Biometrics under attack: On provenance, trust and threat robustness

Guilherme Schardong
UFSM

Tiago Novello
IMPA

Luiz Schirmer
UFSM

Nuno Gonçalves
ISR-UC, INCM

***Abstract*—Facial biometric systems are increasingly used in identity verification, mobile onboarding, border control, access management, and digital credentials. Their trustworthiness, however, cannot be assessed solely through recognition accuracy. Face images may be manipulated through morphing, retouching, synthetic generation, or replay, while defensive mechanisms such as watermarking and provenance verification may themselves affect biometric performance. This work surveys the main concepts, threats, defenses, and evaluation practices associated with trustworthy facial biometric media. It distinguishes identity, integrity, provenance, and recognition performance as complementary but non-equivalent objectives; reviews classical and learning-based face morphing methods; and discusses the role of watermarking in trustworthy biometric pipelines. We argue for a shift from isolated component evaluation toward integrated trust architectures combining secure acquisition, provenance, manipulation detection, watermarking, and biometric recognition. It concludes by identifying priorities involving identity-preserving watermarking, unified benchmarks, adaptive attack models, privacy-preserving provenance, and broader literacy concerning the capabilities and limitations of facial biometric systems.**

***Index Terms*—biometrics, morphing, watermarking**

I. Introduction

Biometric systems increasingly rely on digital media whose trustworthiness cannot be ascertained [1]. Face images, videos, and other biometric signals are routinely captured, compressed, transmitted, enhanced, retouched, and, in some cases, synthetically manipulated. In parallel, media watermarking has long been studied as a means to support authenticity, integrity verification, ownership assertion, and provenance tracking [2]. Yet these two areas have often evolved in isolation, with the biometrics community focusing primarily on recognition accuracy, robustness, and attack resilience, while the watermarking community has emphasized imperceptibility, payload, robustness, and detectability. This separation is problematic because biometric media are now embedded in security-critical workflows where identity, authenticity, and source trust must be considered jointly [1].

This work addresses that gap by bridging media watermarking and biometrics under a common perspective of *trustworthy biometric media*, using face morphing as a means to attack biometric models and systems. The core observation is that these topics are not adjacent. Morphing attacks challenge the identity semantics of biometric evidence by creating images that can match multiple subjects [3], [4]. Watermarking can serve as a mechanism for integrity assurance, provenance support, or tamper indication, but it may also perturb the biometric signal and degrade recognition performance [5], [6]. Understanding these interactions is essential for the design of biometric systems that are not only accurate but also secure, traceable, and trustworthy.

A key premise of this work is that *success cannot be evaluated from a single viewpoint*. In watermarking, performance is often described through image metrics such as perceptual quality, bit error rate, payload size, and robustness to signal processing operations [2]. In biometrics, evaluation centers on genuine and impostor score distributions, false accept and reject trade-offs, equal error rate, and operational performance at low false accept regimes [7]. In media forensics and provenance analysis, the emphasis may instead be on detecting manipulation, validating origin, or reconstructing processing history [1]. In this work, we aim to clarify the concepts of identity preservation, integrity assurance, provenance support, and recognition performance.

Our focus is on the face domain, due to its prevalence in modern society. Face morphing is a well established biometric attack, and face recognition provides a mature framework for studying how embedded distortions affect identity matching. We survey the relevant foundations and literature and provide a structured conceptual framework and an agenda for future work at the intersection of multimedia security, biometrics, and trustworthy AI. The broader goal is to encourage the community to move from isolated optimization of individual components toward integrated designs in which biometric utility, media authenticity, and provenance can be studied and engineered together. Figure 1 provides an overview of the inputs, attack surfaces, operation and evaluation of biometric systems.

Given this breadth, we deliberately prioritize depth over coverage: Sections II and III establish the four trust dimensions and morphing as the canonical attack, while watermarking and provenance are treated as complementary defense layers. Generative-model details are surveyed only to the extent needed to assess threat realism; full treatment is deferred to the referenced literature.

