Conversion to Point Cloud Representation

To enable geometric processing, the Gaussian representation is converted into a point cloud. Specifically, the set of volumetric primitives is reduced to a collection of 3D points:

$$\mathcal{P} = \{\mathbf{p}_i \in \mathbb{R}^3 \mid \mathbf{p}_i = \boldsymbol{\mu}_i\} \tag{6.2}$$

Here, each Gaussian center is treated as a 3D point. Since the reconstructed scene may include background or noise, a depth-based filtering step is applied to retain only relevant surface points by selecting $\mathcal{P}_{filtered} = \{\mathbf{p}_i \in \mathcal{P} \mid z_i < \tau\}$, where z_i is the depth coordinate and τ is a threshold used to remove outliers and background points. The resulting filtered point cloud captures the overall fingerprint geometry, as illustrated in Figure 6.3.

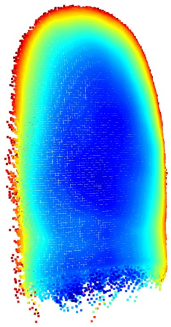


Figure 6.3 Point cloud representation of a flash-captured fingerprint global shape.

Point Cloud Refinement. The filtered point cloud is further refined to guarantee geometric consistency. Downsampling is employed to remove redundancy and ensure more uniform spatial sampling, and noise removal techniques such as statistical or radius-based filtering are utilized to remove sparse

outliers and improve the quality of the reconstructed surface.

3. Surface Normal Estimation

For each point $\mathbf{p}_i$, the surface normal $\mathbf{n}_i$ is estimated using local neighborhood analysis. Given a set of neighboring points $\mathcal{N}(\mathbf{p}_i)$, the normal is computed via Principal Component Analysis (PCA):

$$\mathbf{n}_i = \arg\min_{\|\mathbf{n}\|=1} \sum_{\mathbf{p}_j \in \mathcal{N}(\mathbf{p}_i)} (\mathbf{n}^\top (\mathbf{p}_j - \mathbf{p}_i))^2 \tag{6.3}$$

The eigenvector corresponding to the smallest eigenvalue of the covariance matrix defines the normal direction. To ensure consistency across the surface, normals are oriented such that $\mathbf{n}_i \cdot \mathbf{v} > 0$, where $\mathbf{v}$ is the viewing direction, enforcing outward-facing normals for the finger dome.

4. Mesh Reconstruction via Poisson Surface Reconstruction

The refined point cloud with normals is converted into a continuous surface using Poisson surface reconstruction [88]. This method solves for an indicator function $\chi(\mathbf{x})$ such that $\nabla\chi \approx \mathbf{V}$, where $\mathbf{V}$ is the vector field defined by the oriented normals. The reconstruction is obtained by solving the Poisson equation $\nabla^2\chi = \nabla \cdot \mathbf{V}$. The resulting implicit function is then converted into a triangle mesh by extracting an iso-surface.

Mesh Refinement. The reconstructed mesh is further refined by applying density-based filtering to remove low-confidence vertices corresponding to sparse regions, followed by smoothing using mild Laplacian filtering to reduce noise while preserving global curvature.

5. Output Representation

The final output is a coarse 3D mesh representing the finger dome:

$$\mathcal{M}_{dome} = (\mathcal{V}, \mathcal{E}, \mathcal{F}) \tag{6.4}$$

where $\mathcal{V}$ are vertices, $\mathcal{E}$ edges (*trivial*), and $\mathcal{F}$ triangular faces.

This mesh captures the global curvature and structure of the finger as illustrated in Figure 6.4, which is later combined with fine-scale ridge reconstruction to produce the final detailed 3D fingerprint model.

6.3.2 Ridge Structure Recovery

To capture fine-grained fingerprint details, we reconstruct local surface variations corresponding to ridge-valley structures using an SfS-based pipeline.

1. Shape-from-shading for local geometry:

Given a grayscale and enhanced fingerprint image $I(x, y)$, we assume a simplified reflectance model under frontal illumination. Under Lambertian assumptions(human skin variations are partially ad-

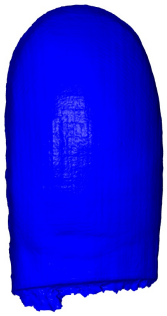


Figure 6.4 Final 3D mesh representation of a flash-captured fingerprint shape, providing a comprehensive basis for low-cost 3D reconstruction.

dressed using CLAHE-based enhancement, following the approach in [89]), image intensity is proportional to the surface normal component along the light direction:

$$I(x, y) \approx \rho \left(\mathbf{n}(x, y) \cdot \mathbf{s} \right) \tag{6.5}$$

where ρ is albedo (assumed constant), $\mathbf{n}$ is the surface normal, and $\mathbf{s}$ is the light direction. Assuming frontal lighting, this simplifies to estimating the z-component of the normal as $n_z \approx I(x, y)$. From this, the gradient magnitude is derived as $\|\nabla z\| = \sqrt{\frac{1}{n_z^2} - 1}$.

