Bit Stability and Hash-Length Trade-Offs in Binary Face Templates

Abdelilah Ganmati, Karim Afdel, Lahcen Koutti
Computer Systems and Vision Laboratory-Faculty of Sciences, Ibn Zohr University, Agadir, Morocco

***Abstract*—Binary face templates are an appealing alternative to floating-point embeddings for face verification in resource-constrained environments because they enable constant-time Hamming matching with minimal storage and input/output (I/O). This paper studies the bit-level behavior of hashes obtained by principal component analysis followed by iterative quantization (PCA–ITQ) at $L \in \{32, 64, 128\}$ derived from a frozen lightweight face encoder. Using subject-disjoint splits on the MORPH longitudinal dataset and an eyeglasses stress protocol on CelebA, the analysis quantifies 1) bit balance and entropy, 2) within-identity bit stability via per-bit flip rates, and 3) verification performance at low false-accept rates in Hamming space. On MORPH, 64-bit PCA–ITQ codes achieve an area under the receiver operating characteristic curve (AUC) of 0.9978 and a true positive rate (TPR) of 96.5% at a false positive rate (FPR) of 1%, compared to 99.1% at 128 bits, while halving the template length; 32-bit codes remain feasible but drop to 85.7% at the same operating point and are more sensitive to nuisance variation. Across both datasets, codes are near-balanced and mostly stable, yet a small minority of bit positions accounts for most flips under the eyeglasses attribute. In this regime, 64-bit hashes offer a favorable size–accuracy trade-off, whereas 128-bit hashes approximate float-embedding behavior and 32-bit hashes require redundancy or additional robustness mechanisms. All evaluations use fixed seeds and subject-disjoint splits; thresholds are selected on validation and held fixed on test to reflect deployment conditions.**

***Keywords*—Binary face templates; face verification; hamming distance; PCA-ITQ; bit stability; eyeglasses attribute; resource-constrained verification; MORPH; CelebA**

I. Introduction

Compact binary face templates are attractive substitutes for floating-point embeddings because they enable deterministic Hamming matching with minimal storage and bandwidth. A widely used route—principal component analysis followed by an orthogonal rotation and element-wise sign (PCA–ITQ)—yields short codes while keeping the mapping deterministic and easy to deploy [1]. Yet shortening the code (e.g., to 32–128 bits) raises questions that are under-reported in biometric verification: *How balanced and informative are individual bit positions?* and *which bits are stable within an identity under common nuisance factors?*

Most hashing work emphasizes retrieval metrics or ranking quality [2], [3], [4], [5]; however, privacy-preserving verification is ultimately governed by bit-level behavior—entropy, balance, and within-identity flip-rates—which determines feasible operating thresholds. Furthermore, the renewability/diversity requirements for protected templates [6], [7] benefit from knowing *where* instability concentrates so that mitigation (e.g., masking) can be targeted rather than uniform.

Problem setting and datasets. Using the MORPH longitudinal database as the primary benchmark and CelebA for attribute-conditioned stress tests (eyeglasses) [9], [10], this study examines binary templates produced from a single frozen mobile-class encoder after standard five-point alignment and PCA–ITQ binarization to $L \in \{32, 64, 128\}$. The analysis covers: 1) distributional properties of individual bits (balance, entropy) as a function of L; 2) within-identity flip-rates; and 3) localization of instability under attribute change (eyeglasses) and its impact on ROC-level operating points in Hamming space.

Novelty relative to prior PCA–ITQ and hashing work. Previous work on PCA–ITQ and deep hashing has mostly reported aggregate retrieval or verification metrics at the code level, such as mean average precision, AUC, or Equal Error Rate (EER) [1], [2], [3], [4], [5]. In contrast, the present study is explicitly designed as a *bit-level* analysis of binary face templates under resource constraints. The key distinctive elements are:

- a systematic characterization of per-bit balance, entropy, and within-identity stability for $32/64/128$-bit PCA–ITQ templates, with the upstream encoder and alignment pipeline held fixed;
- an attribute-localized stress protocol on CelebA that isolates the effect of eyeglasses on per-bit flip-rates and relates the induced instability to low-FAR operating points;
- an operational mapping from observed bit statistics to hash-length choices and threshold selection for privacy-preserving verification in smart-card or embedded settings.

These aspects complement earlier work on mobile-class CNNs and binary hashes for smart-card-constrained authentication [11], ageing drift in binary templates [15], and targeted parity mechanisms for unstable bits [18], by focusing specifically on how bit-level statistics shape feasible operating points for short hashes.

