

A Combined Feature-Based Framework for Disguise and Spoofing Detection in Face Recognition Systems

S. Pararajasingham

Abstract—Face recognition systems are increasingly deployed in access-control, mobile, and financial applications, yet remain vulnerable to two distinct classes of attack: spoofing, in which an impostor presents a photograph or video of an authorized user, and disguise, in which a legitimate user is rejected because their current appearance differs from their enrolled template due to accessories, facial hair, cosmetic change, illumination, or pose. Most prior work addresses these two problems separately, using liveness-detection methods for spoofing and occlusion-robust recognition methods for disguise. This paper proposes and empirically compares five combined feature-extraction and classification pipelines that target both problems within a single framework: PM (PCA and Minimum Euclidean Distance, MED), LPM (Local Binary Patterns, LBP, with PCA and MED), HPM (Histogram of Oriented Gradients, HOG, with PCA and MED), SM (Speeded-Up Robust Features, SURF, with MED), and HM (Harris corner features with MED). Each pipeline follows a common two-phase process comprising pre-processing, feature extraction, feature filtering, and classification. The methods were trained on 115 subjects drawn from the FEI, Disguised Faces Database (DFD), and NUAA databases and evaluated on six test conditions covering mixed appearances, frontal faces, dark illumination, left- and right-turned poses, and photo-spoofing attempts. The HOG-based pipeline (HPM) achieved the most consistent performance across conditions, with 94.59% accuracy on mixed-appearance disguise, 81.5%–93.2% across pose and illumination variants, and 91.67% on spoofing, while the LBP-based pipeline (LPM) achieved the highest spoofing-detection accuracy of the five pipelines (93.2%) but weaker robustness to pose change. These results show a measurable trade-off between spoof sensitivity and disguise robustness among classical feature representations, and motivate the deep-learning and cross-database extensions discussed in the concluding sections.

Index Terms—Face recognition, spoof detection, disguise detection, liveness detection, local binary patterns, histogram of oriented gradients, speeded-up robust features, Harris corner detector, principal component analysis, Euclidean distance classifier.

I. INTRODUCTION

AUTOMATIC face recognition is now embedded in mobile-device unlocking, e-commerce, e-health, border control, and mobile-banking applications, largely because it requires no physical token and can operate at a distance without the active cooperation of the subject. This convenience, however, comes with a security cost: a face template is

comparatively easy to acquire without contact, which exposes recognition systems to two related but distinct failure modes.

The first, spoofing, occurs when an unauthorized individual presents a photograph, replayed video, or other counterfeit likeness of an authorized user in order to obtain that user's access rights [1]. The most common spoofing attacks against 2D systems are carried out with printed photographs and simple video replay, precisely because these require no specialised equipment and can be produced from images gathered from social media or public records. The second failure mode, disguise, occurs when a legitimate, enrolled user is rejected because their presented appearance has changed relative to the enrolled reference image, whether through eyewear, facial hair, cosmetic change, hairstyle, illumination, pose, or long-term ageing. Both failure modes degrade the practical reliability of face recognition, yet they are commonly studied and countered separately: liveness- and texture-based anti-spoofing methods on one hand, and occlusion- or pose-robust recognition methods on the other.

This paper addresses both problems within a single evaluation framework rather than proposing one new descriptor. Four established feature-extraction techniques – Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), Speeded-Up Robust Features (SURF), and the Harris corner detector – are combined with two established filtering strategies (Principal Component Analysis, PCA, and strongest-keypoint selection) and a common Minimum Euclidean Distance (MED) classifier. The resulting five pipelines are evaluated side by side, under identical training and test conditions, so that the accuracy trade-offs between spoof sensitivity and disguise robustness can be measured directly rather than inferred from separate studies that use different protocols.

The contributions of this paper are threefold: (i) a unified two-phase pipeline that treats spoof detection and disguise-robust recognition as a single classification decision rather than two separate subsystems; (ii) a systematic, identical-protocol comparison of five classical feature-extraction and classification combinations across six real-world test conditions; and (iii) an explicit accuracy breakdown by attack and variation type that identifies where each combination is strong and where it fails, used in Section V to motivate directions for follow-up work. Portions of the methodology and evaluation reported here extend an earlier unpublished MCS dissertation on the same topic [24].

Manuscript received [insert date]; revised [insert date].

The author is with the University of Colombo School of Computing (UCSC), University of Colombo, Colombo, Sri Lanka (e-mail: asan935para@gmail.com).