2. Gradient field estimation:

The direction of surface gradients is obtained from image derivatives, given by $g_x = \frac{\partial I}{\partial x}$ and $g_y = \frac{\partial I}{\partial y}$. These are normalized to obtain unit directions and scaled by the estimated gradient magnitude:

$$p(x, y) = \frac{\partial z}{\partial x} = \|\nabla z\| \cdot \frac{g_x}{\sqrt{g_x^2 + g_y^2}}, \quad q(x, y) = \frac{\partial z}{\partial y} = \|\nabla z\| \cdot \frac{g_y}{\sqrt{g_x^2 + g_y^2}} \tag{6.6}$$

The resulting (p, q) represent the surface gradient field corresponding to ridge-valley structures.

3. Depth reconstruction via Frankot–Chellappa:

The estimated gradients within the masked fingerprint area are integrated into a consistent depth map using the Frankot–Chellappa method [90], which enforces integrability in the Fourier domain. The reconstructed surface $z(x, y)$ is obtained by solving:

$$Z(\omega_x, \omega_y) = \frac{-j\omega_x P(\omega_x, \omega_y) - j\omega_y Q(\omega_x, \omega_y)}{\omega_x^2 + \omega_y^2} \tag{6.7}$$

where P and Q are the Fourier transforms of p and q restricted to the mask. The inverse transform yields the depth map as $z(x, y) = \mathcal{F}^{-1}(Z)$. This produces a smooth and physically consistent reconstruction of ridge geometry exclusively within the fingerprint region. The resulting depth map is normalized only within the fingerprint mask and represents the fine-scale ridge-valley structure of the

fingerprint.

4. Depth to Mesh Conversion

The reconstructed depth map within the masked fingerprint area is further converted into a ridge surface mesh. For each valid pixel (x, y) inside the mask, a 3D vertex is defined as $\mathbf{v}(x, y) = (x, y, z(x, y))$.

A triangulated mesh is constructed by connecting neighboring pixels that all lie inside the mask:

$$\mathcal{M}_{ridge} = (\mathcal{V}, \mathcal{F}) \tag{6.8}$$

where $\mathcal{V}$ are vertices derived from the masked depth map and $\mathcal{F}$ are triangular faces formed only over valid neighboring pixels. Optional smoothing is applied to reduce noise while preserving ridge continuity, ensuring that no noise or background is reconstructed in the final mesh.

6.3.3 Flash vs Non-Flash vs FNF Analysis of Ridge Structure

We analyze depth reconstruction under three conditions: non-flash only, flash only, and enhanced flash–minus–non-flash (FNF) which is computed like I_{diff}. Each input produces a depth map through the same ridge structure recovery pipeline, given by:

$$I \rightarrow (p, q) \rightarrow z(x, y) \rightarrow \mathcal{M}_{ridge} \tag{6.9}$$

Observations

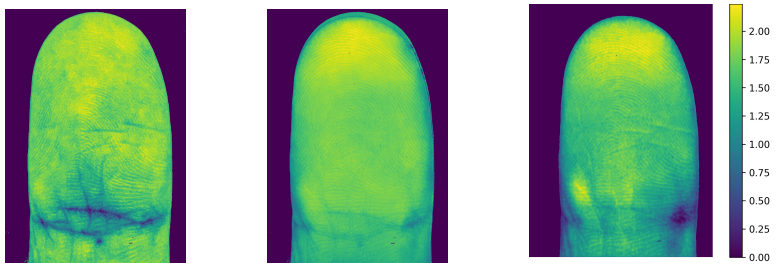
Table 6.1 summarizes the qualitative characteristics of depth reconstruction under the three illumination conditions, highlighting the differences in gradient strength, ridge clarity, and overall reconstruction quality across modalities.

Table 6.1 Summary of depth reconstruction characteristics under different illumination conditions.

Property	**Non-Flash**	**Flash**	**FNF**
Gradient Magnitude	Low	High	High, stable
Ridge Sharpness	Weak	Improved	Maximized
Ridge-Valley Separation	Weak	Moderate	Strong
Depth Smoothness	Smooth, less detailed	Sharper, may have artifacts	Detailed and stable

Depth Map Visualizations

Figure 6.5 shows representative depth maps for the three conditions, along with their corresponding ridge mesh reconstructions. The visual comparison demonstrates how illumination variations impact both the recovered depth structure and the resulting geometric representation.



(a) Non-Flash Depth Map

(b) Flash Depth Map

(c) FNF Depth Map

(d) Non-Flash Ridge Mesh

(e) Flash Ridge Mesh

(f) FNF Ridge Mesh

Figure 6.5 Top row: Depth map reconstructions for Non-Flash, Flash, and FNF conditions. Bottom row: Corresponding ridge mesh reconstructions illustrating ridge-valley geometry. FNF integration produces the most detailed and structurally clear reconstructions.

Ridge-Valley Contrast Analysis

To quantify reconstruction quality, we measure ridge-valley separability in the depth domain. Given a depth map $z(x, y)$, the values are partitioned using a percentile threshold τ into ridge and valley regions defined as $\mathcal{R} = \{z \mid z \geq \tau\}$ and $\mathcal{V} = \{z \mid z < \tau\}$. The ridge-valley contrast is then computed as:

$$C = \text{median}(\mathcal{R}) - \text{median}(\mathcal{V}) \tag{6.10}$$

Empirically, the observed trend for ridge-valley contrast follows Non-Flash $<$ Flash $<$ FNF, indicating that the combined illumination setting produces the most discriminative ridge structure. Figure 6.6 visualizes the comparative ridge-valley contrast across the three conditions.