Contact e-mails: {guilherme.schardong, luiz.schirmer}@ufsm.br, tiago.novello@impa.br, nunogon@deec.uc.pt

979-8-3195-0255-1/26/$31.00 ©2026 IEEE

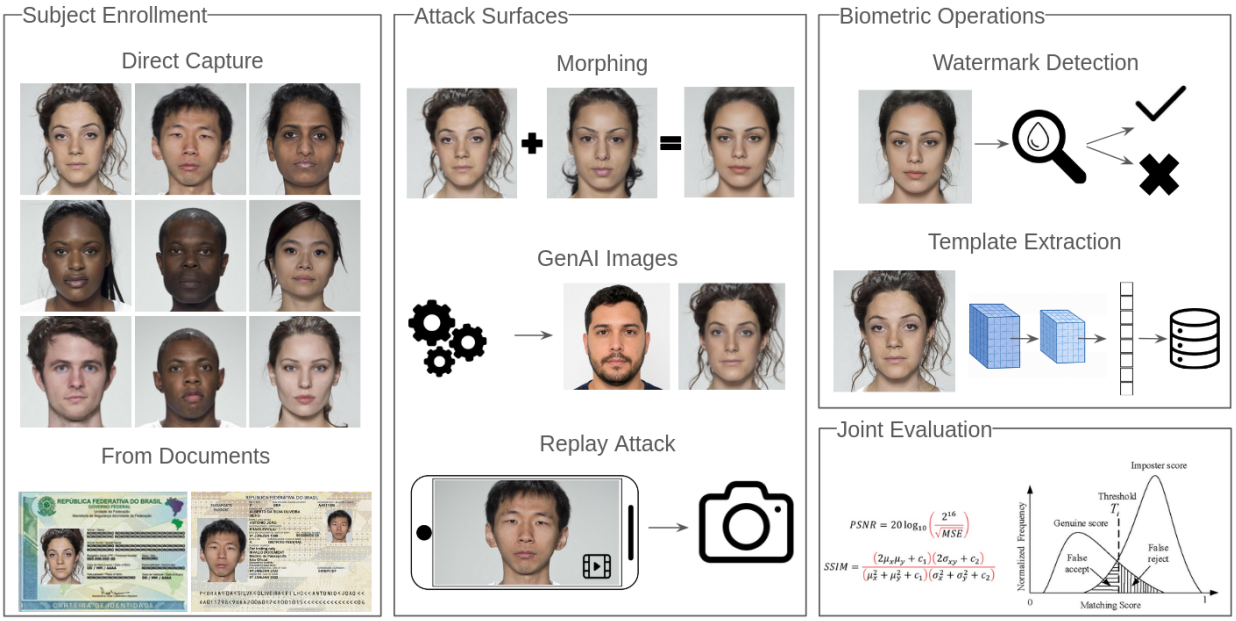


Fig. 1. **Overview of the biometric enrollment process**, its attack surfaces, and evaluation. Face biometrics may be collected from pictures or official documents, altered by morphing, synthetic generation, or replay, and defended through watermarking and secure templates, with image- and biometrics-related metrics.

II. Background and Key Definitions

In this section, we define key concepts in biometrics, watermarking, and attacks. The four main concepts explored throughout the work are *identity*, *provenance*, *integrity*, and *recognition performance*. In the literature, these concepts are often not clearly distinguished, which may confuse adopters and practitioners in general [8].

Identity concerns *who the biometric sample is supposed to represent*. In biometrics, identity protection aims at preserving the link between an image and the underlying person. If an attacker manipulates an image, such as by morphing two faces, and the system accepts both contributors as the same subject, then the identity has been compromised [4]. The question is whether this media object still supports a reliable claim about personhood. This is usually the most straightforward semantic layer. Models such as ArcFace [9] operate on this layer, mapping images into a representation that preserves identity-discriminative features.

Integrity concerns whether the media has been altered, *with no connection to the identity*; an image can preserve identity and still lose integrity: a retouched, recompressed, or cropped face photo may still match the correct person, but it is no longer the original acquisition. The most common watermarking techniques aim to protect this layer [10]. Conceptually, watermarking acts as a "seal" that is broken whenever the media is manipulated [11]. However, recent watermarking methods aim to provide robustness to tampering and noise, especially the print-capture cycle [12], [13]. In this case, such methods may also provide identity protection and store provenance signatures. The irony is that watermarking may compromise the integrity of the original image. Thus, the key idea is to detect *unauthorized modification*. A possible solution for this is to use *reversible watermarking methods*, where a watermark is introduced on capture and removed whenever forensic processing is required [14], [15].