The main contributions can be summarized as follows:

- Bit-level characterization across code lengths. Bit balance, entropy, and within-identity stability are quantified for $L \in \{32, 64, 128\}$ and related to verification performance on MORPH, revealing a favorable size–accuracy frontier at 64 bits.
- Attribute-localized stress test. An eyeglasses protocol on CelebA shows that degradation is concentrated in

a minority of bit positions, rather than uniformly distributed, which supports targeted mitigation strategies.

- Operational guidance grounded in measurements. The measured flip-rate profiles are mapped to threshold selection at low false-accept rates and to practical recommendations on hash length, enrollment redundancy, and template renewability in privacy-preserving systems, in line with ISO/IEC 24745 [6].

II. Related Work

1) Alignment and deep embeddings: Modern face verification maps aligned crops to ℓ_2–normalized embeddings via frozen encoders; robust alignment is a prerequisite to stabilize downstream representations. The experiments adopt the widely used Multitask Cascaded Convolutional Networks (MTCNN) detector with five-point similarity alignment to produce canonicalized inputs [8].

2) Hashing and binary codes: Hashing for visual descriptors spans randomized families such as locality-sensitive hashing (LSH) and SimHash [2], [12], spectral/graph constructions [3], and deep supervised variants (e.g., DPSH, HashNet) [4], [5], with broader overviews in recent surveys [19]. For deployment-constrained verification, deterministic and lightweight transforms are attractive. PCA followed by Iterative Quantization (ITQ) [1] minimizes quantization error under an orthogonal rotation prior to sign binarization, yielding short codes amenable to constant-time Hamming scoring. Beyond classical PCA–ITQ, hybrid quantum PCA methods have been explored to accelerate face verification pipelines [21], and fairness-aware binary descriptors have been proposed for face presentation attack detection using local binary patterns [22]. This paper uses PCA–ITQ specifically to study how *code length* ($L \in \{32, 64, 128\}$) interacts with *per-bit* balance and stability when the upstream encoder is held fixed. The analysis also builds on a prior benchmark of PCA–ITQ binary templates under embedded constraints by the same authors, which established protocol and topline baselines; here the focus is shifted to per-bit balance/stability and attribute-localized robustness [11].

3) Multimodal biometric fusion: Convolutional neural network (CNN)-driven multimodal fusion of face, fingerprint [20], and voice has become increasingly common to enhance robustness under occlusion or sensor variability [23], [24], [25]. Although the present study remains unimodal (face only), the findings on bit stability and attribute-specific drift help clarify how individual face templates might contribute within a larger fusion framework.

4) Protected biometric templates: Template-protection standards emphasize irreversibility, renewability, and unlinkability [6]. Cancelable-biometrics schemes implement parameterized transforms so compromised templates can be reissued without reacquisition [7]. Template interfaces are often kept conservative to limit oracle leakage, and renewability/diversity are handled via cancelable transforms [6], [7]. These works motivate compact binary representations and conservative interfaces, but typically report aggregate ROC/EER at the *code level* rather than analyzing *statistics of individual bit positions*. Recent system-level studies on deep learning-based multifactor authentication and ISO/IEC-compliant match-on-card face verification further illustrate how compact templates and conservative interfaces can be combined in practice [17], [16].

5) Bit-level behavior: Despite the prevalence of binary hashing, systematic measurements of *bit balance*, *within-identity flip rates*, and *attribute-localized instability* (e.g., eyeglasses) remain scarce in the above literatures [2], [3], [4], [5], [6], [7]. Our study targets this gap by quantifying per-bit statistics across $L \in \{32, 64, 128\}$ and relating them directly to operating-point selection in Hamming space.

6) Datasets and protocol: Following established practice, we use *MORPH* as the primary benchmark for verification under longitudinal variability [9] and *CelebA* for attribute-conditioned stress tests (eyeglasses) [10], with a unified MTCNN alignment pipeline [8]. This setup isolates bit-level phenomena independently of classifier tuning.

7) Positioning: Relative to prior art, we keep the encoder and binarizer fixed (single frozen encoder + PCA–ITQ) and move the analytical lens from aggregate ROC summaries to *per-bit* statistics and their operational consequences—precisely the level that governs hash-length choice, threshold setting, and renewal/diversity policies in protected template deployments [6], [7].

III. Methods

A. Data and Splits

1) MORPH (primary): Subject–disjoint splits are formed with identities reserved for threshold selection (validation) and final reporting (test) on longitudinal images [9]. No identity appears in more than one partition.

2) CelebA (eyeglasses stress): Using the official attribute annotations [10], the protocol constructs per–identity pairs that differ only in the `Eyeglasses` attribute (present vs. absent). CelebA identities used for stress testing are disjoint from MORPH identities.

B. Preprocessing and Embeddings

Aligned face crops are produced via standard five–landmark similarity alignment. Aligned images are passed through a *frozen* encoder to obtain ℓ_2–normalized float embeddings. No fine–tuning or test-time augmentation is used. This protocol fixes upstream variability so that downstream analyses isolate bit–level phenomena. The alignment/embedding/hashing setup follows a previously reported benchmark by the same authors and is kept frozen here [11].