6.3.4 Fusion of Global and Local Geometry

The final 3D fingerprint model is obtained by combining the coarse global dome geometry (M_{dome}) with the fine ridge-level surface variations (M_{ridge}), ensuring that both macroscopic finger curvature

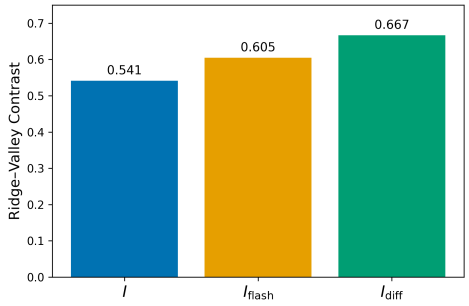


Figure 6.6 Comparison of ridge-valley contrast C for Non-Flash (I), Flash (I_{flash}), and FNF (I_{diff}) reconstructions. FNF achieves the highest discriminability.

and microscopic ridge details are preserved in a unified representation. This involves merging the dome structure with fine ridge depth meshes, aligning and scaling local and global components, and applying UV mapping [87] to project fingerprint texture onto the reconstructed surface.

1. Geometric Interpretation

Let the dome mesh be defined as $\mathcal{M}_{dome} = (\mathcal{V}_d, \mathcal{F}_d)$ and the ridge mesh reconstructed from shape-from-shading be $\mathcal{M}_{ridge} = (\mathcal{V}_r, \mathcal{F}_r)$. The goal is to embed $\mathcal{M}_{ridge}$ onto $\mathcal{M}_{dome}$ such that ridge variations follow the curvature of the finger surface.

2. Alignment and Normalization

Before fusion, the ridge geometry is normalized. Ridge heights are centered to remove global bias:

$$z_r' = z_r - \text{mean}(z_r) \tag{6.11}$$

and the ridge mesh is scaled to match the spatial extent of the dome surface. This ensures that ridge variations represent only local surface deviations rather than absolute depth.

3. Projection via Ray Casting

To align ridge vertices with the dome surface, each ridge vertex is projected onto the dome using ray casting. For each ridge vertex $\mathbf{v}_r = (x, y, z)$, a ray is defined as $\mathbf{o} = (x, y, z_{max}+\delta)$ and $\mathbf{d} = (0, 0, -1)$, where $\mathbf{o}$ is the ray origin placed above the dome and $\mathbf{d}$ is the downward direction.

$$\mathbf{p}_i = \mathbf{o} + t_i \mathbf{d} \tag{6.12}$$

where t_i is the intersection distance obtained via ray-surface intersection. Only valid intersections (i.e., rays that hit the dome) are retained for further processing.

4. Embedding Ridge Geometry onto Dome

Once intersection points $\mathbf{p}_i$ and corresponding surface normals $\mathbf{n}_i$ are obtained, the ridge geometry is embedded by displacing points along the local normal direction:

$$\mathbf{v}_i^{final} = \mathbf{p}_i + \lambda \cdot z'_r(i) \cdot \mathbf{n}_i \tag{6.13}$$

where $z'_r(i)$ is the normalized ridge height, $\mathbf{n}_i$ is the dome surface normal at the intersection point, and λ is a scaling factor controlling ridge prominence. This step ensures that ridge structures conform to the underlying dome curvature rather than remaining planar.

5. Mesh Reconstruction and Cleaning

After embedding, only vertices with valid dome intersections are retained(*highest for FNF*), triangles are reconstructed by preserving connectivity among valid vertices using a structured grid triangulation (regular image-grid triangulation with fixed diagonal splitting) inherited from the depth map, and degenerate and non-manifold elements are removed to ensure mesh integrity. The resulting mesh is defined as $\mathcal{M}_{fuse} = (\mathcal{V}_{fuse}, \mathcal{F}_{fuse})$, which represents a geometrically consistent fusion of global and local structures.

6. Texture Mapping via UV Projection

To enhance realism, the original fingerprint image is mapped onto the reconstructed surface using UV mapping. Each 3D vertex is assigned texture coordinates:

$$(u, v) = \left(\frac{x}{W}, \frac{y}{H}\right) \tag{6.14}$$

where W and H are image dimensions. This allows the 2D fingerprint texture to be projected onto the 3D geometry.

7. Final Output

The final model captures *global geometry* (finger dome curvature), *local detail* (ridge-valley microstructure), and *visual realism* (texture-mapped fingerprint appearance). This fusion produces a structurally consistent and visually meaningful 3D fingerprint representation suitable for downstream tasks such as visualization, simulation, and spoof analysis.