Provenance handles origin and history. It concerns how the media was acquired, its chain of custody, and its processing history. This information may be stored as metadata, which is cryptographically signed [1]. While integrity indicates whether the content has changed, provenance concerns whether the media source and its history can be trusted. A face image submitted by the user for enrollment in a face recognition system may have adequate integrity, but poor provenance since its source is untrusted [16]. Conversely, a trusted capture device may sign or watermark images at acquisition time, allowing downstream systems to verify that the image originated from an approved sensor or workflow [17]. Provenance is especially important once we enter the realm of generative media, document issuance, forensic traceability, and robustness to attacks.

Recognition performance is about system behavior, not about semantics or trust claims directly. It concerns how well the matcher performs under some condition, typically in terms of cosine similarity distributions, True and False Acceptance Rates (TAR and FAR), False Recognition Rate (FRR), Equal Error Rate (EER), Receiver Operating Characteristic (ROC), or Detector Error Tradeoff (DET). A watermark may leave identity intact but still degrade recognition performance to an operationally unacceptable level, leading to an increase in FRR, for instance. "The right person is still in the image" is not the same as "the recognition system still works well". Recognition performance is therefore an empirical systems-level property, whereas identity, integrity, and provenance

are security or trust properties of the media and its use context [18].

A. Biometric Verification Metrics

Face recognition systems are commonly evaluated in either *verification* or *identification* settings. In verification, a system receives two samples and must check whether they belong to the same identity, while in identification, a system receives a single sample and searches a database of known samples for the closest match [19]. Modern face recognition systems typically map each image to a feature vector or embedding, and similarity between two embeddings is then computed using a score such as cosine similarity [9], [18].

Verification performance is assessed by analyzing the distributions of similarity scores for two types of comparisons. *Genuine* comparisons correspond to pairs of samples from the same subject, while *impostor* comparisons correspond to pairs from different subjects. Ideally, genuine scores should be high and impostor scores should be low, with minimal overlap between the two distributions. Any media transformation that disturbs these distributions can affect operational performance [20].

A threshold on the similarity score determines the decision; varying it trades security against usability. Two standard error measures are the *False Accept Rate* (FAR), the proportion of impostor comparisons incorrectly accepted, and the *False Reject Rate* (FRR), the proportion of genuine comparisons incorrectly rejected [20], [21].

Several summary measures are widely used. The *Receiver Operating Characteristic* (ROC) curve describes performance across thresholds by plotting TAR against FAR, while the related *Detection Error Tradeoff* (DET) curve emphasizes the trade-off between false accepts and false rejects [20]. The *Equal Error Rate* (EER) is the operating point at which FAR and FRR are equal, and is commonly used as a scalar summary of verification performance [20]. In practice, however, systems are often deployed at low false accept regimes, so it is also common to report measures such as TAR at a fixed FAR [20].

B. Watermarking: Concepts and Classes

Digital watermarking refers to the embedding of information into media in a manner intended to remain recoverable or verifiable under specified conditions [22]. Watermarking methods are commonly discussed in terms of a few core properties: *imperceptibility*, meaning that the embedded signal should not cause notable visual distortion; *robustness*, meaning that the embedded information should survive expected processing or attacks; *payload*, concerns the amount of information that can be embedded; and *detectability* or recoverability, meaning whether the watermark can be reliably identified or decoded [2], [23], [24].

A common distinction is among *(semi)-fragile* [5], [25], [26] and *robust* watermarking [12], [13], [27]. Fragile watermarking is designed to break under small modifications, thus being useful for tamper detection and integrity verification. Semi-fragile watermarking attempts to tolerate benign operations such as mild compression while breaking under more substantial changes. Robust watermarking aims to preserve the embedded signal under a broader range of operations, including printing and capture [2], [12]. Which class is appropriate depends on the intended trust objective. A fragile watermark may be suitable for integrity assurance, whereas a robust watermark may be more appropriate for source authentication or chain-of-custody support.