C. Binarization: PCA–ITQ

Given an encoder output $\mathbf{f} \in \mathbb{R}^{d_{\text{enc}}}$, we compute

$$\mathbf{x} = (\mathbf{f} - \boldsymbol{\mu}) W_{\text{PCA}} \in \mathbb{R}^L,$$

$$\mathbf{z} = \mathbf{x} R,$$

$$\mathbf{b} = \operatorname{sign}(\mathbf{z}) \in \{-1, +1\}^L, \qquad L \in \{32, 64, 128\},$$

where W_{PCA} retains the top-L principal components fitted on *training* identities only, and R is the orthogonal rotation learned by Iterative Quantization (ITQ) to minimize $\|\mathbf{B} - \mathbf{X}R\|_F^2$ subject to $R^\top R = \mathbf{I}$ with $\mathbf{B} \in \{-1,+1\}^{N\times L}$ [1]. We map $\{-1,+1\} \to \{0,1\}$ and pack bits MSB–first. PCA fitting and ITQ initialization use fixed random seeds.

D. Mathematical Formulation and Decision Metrics

1) Hamming scoring and decision: Let $\mathbf{b}_{\mathrm{enr}}, \mathbf{b}_{\mathrm{prb}} \in \{0,1\}^L$ be enrolled and probe codes. The Hamming distance is

$$D_H(\mathbf{b}_{\mathrm{enr}}, \mathbf{b}_{\mathrm{prb}}) = \sum_{k=1}^{L} \big[b_{\mathrm{enr},k} \oplus b_{\mathrm{prb},k}\big]. \tag{1}$$

A verifier with threshold τ accepts iff $D_H \le \tau$.

a) Operating characteristics: Let $\mathcal{P}^+$ and $\mathcal{P}^-$ be sets of genuine and impostor pairs. Then

$$\mathrm{TPR}(\tau) = \Pr\big[D_H \le \tau \,\big|\, \mathcal{P}^+\big], \tag{2}$$
$$\mathrm{FPR}(\tau) = \Pr\big[D_H \le \tau \,\big|\, \mathcal{P}^-\big], \tag{3}$$
$$\mathrm{FNR}(\tau) = 1 - \mathrm{TPR}(\tau). \tag{4}$$

The receiver operating characteristic (ROC) traces $(\mathrm{FPR}(\tau), \mathrm{TPR}(\tau))$; the Equal Error Rate (EER) satisfies $\mathrm{FPR}(\tau^\star) \approx \mathrm{FNR}(\tau^\star)$. We also report $\mathrm{TPR@FPR} = \alpha$ (e.g., $\alpha = 1\%$).

b) Bit balance and entropy: For bit k,

$$m_k = \frac{1}{N}\sum_{n=1}^{N} b_{nk}, \qquad p_k = \tfrac{1+m_k}{2}, \tag{5}$$
$$H_k = -p_k \log_2 p_k \;-\; (1-p_k)\log_2(1-p_k). \tag{6}$$

Well-behaved codes have $m_k \approx 0$ and $H_k \approx 1$ bit.

c) Within–identity stability and eyeglasses sensitivity: For same-identity pairs $\mathcal{S}$, the per-bit flip-rate is

$$\phi_k = \frac{1}{|\mathcal{S}|}\sum_{(i,j)\in\mathcal{S}} \mathbb{1}\big[b_{ik} \neq b_{jk}\big]. \tag{7}$$

On CelebA, compute $\phi_k^{(\mathrm{glass})}$ on cross-condition pairs (with vs. without eyeglasses) and $\phi_k^{(\mathrm{ctrl})}$ on same-condition pairs. The difference $\Delta\phi_k = \phi_k^{(\mathrm{glass})} - \phi_k^{(\mathrm{ctrl})}$ localizes eyeglasses-induced instability.

E. Eyeglasses Stress Protocol (CelebA)

For identities with both conditions, cross–condition pairs are formed to isolate attribute-induced instability and same–condition controls are used to factor out identity/time variance. We compare the distributions of $\{\phi_k^{(\mathrm{glass})}\}$ and $\{\phi_k^{(\mathrm{ctrl})}\}$ and quantify their impact on ROC/EER at each L.