6.4 Results

This section presents *qualitative* results of the proposed reconstruction pipeline, highlighting the effectiveness of combining flash and non-flash information for recovering both global and local fingerprint geometry. A visual comparison of reconstructed depth maps for flash, non-flash, and FNF inputs demonstrates improved ridge-valley separation in the fused setting, as shown in Figure 6.7. The fused FNF reconstructions exhibit enhanced ridge visibility and sharper structural definition compared to in-

dividual inputs. In addition, realistic finger dome geometry is obtained from monocular reconstruction, effectively capturing the overall curvature of the finger surface. The use of gradient integration via the Frankot–Chellappa method further improves the continuity and smoothness of the reconstructed surfaces. Finally, consistent alignment between the global dome shape and local ridge structure before fusion results in a coherent and geometrically meaningful 3D fingerprint model.

(a) Aligned Ridge + Dome

(b) Fused Mesh Frontal View

(c) Fused Mesh Lateral View

Figure 6.7 Results of the 3D reconstruction pipeline: (a) Aligned ridge and dome after UV mapping, and (b–c) final fused fingerprint meshes from different views. This illustrates the completion of the low-cost 3D reconstruction process.

6.4.1 Limitations

Despite promising qualitative results, several limitations remain:

Lack of Ground Truth: Since there is no available ground-truth 3D fingerprint data for rigorous quantitative evaluation, qualitative assessment is used to make the current conclusions. However, the visual consistency and realism of the ridge patterns in the reconstructed geometry are quite high.

Capture Variability: The method requires consistent capture conditions, and changes in pose, distance, motion blur, and ambient lighting can affect reconstruction quality results significantly.

Non-Lambertian Skin Reflectance: The method is based on Lambertian assumptions for SfS, though it is known that human skin is not Lambertian due to issues with specular reflections and subsurface scattering. While applying CLAHE can improve the contrast, it does not solve the issue with inconsistent reflectance.

Limited Observations: Using only two images (flash and non-flash) restricts the robustness of photometric estimation compared to classical multi-light photometric stereo setups.

Specular Noise and Artifacts: The flash images can have strong highlights, which can cause errors during gradient estimation, thereby causing errors during local reconstruction.

Dependency on Consistent Masking: Foreground segmentation is an important task, and errors during this process can affect the quality of depth reconstruction and the mesh.

6.5 Conclusion

This chapter presents a **low-cost, computationally efficient approach for reconstructing 3D fingerprint geometry** using only flash and non-flash image pairs captured on mobile devices. Despite operating under constrained conditions, the proposed pipeline demonstrates that meaningful global shape and fine ridge structures can be recovered and fused into a good quality 3D representation. Some key implications of this work include:

Style Transfer for Spoof Materials: By applying material-specific transformations (e.g., silicone, latex, or display-based artifacts) onto reconstructed geometry, it becomes possible to simulate different spoof types and enable fully digital spoof analysis.

Low-Cost Spoof Recreation: Spoof generation may also be carried out entirely in software without physical fabrication of spoofing artifacts, reducing costs and effort.

Understanding Spoof Characteristics: Digital reconstruction allows experimentation on how different materials affect ridge structure, reflectance, and depth information.

Moreover, the reconstructed models in 3D space can be utilized to carry out data augmentation for the development of effective recognition and PAD systems, to improve the PAD systems by adding depth-based features, and to narrow the gap between contact-based and contactless fingerprint systems by utilizing geometric information in matching.

6.5.1 Future work

Further, we can focus on the incorporation of learning-based photometric stereo approaches for better handling of non-Lambertian skin reflectance, improving robustness for unconstrained capture, as well as extending this work for real-time mobile deployment, and finally, the incorporation of neural rendering approaches for increased realism.

Overall, this work shows that 3D fingerprint reconstruction can be accomplished with minimal hardware, low computational cost, and simple acquisition schemes. The ability to digitally manipulate the geometry of the fingerprint can lead to many avenues for biometric security, especially for spoof detection and simulation. Code is available in the repository[1].

[1] Low-cost 3D Reconstruction Source Code

Chapter 7

Conclusion

The thesis aims to provide a unified framework for improving contactless fingerprint recognition and detection of spoofs through the effective utilization of paired flash–non-flash images. The main idea is that instead of being considered a noise source, variations in illumination are actually useful information for understanding the ridge structure, material, and photometric properties of fingerprints. By treating paired acquisition as a lightweight active sensing mechanism, this work aims to provide a novel integrated pipeline for image enhancement, recognition, matching, and security within a single framework.

7.1 Key Contributions

The Flash–Non-Flash Fingerphoto (FNF) Database is a significant contribution that facilitates the analysis of illumination impact for contactless fingerprint imaging. Comprising 3,140 captured and 12,560 processed paired images from 79 individuals, it provides a controlled benchmark for studying complementary information across illumination conditions and supports reproducible evaluation of proposed methods.

Fusion2Print (F2P) proposes an end-to-end deep learning approach for enhancing and embedding fingerprints by developing optimal fusion strategies across spatial, frequency, and color domains. It demonstrates consistent improvements over both minutiae-based and learning-based matching approaches with significant improvements in texture-aware representations and matching across domains of contact and contactless fingerprints, addressing a critical interoperability challenge.

The illumination-aware presentation attack detection framework exploits physics-informed features such as inter-channel correlation, specular reflection, texture realism, differential imaging, and material characteristics to differentiate between genuine fingerprints and spoof samples. These cues show that the photometric response for spoof samples varies across illumination conditions, thus creating a strong interpretable foundation for spoof detection.