Watermarking methods can also be categorized by embedding domain. *Spatial-domain* methods modify pixel values directly, often in simple and computationally efficient ways, but they are typically less robust to compression and common image transformations [22], [23], [28]–[30]. *Transform-domain* methods embed information into coefficients obtained from transforms such as the Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT), and they often provide a more favorable balance between robustness and perceptual quality [31]–[33]. Recent learning-based methods use neural networks to optimize embedding and recovery jointly [12], [13], [25], [27], [34], [35]. Regardless of implementation, all watermarking approaches must balance trade-offs between signal fidelity and watermark reliability.

In the biometrics context, these watermarking trade-offs take on additional significance. A watermark that is visually imperceptible may still disturb identity-bearing structures used by a face recognition model, and a watermark that is robust to compression may still degrade verification performance by modifying discriminative regions or frequencies. Consequently, biometric utility must be considered alongside conventional watermarking objectives, since the latter becomes a controlled perturbation of a recognition-relevant signal.

III. Threats and Trust in Biometric Media

This section describes the main attack surfaces in biometric media workflows, using face morphing as an attack, and discusses how watermarking may support integrity and provenance in this landscape.

A. Biometric Media Workflows and Attack Surfaces

A biometric system relies on a sequence of stages, which in the face domain may include image acquisition, face detection and alignment, enhancement or compression, template generation, storage, transmission, enrollment, and later comparison against newly acquired samples [9], [18]. It is important to keep these in mind, as each exposes different vulnerabilities. A perturbation introduced at capture time differs fundamentally from one introduced after enrollment or during transmission [1].

In this setting, several attack surfaces arise naturally. At the point of acquisition, an attacker may attempt to submit manipulated or synthetically generated media instead of an authentic sample (see Fig. 2 for illustrations). During enrollment or document issuance, an attacker may alter the image associated with an identity record, potentially compromising subsequent verification. During storage and transmission, the media may

be recompressed, retouched, degraded, or intentionally modified in ways that affect its biometric utility. Even benign transformations, such as resizing or contrast adjustment, may become relevant if they interact with watermark detection or recognition behavior [12], [13]. These cases show that trust in biometric media is not determined solely by the final matching score. Still, it also depends on the conditions under which the media was produced and handled [18].

The interpretation of these risks depends on the threat model. An attacker may attempt to impersonate an identity using falsified media. A malicious user may seek to enroll an image that is valid enough to pass document issuance or biometric checks while encoding a future advantage. An insider may tamper with stored records, alter issuance workflows, or suppress provenance information. The same technical mechanism may therefore serve different security purposes depending on which attacker is assumed. A provenance mechanism may be useful against disputes over origin, but insufficient against content-level manipulation; a tamper-evident mechanism may detect unauthorized modification, but may not prevent enrollment of already manipulated content. For this reason, any claim about trustworthy biometric media must be grounded in a clear model of where the attacker intervenes, what they can modify, and what property is meant to remain protected [1], [18].

B. Morphing for Identity Manipulation

Among the different manipulations of biometric media, face morphing is noteworthy since it directly challenges the identity semantics of the sample. Morphing combines information from multiple identities so that the resulting image resembles each of them sufficiently to pass visual inspection or biometric comparison [3], [7]. Morphing goes beyond image editing, as it is not only content alteration but also identity-level manipulation. Its significance lies in the fact that it results in an image whose relationship to personhood is ambiguous.

Face morphing has evolved from classical image interpolation to modern generative and implicit-neural approaches [36]–[41]. Early morphing methods were landmark-based geometric warping and appearance blending. This approach still underlies many biometric morphing pipelines [3], [7].

This problem is notable in document issuance and border-control settings [42], [43]. Face recognition has become a central component of automatic border control and digital identity validation systems, making the quality and trustworthiness of the enrolled face image operationally critical [43], [44]. If a morphed image is accepted during enrollment or document issuance, it may later enable multiple individuals to use a single credential or to benefit from a shared biometric representation, thus weakening the one-person/one-identity assumption of biometric systems [7]. The problem is therefore not merely that the image has been altered, but that the *biometric evidence itself has been semantically compromised*.

Recent work has emphasized the operational aspect of this threat. Studies on morphing attack potential quantify the extent to which different strategies succeed against biometric verification systems, shifting the discussion from visual realism alone to measurable system vulnerability [45]. Likewise, benchmark-oriented efforts such as MorFacing [7] aim to estimate face-recognition robustness under morphing attacks, reinforcing the need to evaluate morphs as operational adversarial objects rather than only as synthetic images.