Input: MORPH (train/val/test), CelebA (eyeglasses splits); $L \in \{32, 64, 128\}$; optional instability quota q (masking)
Output: Bit statistics and verification metrics (ROC/AUC, EER, TPR@FPR=1%)
Initialize $\boldsymbol{\mu}$, W_{PCA} (top L), ITQ rotation R (fixed seeds)
Fit PCA–ITQ on MORPH train; learn R by ITQ [1]
for each $L \in \{32, 64, 128\}$ **do**
  Validation (MORPH): binarize; sweep τ; record $\tau^\star$ (AUC/EER)
  Bit statistics (MORPH): compute m_k, H_k, and flip-rate ϕ_k
  Eyeglasses (CelebA): cross- vs same-condition; compute $\phi_k^{(\mathrm{glass})}$, $\phi_k^{(\mathrm{ctrl})}$, $\Delta\phi_k$
  if masking enabled **then**
    rank by instability; define mask $\mathcal{M}$ (top-q%); re-estimate $\tau^\star_{\mathrm{mask}}$ on $\overline{\mathcal{M}}$
  end if
  Test (MORPH): evaluate baseline (and masked) with frozen thresholds
  Report (CelebA): summarize $\{\phi_k^{(\mathrm{glass})}\}$ vs $\{\phi_k^{(\mathrm{ctrl})}\}$ and relate to MORPH operating points
end for

Fig. 1. Protocol for bit-level analysis and verification.

TABLE I. MORPH: EFFECT OF BIT LENGTH L (PCA-ITQ)

L (bits)	AUC	EER (%)	TPR@1% (%)	TPR@0.1% (%)
128	0.9994	0.90	99.13	96.64
64	0.9978	2.04	96.52	88.96
32	0.9887	5.17	85.68	49.02

F. Scoring and Operating Points

For each $L \in \{32, 64, 128\}$ a single threshold τ is set on the MORPH validation ROC and then applied unchanged to the MORPH test split. CelebA is used to report bit-level stability and to assess how eyeglasses-conditioned instability correlates with shifts in operating points (thresholds remain those fixed on MORPH validation).

G. Reproducibility and Privacy Context

All random choices (PCA/ITQ seeds, pair sampling) are fixed; identities are disjoint across train/validation/test; and CelebA attribute pairing follows the official annotations [10]. The discussion touches on renewability and diversity of templates in the sense of ISO/IEC 24745 [6], [7], but deliberately excludes transport or on-card API details from this short paper.

The complete protocol, including training of PCA–ITQ, threshold calibration, bit-statistics computation, and CelebA stress testing, is summarized in Fig. 1.

IV. RESULTS

All evaluations follow Section III. Thresholds are selected on MORPH validation and held fixed for the corresponding test reports unless stated otherwise. When shown, confidence intervals are nonparametric (percentile; 10,000 bootstrap resamples).

A. Hash Length vs. Verification on MORPH

Takeaway. A 64-bit template offers a strong size–accuracy trade-off: near-saturated AUC with $\sim 2\times$ smaller footprint than 128-bit, at a modest cost at FPR$= 0.1\%$. The 32-bit regime substantially narrows the genuine–impostor margin.

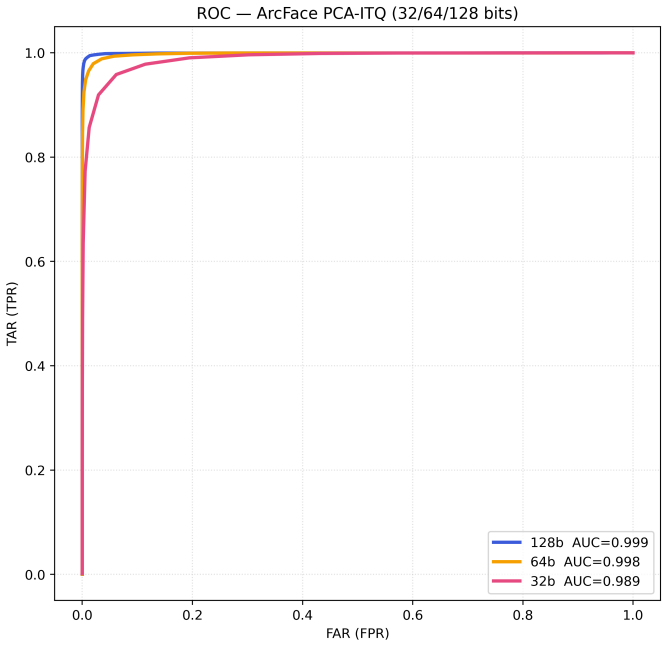

Fig. 2. MORPH ROC at $L \in \{32, 64, 128\}$ (PCA–ITQ). Global view complements the low-FAR zoom in Fig. 3.