The proposed low-cost 3D reconstruction approach shows that it is possible to recover meaningful surface geometry with just two images without the need for any specialized hardware. The approach uses monocular reconstruction for global shapes and shape-from-shading(SfS) for ridge details. This

allows coherent 3D models to be created that are useful for digital spoof simulation and material-aware data generation, reducing the need for extensive physical spoof datasets.

Impact

Integrating enhancement, matching, and spoof detection reduces redundancy in acquisition and improves usability in real-world scenarios. The illumination-aware cues improve the interpretability of the results, allowing for a better analysis of the system's behavior and failure modes. Using photometric properties dependent on materials creates an additional, strong layer of security against advanced spoofing attacks, as attackers find it hard to ensure consistency across multiple illumination-dependent cues. A smartphone-based solution offers tremendous flexibility without relying on specialized hardware, allowing the solution to be deployed in border control, law enforcement, financial services, and smart infrastructure, thus broadening the use cases while maintaining high security and usability.

7.2 Limitations and Future Work

The FNF database, although varied, can be extended even further to include a wider range of devices, which would help in capturing varied sensor properties. Moreover, there might exist some issues related to the temporal misalignment of flash and non-flash images, which would require improved acquisition methodologies. The existing spoof detection framework relies majorly on illumination-based features. The inclusion of texture, temporal, and material features might help in achieving improved robustness. In addition, the PAD methodology must adapt to the ever-evolving spoofing techniques, and broader evaluation across devices and environments is necessary to strengthen generalization. The overall reconstruction methodology relies on the Lambertian assumption in the context of SfS, which might not accurately represent the non-Lambertian properties of human skin. The inclusion of learning-based photometric stereo for improved reflectance modeling could achieve improved results. The overall performance under extreme pose, motion blur, and dynamic illumination must be explored as well. The evaluation methodology is also restricted by the lack of ground-truth 3D fingerprint data, which would require the use of high-fidelity scanning.

Future directions include extending it to multi-illumination settings other than flash/non-flash, developing fully end-to-end joint optimization of enhancement, embedding, and spoof detection, incorporating video-based acquisition for robustness, exploring cross-spectral modalities like thermal imaging, and developing mechanisms for privacy-preserving biometric data handling.

7.3 Summary

This thesis proposes a comprehensive framework for contactless fingerprint recognition, where the change in illumination is used as a source of discriminative information. The flash–non-flash paradigm moves the community from single-image processing to paired multi-illumination sensing, which helps

improve recognition, spoofing, and 3D modeling. The proposed framework is validated using classical and learning-based recognition systems, which proves the robustness and generality of the approach. The approach is practical, deployable, and solves many problems faced by contactless biometric systems. As contactless biometrics continue to expand in response to hygiene, accessibility, and mobile imaging advancements, this work provides a foundation for building secure, scalable, and user-friendly systems, enabling the deployment of reliable mobile fingerprint-based authentication in diverse applications.

Bibliography

[1] Y. Yu, Q. Niu, X. Li, J. Xue, W. Liu, and D. Lin, "A review of fingerprint sensors: Mechanism, characteristics, and applications," *Micromachines (Basel)*, vol. 14, no. 6, p. 1253, Jun. 2023.

[2] R. Donida Labati, A. Genovese, V. Piuri, and F. Scotti, "A scheme for fingerphoto recognition in smartphones," in *Selfie Biometrics*, ser. Advances in computer vision and pattern recognition. Cham: Springer International Publishing, 2019, pp. 49–66.

[3] Y. K. G, S. Bhattacharya, and N. Aishwarya, "Remote photoplethysmography (rPPG) for contactless blood oxygen saturation monitoring." IEEE, Jul. 2024, pp. 1–6.

[4] A. M. M. Chowdhury and M. H. Imtiaz, "Contactless fingerprint recognition using deep learning–a systematic review," *Journal of Cybersecurity and Privacy*, vol. 2, no. 3, pp. 714–730, 2022. [Online]. Available: https://www.mdpi.com/2624-800X/2/3/36

[5] M. Derawi, B. Yang, and C. Busch, "Fingerprint recognition with embedded cameras on mobile phones," vol. 94, 01 2012, pp. 136–147.

[6] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*, 2nd ed. Springer, 2009.

[7] A. K. Jain, A. A. Ross, and K. Nandakumar, *Fingerprint Recognition*. Boston, MA: Springer US, 2011, pp. 51–96. [Online]. Available: https://doi.org/10.1007/978-0-387-77326-1_2

[8] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, "Synthetic fingerprint generation," in *Handbook of Fingerprint Recognition*. London, UK: Springer, 2009, pp. 271–302.

[9] H. Choi, K.-S. Choi, and J. Kim, "Mosaicing touchless and mirror-reflected fingerprint images," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 52–61, 2010.

[10] S. Sreehari and S. M. Anzar, "Touchless fingerprint recognition: A survey of recent developments and challenges," *Comput. Electr. Eng.*, vol. 122, no. 109894, p. 109894, Mar. 2025.