A recent trend is the use of generative models to create morphings. GAN-based methods such as MIPGAN [46] introduced identity-prior-driven generation of visually plausible morphs that challenge deployed recognition systems [47]; latent-structure formulations such as MLSD-GAN [48] and diffusion-based MorDIFF [49] extend this line, the latter also studying detectability. LADIMO [50] generates morphs through biometric template inversion with latent diffusion, and DCMorph [41] uses dual-stream cross-attention to improve attack effectiveness.

Another line of work approaches morphing from the perspective of Implicit Neural Representations (INRs). *ifmorph* formulates face morphing using coordinate-based neural networks that represent warping and blending as smooth, time-dependent transformations [39]. This work connects classical ideas from geometric morphing and variational regularization [36], [37], [51] with modern implicit representations, enabling smooth, continuous morphs over the image domain. *FLOWING* extended this idea by treating morphing as the construction of a neural flow, emphasizing structure preservation and principled control of the deformation process [40]. Although these methods are not designed exclusively for biometric attacks, they can produce high-quality, difficult-to-detect morphings that detection systems are not yet tuned for. Related work [38] also points in this direction by framing transformation sequences in terms of affine-diffeomorphic regression, suggesting a broader movement toward smoother and more structured continuous morphing models.

A morphed image may preserve visual quality and even remain internally consistent as a file, so that common integrity checks do not expose it; yet its identity semantics are compromised, and the recognition system may still accept it against one or more contributors [3], [45]. This motivates a joint treatment of media security and biometrics.

C. Morph Detection as an Alternative Defense

While watermarking protects media *proactively*, morph detection operates *reactively*: manipulation detection methods analyze submitted images for morph artifacts, ranging from texture and frequency-domain cues to learned detectors [3], [4], [7]. Notably, frequency-based detectors and frequency-domain watermarks exploit the same spectraldecomposition of the face image, making them natural competitors: both alter or inspect the frequency content that face recognition embeddings rely on. A watermarking scheme that degrades recognition performance may be preferable to a detector with poor morphing attack potential, and vice versa. Critically, detection requires no cooperation from the capture device, whereas watermarking requires trusted acquisition. These trade-offs motivate a joint

evaluation framework, which we identify as an open problem in Section IV.

D. Watermarking for Integrity and Provenance

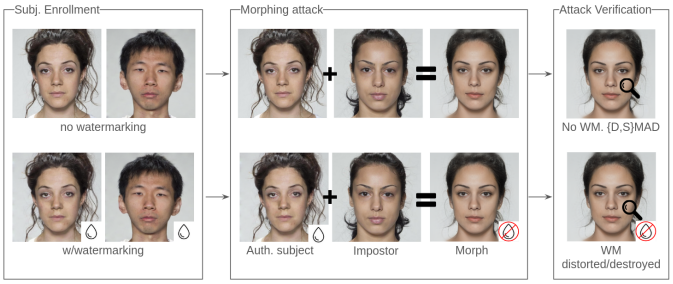


Fig. 2. Watermarking on capture provides an additional protection layer. A morphing attack with a non-watermarked image of an enrolled subject (top row) may be successful, since more sophisticated methods must be employed to check for manipulations. On the other hand, on a watermarked image of the same subject (bottom row), most attacks, including morphing, alter or destroy the watermark, which may be used as a flag for manual checking.

Within this threat landscape, watermarking can be viewed as a family of mechanisms for supporting trust in biometric media, though not as a universal solution. The most immediate role of watermarking is in integrity assurance. Fragile or semi-fragile watermarking can be used to indicate whether a biometric image has been altered after acquisition, certification, or enrollment [5]. In such settings, the watermark acts as a signal tied to expected content structure, and deviations from said structure may reveal tampering. This is useful when the goal is to detect modification rather than to infer identity directly [1], [2]. Figure 2 shows a simplified comparison of employing watermarking during enrollment and the effects on threat verification. In this case, the threat deployed is face morphing using an enrolled subject image, with an unenrolled third party.