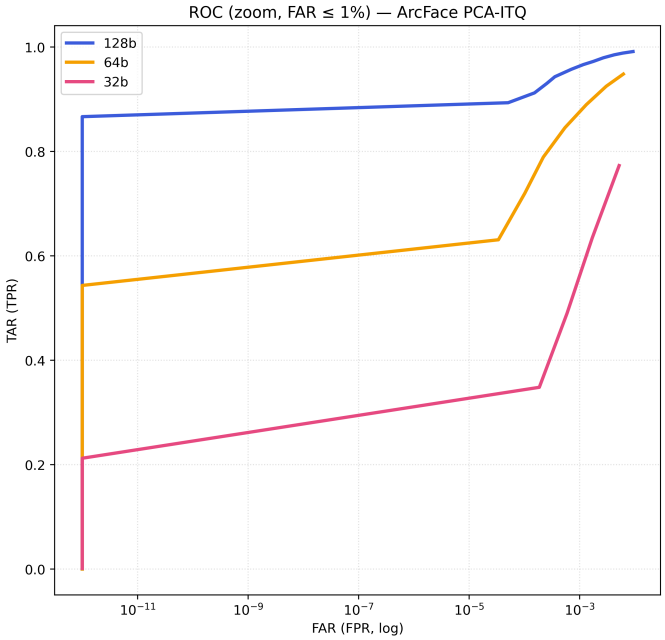

Fig. 3. MORPH ROC zoomed to FPR $\leq 1\%$ for $L \in \{32, 64, 128\}$ (PCA–ITQ). The 64-bit code retains most of the 128-bit low-FPR advantage; 32-bit degrades at strict FPR.

Fig. 2 shows the global ROC curves on MORPH for $L \in \{32, 64, 128\}$, while Fig. 3 zooms into the low-FAR region (FPR $\leq 1\%$) that is most relevant for deployment. Overall, these results indicate that moving from 128 to 64 bits yields a nearly two-fold reduction in template size with only a modest drop in low-FAR performance (about 2.6 percentage points at FPR=0.1%), whereas the transition from 64 to 32 bits incurs

TABLE II. Bit Statistics and Operating Thresholds on MORPH. Mean Balance is the Ones Ratio; Stability is Mean Within-ID Agreement

L	Mean balance	Mean stability	τ_{EER}	EER (%)	$\tau_{1\%}$ / TPR@1%
32	0.5016	0.876	5	41.19	3 / 17.65
64	0.5011	0.868	15	16.66	13 / 63.88
128	0.5003	0.860	37	5.33	35 / 91.66

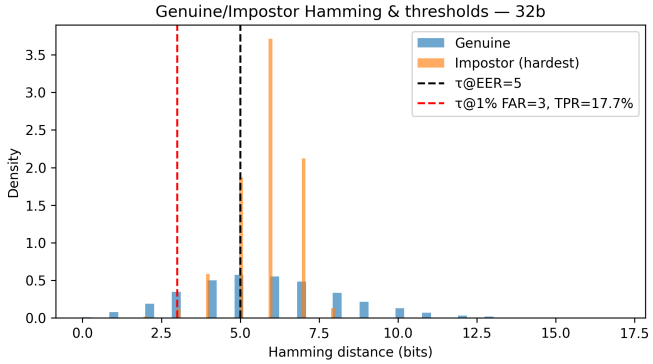

Fig. 4. MORPH (32-bit): Genuine vs. hardest impostor with τ_{EER} and $\tau_{1\%}$. Shows margin compression relative to 64-bit (Fig. 5).

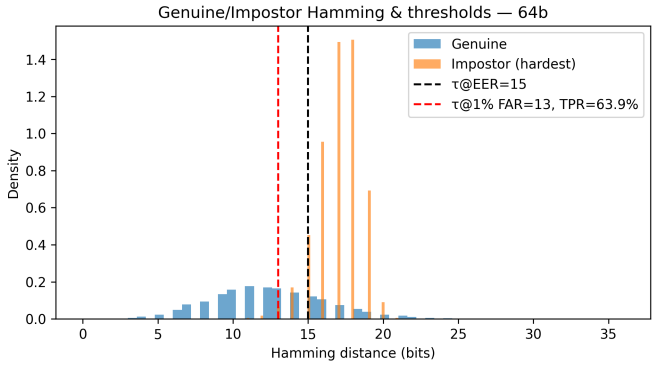

Fig. 5. MORPH (64-bit): Hamming distributions for genuine vs. hardest impostor with τ_{EER} and $\tau_{1\%}$ marked. Visualizes the margin underlying Table II.

a much sharper loss (around 39.9 percentage points at the same operating point). This pattern is consistent with previous observations that excessively short binary codes compress both intra-class and inter-class variation [1], [11], and it provides concrete bounds for deployments that must trade memory and I/O against verification accuracy.

B. Bit Balance, Stability, and Operating Points

Observation. Codes remain well balanced (≈0.5) across L; stability declines only mildly with longer codes. Operating points are governed by the shift of genuine mass relative to hardest impostors (Fig. 5).