[11] Y. Song, C. Lee, and J. Kim, "A new scheme for touchless fingerprint recognition system," in *Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Seoul, Korea, 2004.

[12] A. Kumar, "Introduction to trends in fingerprint identification," in *Contactless 3D Fingerprint Identification*. Berlin/Heidelberg, Germany: Springer, 2018, pp. 1–15.

[13] Y. Wang, L. G. Hassebrook, and D. L. Lau, "Data acquisition and processing of 3-d fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 5, pp. 750–760, 2010.

[14] B. Stanton, B. Stanton, M. Theofanos, S. Furman, and P. Grother, "Usability testing of a contactless fingerprint device: Part 2," US Department of Commerce, National Institute of Standards and Technology, Gaithersburg, MD, USA, Tech. Rep., 2016.

[15] A. Kumar and Y. Zhou, "Contactless fingerprint identification using level zero features," in *Proceedings of the IEEE CVPR 2011 Workshops*, Colorado Springs, CO, USA.

[16] C. Lee, S. Lee, and J. Kim, "A study of touchless fingerprint recognition system," in *Structural, Syntactic, and Statistical Pattern Recognition*, D.-Y. Yeung, J. T. Kwok, A. Fred, F. Roli, and D. de Ridder, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 358–365.

[17] G. Parziale and Y. Chen, "Advanced technologies for touchless fingerprint recognition," in *Handbook of Remote Biometrics*. London, UK: Springer, 2009, pp. 83–109.

[18] J. Priesnitz, C. Rathgeb, N. Buchmann, C. Busch, and M. Margraf, "An overview of touchless 2D fingerprint recognition," *EURASIP J. Image Video Process.*, vol. 2021, no. 1, Dec. 2021.

[19] A. Malhotra, A. Sankaran, M. Vatsa, and R. Singh, "On matching finger-selfies using deep scattering networks," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 4, pp. 350–362, 2020.

[20] S. Chopra, A. Malhotra, M. Vatsa, and R. Singh, "Unconstrained fingerphoto database," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Jun. 2018.

[21] B. Jawade, D. D. Mohan, S. Setlur, N. Ratha, and V. Govindaraju, "Ridgebase: A cross-sensor multi-finger contactless fingerprint dataset," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, Oct. 2022, pp. 1–9.

[22] C.-T. Lin and A. Kumar, "Matching contactless and contact-based conventional fingerprint images for biometrics identification," *IEEE Transactions on Image Processing*, vol. 27, pp. 2008–2021, 2018.

[23] J. Kolberg, J. Priesnitz, C. Rathgeb, and C. Busch, "Colfispoof: A new database for contactless fingerprint presentation attack detection research," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2023, pp. 653–661.

[24] A. Taneja, A. Tayal, A. Malhorta, A. Sankaran, M. Vatsa, and R. Singh, "Fingerphoto spoofing in mobile devices: A preliminary study," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–7.

[25] S. Purnapatra, C. Miller-Lynch, S. Miner, Y. Liu, K. Bahmani, S. Dey, and S. Schuckers, "Presentation attack detection with advanced cnn models for noncontact-based fingerprint systems," 2023. [Online]. Available: https://arxiv.org/abs/2303.05459

[26] W. Chen and Y. Gao, "A minutiae-based fingerprint matching algorithm using phase correlation," in *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*. IEEE, Dec. 2007.

[27] J, Ravi., K. B. Raja, and R, Venugopal. K., "Fingerprint recognition using minutia score matching," 2010.

[28] Kumaran, O. Agrawal, J. Paul, and Rajesh, "Contactless fingerprint recognition using SIFT enhanced minutiae matching," in *2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*. IEEE, May 2025, pp. 1–8.

[29] Erwin, N. N. B. Karo, A. Yulia Sari, N. Aziza, and H. Kesuma Putra, "The enhancement of fingerprint images using gabor filter," *J. Phys. Conf. Ser.*, vol. 1196, p. 012045, Mar. 2019.

[30] S. Bakheet, S. Alsubai, A. Alqahtani, and A. Binbusayyis, "Robust fingerprint minutiae extraction and matching based on improved SIFT features," *Appl. Sci. (Basel)*, vol. 12, no. 12, p. 6122, Jun. 2022.

[31] N. Giansiracusa, R. Giansiracusa, and C. Moon, "Persistent homology machine learning for fingerprint classification," 2017. [Online]. Available: https://arxiv.org/abs/1711.09158

[32] P. D. Devkar, N. Ajwani, S. Malla, A. Bhatt, M. Dave, N. Sharma, M. Panday, and C. S. Yajnik, "Topological analysis of dermatoglyphic patterns using synthetic and real data," in *Proceedings of the 26th International Conference on Distributed Computing and Networking*, ser. ICDCN '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 295–301. [Online]. Available: https://doi.org/10.1145/3700838.3703694

[33] H. Kekre and V. Bharadi, "Fingerprint's core point detection using orientation field," in *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, 2009, pp. 150–152.