A practical concern is that benign pre-processing (resizing, recompression, contrast normalization, or face cropping) may destroy a fragile watermark even when no attack occurred, producing false tampering alarms. This motivates semi-fragile schemes that tolerate a specified set of legitimate operations, and reversible watermarking, where the payload is removed before forensic processing. The choice of watermarking class must therefore match the expected processing pipeline of the deployment scenario, not only the threat model.

A second role of watermarking lies in provenance support. If a watermark is inserted at capture time, it may encode metadata about source, processing context, or chain of custody. In this case, the watermark contributes evidence not simply about whether the current content matches a previous version, but about the media origin and whether it passed through trusted stages [5]. Provenance support is attractive in biometric ecosystems involving digital credentials, remote verification, or document workflows, where trust in source and handling may be as important as the visual content itself [1]. Approaches such as StegaStamp [35] illustrate this, with a learned encoder–decoder pair that embeds a multi-bit payload which survives printing and capture by camera, using a differentiable approximation of the print–capture pipeline during training [35]. This is relevant to biometric credentials, where the medium may be physically printed and later re-imaged, and where a payload of tens of bits is sufficient to carry a device identifier or a pointer to a signed provenance record [1], [12], [13], [34].

These robustness gains come at a perceptual and semantic cost. Learned embedding schemes distribute the payload globally across the image, and the redundancy that grants survival under print–capture also increases the payload size [12]. In face imagery, these perturbations overlap with the mid- and high-frequency detail that recognition embeddings rely on, making the interaction between payload robustness and biometric utility an empirical question to be handled experimentally [13]. At the same time, watermarking introduces its own trade-offs and limitations. More fundamentally, watermark embedding changes the signal that is later used by the recognition system, meaning that a mechanism intended to strengthen integrity or provenance may also alter recognition performance. Watermarking should therefore be understood as one component in a larger trust architecture, not as a standalone guarantee [2], [5].

IV. Conclusions and Future Directions

This work has reviewed the main concepts associated with trustworthy facial biometrics. Its central argument is that face recognition, watermarking, integrity verification, and provenance should no longer be treated as isolated research topics, as they operate on the same media, interact within the same operational workflows, and influence different dimensions of trust. These distinctions must become requirements when designing and assessing facial biometric systems.

This need is becoming more urgent as facial biometrics move into routine social and institutional use. Identity verification, mobile onboarding, border control, access management, financial services, and digital credentials increasingly rely on face images as evidence. The expansion of these systems requires not only technical improvements in recognition accuracy, but also sustained research efforts regarding robustness and trust.

The field should consequently move toward an integrated notion of *trustworthy biometric media*. Under this view, a biometric sample should not be evaluated only by whether it matches a reference template, but also according to how it was captured, whether it has been altered, its processing history, and whether defensive mechanisms remain effective under realistic attacks. This requires a transition from component-level evaluation toward end-to-end trust architectures combining secure capture, biometric recognition, attack detection, watermarking, cryptographic provenance, and explicit decision policies.

Acknowledgments: This project has received funding from the European Union's Horizon Europe research and innovation programme ACHILLES under Grant Agreement No 101189689 and FCT - Fundação para a Ciência e a Tecnologia, under the project UID/00048 - Instituto de Sistemas e Robótica - Coimbra (ISR-UC).

REFERENCES

[1] J. Young *et al.*, "Media integrity and authentication: Status, directions, and futures," 2026. [Online]. Available: https://arxiv.org/abs/2602.18681

[2] W. Wan *et al.*, "A comprehensive survey on robust image watermarking," *Neurocomputing*, vol. 488, pp. 226–247, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222002533

[3] S. Venkatesh *et al.*, "Face morphing attack generation and detection: A comprehensive survey," *IEEE Transactions on Technology and Society*, vol. 2, no. 3, pp. 128–145, 2021.

[4] U. Scherhag *et al.*, "Face Recognition Systems Under Morphing Attacks: A Survey," *IEEE Access*, vol. 7, pp. 23 012–23 026, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8642312/

[5] M. Schäler *et al.*, "Reliable provenance information for multimedia data using invertible fragile watermarks," in *Advances in Databases*, A. A. A. Fernandes *et al.*, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 3–17.

[6] Z. Wang *et al.*, "Data hiding with deep learning: A survey unifying digital watermarking and steganography," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 6, pp. 2985–2999, 2023.