C. CelebA Stress Tests and Alignment Effects

1) Eyeglasses Paired stability and localization: Finding. Eyeglasses increase flip-rates modestly on average (Table IV),

TABLE III. CELEBA (UNALIGNED VS. ALIGNED): MEAN STABILITY, EER, AND TPR@1%. ALIGNMENT YIELDS A SMALL BUT CONSISTENT BENEFIT AT 64 BITS

Setting	L	Stability	EER	TPR@1%
Unaligned	32	0.889	0.220	0.472
Unaligned	64	0.846	0.209	0.582
Unaligned	128	0.799	0.274	0.455
Aligned	32	0.893	0.264	0.408
Aligned	64	**0.856**	**0.196**	**0.591**
Aligned	128	0.804	0.259	0.555

TABLE IV. CELEBA (ALIGNED): PAIRED WITHIN-ID STABILITY WITH EYEGLASSES OFF/ON; Δ SHOWS THE MEAN DIFFERENCE WITH 95% CI

L	Off	On	Δ (95% CI)	$\tau_{1\%}$	TPR@1%	τ_{EER}/EER
32	0.899	0.887	**+0.012** [+0.0026,+0.0214]	4	0.269	8 / 0.382
64	0.870	0.854	**+0.016** [+0.0096,+0.0228]	15	**0.326**	21 / **0.302**
128	0.819	0.815	+0.004 [–0.0015,+0.0097]	40	0.288	48 / 0.386

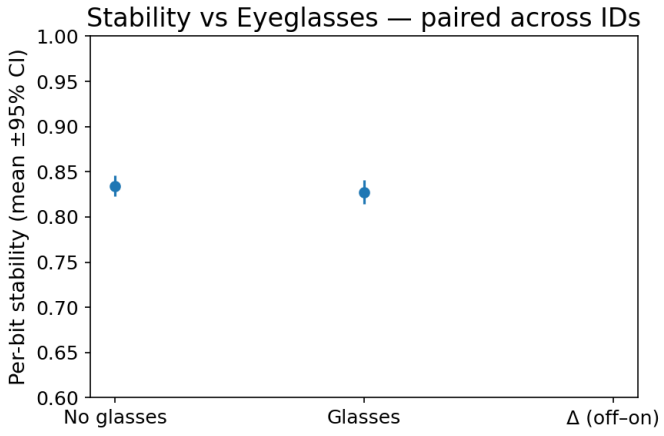

Fig. 6. CelebA (aligned): Per-bit stability (mean ±95% CI) with eyeglasses off vs. on. Instability is concentrated in a minority of positions.

but the effect is localized (Fig. 6), supporting targeted mitigation.

V. DISCUSSION

A. From Bit Statistics to Operating Points

The measurements link per–bit behavior to verification outcomes in a simple margin model. Let D_H^- denote impostor Hamming distance and D_H^+ genuine Hamming distance. For random, independent codes with mean balance ≈ 0.5, $D_H^- \sim \mathrm{Binom}(L, 0.5)$, so $\mathbb{E}[D_H^-] = L/2$ and $\mathrm{sd}(D_H^-) = \sqrt{L}/2$. Genuine distance can be written $D_H^+ = \sum_{k=1}^{L} X_k$ with $X_k \sim \mathrm{Bernoulli}(\phi_k)$, where ϕ_k is the within–ID flip–rate of bit k. Thus

$$\mathbb{E}[D_H^+] = \sum_k \phi_k, \qquad \mathrm{sd}(D_H^+) = \sqrt{\sum_k \phi_k(1-\phi_k)}.$$

Separation at a fixed FAR is governed by the gap $\Delta = \mathbb{E}[D_H^-] - \mathbb{E}[D_H^+]$ relative to the spreads. Increasing L moves *both* means apart, but also increases impostor concentration ($\propto \sqrt{L}$). Empirically, (Table II) higher L reduces EER because the impostor distribution tightens, whereas too small L (32 b) narrows Δ sharply and hurts low-FAR TPR. This explains *why* 64 b sits on a favorable size–accuracy frontier on MORPH (Table I) and why 128 b minimizes EER when footprint is less constrained. Compared to the benchmark-level study in [11], which contrasted different mobile-class encoders and hash lengths primarily via aggregate ROC and EER metrics, the present analysis exposes how a small subset of unstable bits dominates the degradation when codes are shortened. This observation aligns with the ageing-drift patterns reported for binary templates in [15], and it provides the statistical rationale for targeted error-correction schemes such as the unstable-bit Reed–Solomon approach in [18]. In contrast to general surveys on biometric hashing and template protection [19], [13], [14], the current work quantifies the bit-level conditions under which short codes (32–64 bits) remain viable in privacy-preserving, decision-only settings.