[34] G. Bahgat, A. Khalil, N. Abdel Kader, and S. Mashali, "Fast and accurate algorithm for core point detection in fingerprint images," *Egyptian Informatics Journal*, vol. 14, no. 1, pp. 15–25, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110866513000030

[35] J. Priesnitz, R. Huesmann, C. Rathgeb, N. Buchmann, and C. Busch, "Mobile contactless fingerprint recognition: Implementation, performance and usability aspects," *Sensors (Basel)*, vol. 22, no. 3, p. 792, Jan. 2022.

[36] R. Raghavendra, C. Busch, and B. Yang, "Scaling-robust fingerprint verification with smartphone camera in real-life scenarios," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. IEEE, Sep. 2013.

[37] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh, "On smartphone camera based fingerphoto authentication," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, Sep. 2015.

[38] G. Abramovich, M. Ganesh, K. Harding, S. Manickam, J. Czechowski, X. Wang, and A. Vemury, "A spoof detection method for contactless fingerprint collection utilizing spectrum and polarization diversity," in *Proceedings of SPIE: Next-Generation Spectroscopic Technologies III*, vol. 7680, April 2010, p. 768005.

[39] L. Leng, M. Li, C. Kim, and X. Bi, "Dual-source discrimination power analysis for multi-instance contactless palmprint recognition," *Multimedia Tools Appl.*, vol. 76, no. 1, p. 333–354, Jan. 2017. [Online]. Available: https://doi.org/10.1007/s11042-015-3058-7

[40] Y. Wu, L. Leng, and H. Mao, "Non-contact palmprint attendance system on pc platform," *Journal of Multimedia Information System*, vol. 5, no. 3, pp. 179–188, 2018. [Online]. Available: https://doi.org/10.9717/JMIS.2018.5.3.179

[41] S. A. Grosz, J. J. Engelsma, E. Liu, and A. K. Jain, "C2CL: Contact to contactless fingerprint matching," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 196–210, 2022.

[42] P. Kaplesh, A. Gupta, D. Bansal, S. Sofat, and A. Mittal, "Vision transformer for contactless fingerprint classification," *Multimed. Tools Appl.*, vol. 84, no. 26, pp. 31 239–31 259, Nov. 2024.

[43] H. Tan and A. Kumar, "Towards more accurate contactless fingerprint minutiae extraction and pose-invariant matching," *IEEE Trans. Inf. Forensics Secur.*, pp. 1–1, 2020.

[44] Y. Shi, Z. Zhang, S. Liu, and M. Liu, "Towards more accurate matching of contactless fingerprints with a deep geometric graph convolutional network," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5, no. 1, pp. 29–38, 2023.

[45] C. Lin and A. Kumar, "Multi-siamese networks to accurately match contactless to contact-based fingerprint images," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 277–285.

[46] H. Tan and A. Kumar, "Minutiae attention network with reciprocal distance loss for contactless to contact-based fingerprint identification," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3299–3311, 2021.

[47] S. Peddi, S. Bandyopadhyay, and D. Samanta, "G-msginet: A grouped multi-scale graph-involution network for contactless fingerprint recognition," 2025. [Online]. Available: https://arxiv.org/abs/2505.08233

[48] P. Wasnik, R. Ramachandra, K. Raja, and C. Busch, "Presentation Attack Detection for Smartphone Based Fingerphoto Recognition Using Second Order Local Structures," in *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2018, pp. 241–246.

[49] E. Marasco and A. Vurity, "Late deep fusion of color spaces to enhance finger photo presentation attack detection in smartphones," *Applied Sciences*, vol. 12, no. 22, p. 11409, 2021. [Online]. Available: https://doi.org/10.3390/app122211409

[50] B. Adami, S. Tehranipoor, N. Nasrabadi, and N. Karimian, "A universal anti-spoofing approach for contactless fingerprint biometric systems," 2023. [Online]. Available: https://arxiv.org/abs/2310.15044

[51] B. Adami and N. Karimian, "Contactless fingerprint biometric anti-spoofing: An unsupervised deep learning approach," 2023. [Online]. Available: https://arxiv.org/abs/2311.04148

[52] ——, "Gru-aunet: A domain adaptation framework for contactless fingerprint presentation attack detection," 2025. [Online]. Available: https://arxiv.org/abs/2504.01213

[53] J. J. Engelsma, K. Cao, and A. K. Jain, "Raspireader: An open source fingerprint reader facilitating spoof detection," 2017. [Online]. Available: https://arxiv.org/abs/1708.07887

[54] C. Dong and A. Kumar, "Bridging dimensions in fingerprints to advance distinctiveness: Recovering 3d minutiae from a single contactless 2d fingerprint image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 9, pp. 7812–7831, 2025.

[55] F. Liu and D. Zhang, "3d fingerprint reconstruction system using feature correspondences and prior estimated finger model," *Pattern Recognition*, vol. 47, no. 1, pp. 178–193, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320313002616

[56] C. Lin and A. Kumar, "Tetrahedron based fast 3d fingerprint identification using colored leds illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 3022–3033, 2018.

[57] Q. Miao, H. Wang, H. Sun, and Y. Zhang, "Print2volume: Generating synthetic oct-based 3d fingerprint volume from 2d fingerprint image," 2025. [Online]. Available: https://arxiv.org/abs/2508.21371

[58] J. Galbally, G. Bostrom, and L. Beslay, "Full 3D touchless fingerprint recognition: Sensor, database and baseline performance," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, Oct. 2017.