[7] I. Medvedev and N. Gonçalves, "MorFacing: A benchmark for estimation face recognition robustness to face morphing attacks," in *2024 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2024, pp. 1–10. [Online]. Available: https://ieeexplore.ieee.org/document/10744449/

[8] S. Bharadwaj *et al.*, "Biometric quality: a review of fingerprint, iris, and face," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, p. 34, 12 2014.

[9] J. Deng *et al.*, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 06 2019, pp. 4690–4699.

[10] A. Noore *et al.*, "Enhancing security of fingerprints through contextual biometric watermarking," *Forensic Science International*, vol. 169, no. 2, pp. 188–194, 2007.

[11] C.-H. Huang and J.-L. Wu, "Slic: Secure learned image codec through compressed domain watermarking to defend image manipulation," in *Proceedings of the 6th ACM International Conference on Multimedia in Asia*, ser. MMAsia '24. New York, NY, USA: Association for Computing Machinery, 2024.

[12] A. Cruz *et al.*, "Riemstega: Covariance-based loss for print-proof transmission of data in images," in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2025, pp. 7572–7581.

[13] F. Shadmand *et al.*, "Stampone: Addressing frequency balance in printer-proof steganography," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, WA: IEEE, 06 2024, pp. 4367–4376.

[14] R. Caldelli *et al.*, "Reversible watermarking techniques: An overview and a classification," *EURASIP Journal on Information Security*, vol. 2010, no. 1, p. 134546, 2010.

[15] A. Khan *et al.*, "A recent survey of reversible watermarking techniques," *Information Sciences*, vol. 279, pp. 251–272, 2014.

[16] P. Korus, "Digital image integrity – a survey of protection and verification techniques," *Digital Signal Processing*, vol. 71, pp. 1–26, 2017.

[17] M. Barbier *et al.*, "Image watermarking with biometric data for copyright protection," in *10th International Conference on Availability, Reliability and Security*, 2015, pp. 618–625.

[18] A. Jain *et al.*, "Biometrics: Trust, but verify," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 3, pp. 303–323, 2022.

[19] A. J. O'Toole and H. Abdi, "Face Recognition Models," in *International Encyclopedia of the Social & Behavioral Sciences*, N. J. Smelser and P. B. Baltes, Eds. Oxford: Pergamon, 2001, pp. 5223–5226. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B0080430767006902

[20] Y. Lee *et al.*, "Ocular and iris recognition baseline algorithm," National Institute of Standards and Technology, NIST Interagency Report NIST IR 7828, 2011. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7828.pdf

[21] M. Ngan and P. Grother, "Face Analysis Technology Evaluation (FATE) MORPH 4B: Considerations for implementing morph detection in operations," National Institute of Standards and Technology, NIST Interagency Report NIST IR 8584, 2025. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8584.pdf

[22] I. Cox *et al.*, "Watermarking as communications with side information," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127–1141, 1999.

[23] P. C. Mandal *et al.*, "Digital image steganography: A literature survey," *Information Sciences*, vol. 609, pp. 1451–1488, 09 2022.

[24] K. Hu *et al.*, "Learning-based image steganography and watermarking: A survey," *Expert Systems with Applications*, vol. 249, p. 123715, 2024.

[25] P. Fernandez *et al.*, "Watermarking images in self-supervised latent spaces," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 3054–3058.

[26] C. Zhang *et al.*, "A survey on reversible data hiding for uncompressed images," *ACM Comput. Surv.*, vol. 56, no. 7, Apr. 2024.

[27] T. Bui *et al.*, "Rosteals: Robust steganography using autoencoder latent space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 933–942.

[28] M. T. B. O. Munirah Gaaed Almutiri, "Digital image watermarking based on lsb techniques: A comparative study," *International Journal of Computer Applications*, vol. 181, no. 26, pp. 30–36, 11 2018.

[29] B. Chen and G. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[30] I. Cox *et al.*, "Secure spread spectrum watermarking for multimedia," *Trans. Img. Proc.*, vol. 6, no. 12, p. 1673–1687, 12 1997.

[31] A. Bors and I. Pitas, "Image watermarking using dct domain constraints," in *Proceedings of 3rd IEEE International Conference on Image Processing*, vol. 3, 1996, pp. 231–234 vol.3.