B. Attribute Localization and Alignment

Eyeglasses increase mean flip-rates only modestly (Table IV) but the effect is *concentrated* in a minority of positions (Fig. 6). This structured instability validates the premise of targeted mitigation. Alignment yields a small but consistent benefit at 64 b (Table III), suggesting that canonicalization reduces attribute spillover into otherwise stable positions—useful when parity budgets are tight.

C. Practical Guidance for Deployment

- Code length: Use L=64 as a default when template footprint and I/O are constrained; choose L=128 when minimizing EER is paramount and memory allows.
- Enrollment: Prefer $n \geq 3$ enrollment samples to stabilize estimates and reduce variance of genuine comparisons.
- Mitigation: Avoid large masks: small quotas may help at lenient thresholds, but aggressive masking collapses discrimination at low FAR.
- Thresholds: Calibrate a single τ per L on validation and freeze it for test to maintain reproducibility and operational simplicity.

D. Security and Privacy Considerations

The present study focuses on representation and decision statistics, not protocol hardening. That said, compact binary templates with near-balanced bits limit leakage from simple score-based queries. Refreshing the unstable-bit set $\mathcal{U}$ enables diversification and renewability in the spirit of template protection standards [6], while preserving unlinkability when identifiers are rotated. A thorough adversarial analysis (e.g., score-oracle exploitation, side-channel I/O) is orthogonal and left to the companion system paper.

E. Limitations and External Validity

The upstream encoder is intentionally fixed and the analysis focuses on MORPH (longitudinal) and CelebA (eyeglasses) to isolate bit-level effects. Outcomes may differ with other

encoders, demographics, or attributes (e.g., occlusions beyond glasses). CelebA attribute labels carry annotation noise; nevertheless, the localization effect is strong enough to survive that noise. Finally, parity was explored on short Reed–Solomon (RS) designs over a small unstable window; other lightweight constructions may offer different trade-offs.

F. Implications and Next Steps

Bit-level analysis is actionable: 1) it explains the observed hash-length trade-off through a margin lens, 2) it shows that instability is structured, not uniform, and 3) it motivates *localized* robustness that preserves discrimination. Potential next steps include extending attribute stress beyond eyeglasses, learning encoder-aware rotations that explicitly balance stability and entropy at target L, and evaluating parity/masking under formal privacy budgets and attack models.

Bottom line. Binary codes from a fixed encoder are already well balanced and mostly stable; the errors that matter come from a small, identifiable subset of bits. Masking discards signal and saturates quickly. For constrained deployments, 64 b with $n\geq 3$ enrollment samples is a principled operating point; 128 b is the accuracy ceiling when resources permit.

VI. Conclusion

This paper analyzes binary face templates produced by PCA–ITQ from a fixed, mobile-class encoder across hash lengths $L \in \{32, 64, 128\}$, linking *per-bit* statistics to verification behavior. Empirically, codes are near-balanced and mostly stable; the residual errors are driven by a small, identifiable subset of positions whose flip-rates rise under attribute change (eyeglasses). This structure explains the observed hash-length trade-off on MORPH and CelebA: 64 bits sit on a favorable size–accuracy frontier, while 128 bits minimize EER when footprint allows; 32 bits generally require redundancy.

Simple masking removes both noise and identity signal and saturates, especially at strict FPR.

- Guidance: For constrained deployments, use $L{=}64$ with $n\geq 3$ enrollment samples; prefer $L{=}128$ when accuracy dominates. Monitor per-bit flip profiles to refresh the unstable set and preserve renewability.
- Limitations and outlook: The study fixes the encoder and focuses on MORPH (longitudinal) and CelebA (eyeglasses). Extending to additional attributes and demographics, learning rotations that trade entropy for stability at target L, and formalizing adversarial/privacy analyses are promising next steps. We release manifests and pairing lists to support reproducibility and independent verification.

Acknowledgment

The authors thank colleagues at the Laboratory of Computer Systems and Vision (LabSIV), Faculty of Sciences of Agadir, Doctoral Studies Center, Ibn Zohr University, for helpful discussions and feedback on this work.

Access to the MORPH dataset was provided under an academic license to Ibn Zohr University for research use [9]; CelebA is publicly available for non-commercial research [10]. The authors also acknowledge the creators and maintainers of the tools used in the pipeline, including MTCNN for alignment [8] and PCA–ITQ for hashing [1].

Declaration on Generative AI

A language model was used for copy-editing and formatting assistance; all experimental design, analyses, and conclusions were conducted and verified by the authors.

References

[1] Y. Gong, S. Lazebnik, A. Gordo and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.

[2] A. Gionis, P. Indyk and R. Motwani, "Similarity search in high dimensions via hashing," in *Proc. VLDB*, pp. 518–529, 1999.