[59] L. Mescheder, W. Dong, S. Li, X. Bai, M. Santos, P. Hu, B. Lecouat, M. Zhen, A. Delaunoy, T. Fang, Y. Tsin, S. R. Richter, and V. Koltun, "Sharp monocular view synthesis in less than a second," 2026. [Online]. Available: https://arxiv.org/abs/2512.10685

[60] N. Zaeri, "Minutiae-based fingerprint extraction and recognition," in *Biometrics*. InTech, 2011.

[61] S. Singh, D. K. Nishad, and S. Khalid, "Enhancing fingerprint identification using fuzzy-ann minutiae matching," *Measurement: Sensors*, vol. 37, p. 101809, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2665917425000030

[62] "Camera — expo documentation," https://docs.expo.dev/versions/latest/sdk/camera/.

[63] M. Taylor, R. A. Hicklin, and G. Kiebuzinski, "Best practices in the collection and use of biometric and forensic datasets," National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep. NIST Interagency/Internal Report 8361, 2021.

[64] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness." IEEE Comput. Soc. Press, 2002.

[65] T. Li and J. Chen, "A fingerprint image enhancement method based on gabor filter." IEEE, Jul. 2011.

[66] Neurotechnology, "Verifinger sdk 2025.1," https://www.neurotechnology.com/verifinger.html, 2025.

[67] J. J. Engelsma, K. Cao, and A. K. Jain, "Learning a fixed-length fingerprint representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 1981–1997, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202718922

[68] T. Rohwedder, "Fixed-length fingerprint extractors," https://github.com/tim-rohwedder/fixed-length-fingerprint-extractors, 2021, accessed: September 2025.

[69] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[70] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015. [Online]. Available: https://arxiv.org/abs/1505.04597

[71] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, p. 94–114, Apr. 2020. [Online]. Available: http://dx.doi.org/10.1016/j.isprsjprs.2020.01.013

[72] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016, pp. 770–778.

[73] O. Bausinger and E. Tabassi, "Fingerprint sample quality metric nfiq 2.0." 01 2011, pp. 167–171.

[74] S. Patel, "Fingerprint verification system (fvs)," https://fvs.sourceforge.net/, 2002.

[75] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," 2017. [Online]. Available: https://arxiv.org/abs/1703.06868

[76] E. Lim, X. Jiang, and W. Yau, "Fingerprint quality and validity analysis," in *Proceedings. International Conference on Image Processing*, vol. 1, 2002, pp. I–I.

[77] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun, "A comparative study of fingerprint image-quality estimation methods," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4. [Online]. Available: http://dx.doi.org/10.1109/TIFS.2007.908228

[78] C. Kauba, D. Söllinger, S. Kirchgasser, A. Weissenfeld, G. Fernández Domínguez, B. Strobl, and A. Uhl, "Towards using police officers' business smartphones for contactless fingerprint acquisition and enabling fingerprint comparison against contact-based datasets," *Sensors*, vol. 21, no. 7, 2021.

[79] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, M. Assran, N. Ballas, W. Galuba, R. Howes, P.-Y. Huang, S.-W. Li, I. Misra, M. Rabbat, V. Sharma, G. Synnaeve, H. Xu, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski, "Dinov2: Learning robust visual features without supervision," 2024.

[80] D. Niu, R. Guo, and Y. Wang, "Moiré attack (ma): A new potential risk of screen photos," *arXiv preprint arXiv:2110.10444*, 2021.

[81] Z. Sedaghatjoo, H. Hosseinzadeh, and B. S. Bigham, "Local binary pattern(lbp) optimization for feature extraction," 2024. [Online]. Available: https://arxiv.org/abs/2407.18665

[82] A. R. Zubair and O. A. Alo, "Grey level co-occurrence matrix (glcm) based second order statistics for image texture analysis," 2024. [Online]. Available: https://arxiv.org/abs/2403.04038

[83] R. R. Anderson and J. A. Parrish, "The optics of human skin," *Journal of Investigative Dermatology*, vol. 77, no. 1, pp. 13–19, 1981.

[84] S. Marschner, E. Lafortune, S. Westin, K. Torrance, and D. Greenberg, "Image-based brdf measurement," 08 1999.

[85] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," in *ACM SIGGRAPH 2023 Conference Proceedings*. ACM, 2023.

[86] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," 2020. [Online]. Available: https://arxiv.org/abs/2003.08934

[87] J. F. Blinn and M. E. Newell, "Texture and reflection in computer generated images," *Communications of the ACM*, vol. 19, no. 10, pp. 542–547, 1976.

[88] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson Surface Reconstruction," in *Symposium on Geometry Processing*, A. Sheffer and K. Polthier, Eds. The Eurographics Association, 2006.

[89] A. Srivastava and A. Namboodiri, "Split and knit: 3d fingerprint capture with a single camera," in *Proceedings of the Thirteenth Indian Conference on Computer Vision, Graphics and Image Processing*, ser. ICVGIP '22. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3571600.3571618

[90] R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 439–451, 1988.