[32] S. N. Biswas *et al.*, "Mpeg-2 digital video watermarking technique," in *2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings*. IEEE, 2012, pp. 225–229.

[33] S. Gujjunoori and B. Amberker, "Dct based reversible data embedding for mpeg-4 video using hvs characteristics," *Journal of information security and applications*, vol. 18, no. 4, pp. 157–166, 2013.

[34] J. Zhu *et al.*, "HiDDeN: Hiding Data with Deep Networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Sep. 2018.

[35] M. Tancik *et al.*, "StegaStamp: Invisible Hyperlinks in Physical Photographs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020.

[36] T. Beier and S. Neely, "Feature-based image metamorphosis," in *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*. ACM, Jul. 1992, pp. 35–42. [Online]. Available: https://dl.acm.org/doi/10.1145/133994.134003

[37] J. Liao *et al.*, "Automating Image Morphing Using Structural Similarity on a Halfway Domain," *ACM Transactions on Graphics*, vol. 33, no. 5, pp. 1–12, 09 2014. [Online]. Available: https://dl.acm.org/doi/10.1145/2629494

[38] M. Lohr and L. Younes, "FineMorphs: Affine-diffeomorphic sequences for regression," 05 2023, arXiv:2305.17255 [stat]. [Online]. Available: http://arxiv.org/abs/2305.17255

[39] G. Schardong *et al.*, "Neural Implicit Morphing of Face Images," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE Computer Society, 06 2024, pp. 7321–7330. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR52733.2024.00699

[40] A. Bizzi *et al.*, "FLOWING: Implicit Neural Flows for Structure-Preserving Morphing," in *Advances in Neural Information Processing Systems*, D. Belgrave *et al.*, Eds., vol. 38. Curran Associates, Inc., 2025, pp. 153 041–153 065.

[41] T. Chettaoui *et al.*, "DCMorph: Face Morphing via Dual-Stream Cross-Attention Diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 06 2026, pp. 1556–1566.

[42] E. Sarkar *et al.*, "Vulnerability Analysis of Face Morphing Attacks from Landmarks and Generative Adversarial Networks," 12 2020, arXiv:2012.05344 [cs]. [Online]. Available: http://arxiv.org/abs/2012.05344

[43] F. Hidayat *et al.*, "Face Recognition for Automatic Border Control: A Systematic Literature Review," *IEEE Access*, vol. 12, pp. 37 288–37 309, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10459007/

[44] I. Medvedev *et al.*, "Towards Facial Biometrics for ID Document Validation in Mobile Devices," *Applied Sciences*, vol. 11, no. 13,

p. 6134, 07 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/13/6134

[45] M. Ferrara *et al.*, "Morphing Attack Potential," in *2022 International Workshop on Biometrics and Forensics (IWBF)*, 04 2022, pp. 1–6, arXiv:2204.13374 [cs]. [Online]. Available: http://arxiv.org/abs/2204.13374

[46] H. Zhang *et al.*, "MIPGAN – Generating Strong and High Quality Morphing Attacks Using Identity Prior Driven GAN," 04 2021, arXiv:2009.01729 [cs]. [Online]. Available: http://arxiv.org/abs/2009.01729

[47] E. Sarkar *et al.*, "Are GAN-based morphs threatening face recognition?" in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore, Singapore: IEEE, 05 2022, pp. 2959–2963. [Online]. Available: https://ieeexplore.ieee.org/document/9746477/

[48] A. R. P N *et al.*, "MLSD-GAN - Generating Strong High Quality Face Morphing Attacks Using Latent Semantic Disentanglement," in *2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, 12 2023, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10464945

[49] N. Damer *et al.*, "MorDIFF: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Diffusion Autoencoders," 03 2023, arXiv:2302.01843 [cs.CV]. [Online]. Available: http://arxiv.org/abs/2302.01843

[50] M. Grimmer and C. Busch, "LADIMO: Face Morph Generation through Biometric Template Inversion with Latent Diffusion," in *2024 IEEE International Joint Conference on Biometrics (IJCB)*, 09 2024, pp. 1–7, iSSN: 2474-9699. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10744444

[51] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Symposium on Geometry processing*, vol. 4. Citeseer, 2007, pp. 109–116.