[3] Y. Weiss, A. Torralba and R. Fergus, "Spectral hashing," in *Proc. NIPS*, vol. 21, pp. 1753–1760, 2008.

[4] W. Liu, J. Wang, R. Ji, Y.-G. Jiang and S.-F. Chang, "Supervised hashing with deep neural networks," in *Proc. SIGKDD*, pp. 1735–1744, 2016.

[5] Z. Cao, M. Long, J. Wang and P. S. Yu, "HashNet: Deep learning to hash by continuation," in *Proc. ICCV*, pp. 5609–5618, 2017.

[6] ISO/IEC JTC 1/SC 27, "Information technology—Security techniques—Biometric information protection," ISO/IEC 24745:2011, International Organization for Standardization, 2011.

[7] N. K. Ratha, S. Chikkerur, J. H. Connell and R. M. Bolle, "Generating cancelable fingerprint templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 561–572, Apr. 2007.

[8] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[9] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Proc. IEEE FG*, pp. 341–345, 2006.

[10] Z. Liu, P. Luo, X. Wang and X. Tang, "Deep learning face attributes in the wild," in *Proc. ICCV*, pp. 3730–3738, 2015.

[11] A. Ganmati, K. Afdel, and L. Koutti, "Binary Face Templates with Mobile-Class CNNs: A Reproducible Benchmark for Smart-Card-Constrained Authentication," *Data and Metadata*, 2025. doi: 10.56294/dm20251223.

[12] M. Charikar, "Similarity estimation techniques from rounding algorithms," in *Proc. STOC*, pp. 380–388, 2002.

[13] J. Bringer, H. Chabanne and B. Kindarji, "The best of both worlds: Applying sound biometric principles to commercial biometric systems," in *Proc. IEEE BIOSIG*, pp. 1–10, 2008.

[14] J. Yang and J. Li, "A secure biometric authentication scheme based on cancelable biometrics and threshold homomorphic cryptosystem," in *Proc. IEEE ICC*, pp. 1–6, 2016.

[15] A. Ganmati, K. Afdel, and L. Koutti, "Ageing Drift in Binary Face Templates: A Bits-per-Decade Analysis," arXiv:2510.21778, 2025. [Online]. Available: https://arxiv.org/abs/2510.21778

[16] A. Ganmati, K. Afdel, and L. Koutti, "ISO/IEC-Compliant Match-on-Card Face Verification with Short Binary Templates," arXiv:2510.16078, 2025. [Online]. Available: https://arxiv.org/abs/2510.16078

[17] A. Ganmati, K. Afdel, and L. Koutti, "Deep Learning-Based Multi-Factor Authentication: A Survey of Biometric and Smart Card Integration Approaches," arXiv:2510.05163, 2025. [Online]. Available: https://arxiv.org/abs/2510.05163

[18] A. Ganmati, K. Afdel, and L. Koutti, "Targeted Reed-Solomon Parity for Binary Face Templates: Repairing Unstable Bits with Negligible On-Card Cost," Authorea, 2025, doi: 10.22541/au.176063118.87290030/v1.

[19] A. Ganmati, K. Afdel, and L. Koutti, "A Survey of Hash Techniques for Image Retrieval: Advancements, Challenges, and Applications," 2025. [Online]. Available: hal-05283675.

[20] A. Ganmati, K. Afdel, and L. Koutti, "Optimizing Facial Recognition: A Comparative Analysis of Transfer Learning Models on the Georgia Tech Face Database," in *Proc. ICRAMCS 2024*, Marrakech, Morocco, 2024, p. 396. [Online]. Available: https://icramcs2024.sciencesconf.org/data/Proceeding2024.pdf

[21] S. Singh, A. Zhang, M. Hossain, and A. D. Singh, "Accelerating Face Biometric Verification via Quantum PCA and Hybrid Processing," *IEEE Transactions on Quantum Computing*, vol. 3, no. 2, pp. 110–122, 2025.

[22] J. D. Ndibwile, P. Y. Sadewo, and K. D. M. Tanzila, "Fairness-Aware Face Presentation Attack Detection Using Local Binary Patterns," *Neural Computing and Applications*, 2025.

[23] U. A. Gimba and M. T. Tabot, "Enhancing Biometric Authentication through Face and Fingerprint Using CNNs," *Journal of Computer Engineering*, vol. 12, no. 3, pp. 44–53, 2025.

[24] S. S. Girajala and V. K. Bairi, "Dual-Model Biometric Authentication with Face and Voice," in *Proc. Int. Conf. on Smart Computing*, pp. 255–260, 2025.

[25] A. F. Muhammad, "Multimodal Biometric Authentication: Face, Fingerprints, and Voice Using AI," *Journal of Information Systems and Technology*, vol. 7, no. 1, pp. 19–31, 2025.