II. RELATED WORK

A. Spoofing Attacks and Countermeasures

Motion-based countermeasures analyse subconscious liveness cues such as eye blinking and head or mouth movement [4]. Because such cues occur at a physiological rate of roughly 0.2–0.5 Hz, these methods typically require several seconds of video to accumulate a stable decision [4]; eye-blink-specific detectors exploit the fact that a live subject blinks approximately every two to four seconds, a cue that is difficult to reproduce convincingly with a static photograph or simple replay [2]. Texture-based countermeasures instead operate on a single image, analysing the frequency-domain or micro-texture differences between a live face and a recaptured photograph or screen replay [11]. Local Binary Pattern Variance (LBPV) is one such descriptor, jointly encoding local texture pattern and contrast information from sampling points around a circular neighbourhood [12]; more broadly, LBP-, DoG-, and HOG-based texture descriptors have been shown to discriminate recapture artefacts from genuine skin texture using only a single frame [15]. Image-quality-analysis methods generalise this idea further, designing features that are explicitly sensitive to the quality loss introduced by print or replay, and have reported strong within-database performance on the Idiap REPLAY-ATTACK and CASIA-FASD benchmarks [16]. Methods that fuse additional sensing modalities, such as 3D depth, infrared, or contextual scene cues, can achieve high robustness but impose extra hardware or protocol requirements, and specific contextual cues have been shown to be circumventable once the spoofing mechanism is known [17].

Table I summarises these four broad categories of spoofing countermeasure with their principal strengths, limitations, and representative reported performance.

B. Disguise Variations in Face Recognition

Disguise-robust recognition is complicated by low-quality or non-cooperative capture, temporal appearance change, and deliberate occlusion by accessories. Eigen-space approaches restricted to the periocular region (“eigen-eyes”) have been used to reduce sensitivity to occlusion below the eyes, reporting 87.5% accuracy on the Yale face database [6], [18], while feature-fusion approaches combining RGB and depth data from Kinect sensors have reported improved robustness to pose and partial occlusion relative to 2D-only baselines [7]. Benchmarking on the AR Face Database, which contains variation in eyewear, expression, and illumination, shows that even specialised occlusion-handling algorithms remain sensitive to the specific accessory and illumination combination presented [1], [10]; a nearest-neighbour classifier using skin-correlation features, for example, achieved only 45.8% accuracy on this benchmark. This body of work motivates evaluating disguise robustness across explicit, separated variation types rather than reporting a single pooled accuracy figure, the approach adopted in Section IV.

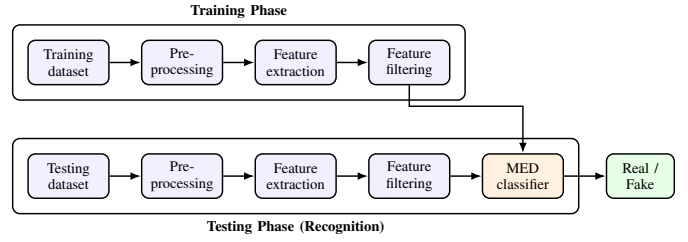


Fig. 1. Overall two-phase process. Training-phase filtered features (PCA eigenspace or strongest keypoints) supply the reference model used by the MED classifier in the testing phase.

III. MATERIALS AND METHODS

A. Datasets

Three public face databases were used. FEI [21] provides pose and expression variation for disguise evaluation; the Disguised Faces Database (DFD) provides accessory, facial-hair, and cosmetic variation; and NUAA provides both photo-spoofing attempts and disguise variation. Table II summarises their composition. The training set combined 115 subjects (15 from FEI, 50 from NUAA, 50 from DFD). The corresponding test set was partitioned into six condition-specific subsets: mixed appearances (37 images), frontal faces (200), left-turned faces (250), right-turned faces (250), dark-illumination faces (50), and spoofing attempts (60).

B. System Overview

The proposed framework follows a common two-phase process, training and testing, each composed of the same four stages: pre-processing, feature extraction, feature filtering, and classification. Five combined pipelines are instantiated from this common process by varying the feature-extraction and filtering stage: PM (PCA + MED), LPM (LBP + PCA + MED), HPM (HOG + PCA + MED), SM (SURF + MED), and HM (Harris + MED). Fig. 1 illustrates the overall process.

C. Pre-processing

Input images are 640×480 RGB colour images. Each image is cropped to a fixed 280×380 region, with crop origin (180, 40), then converted from RGB to greyscale prior to feature extraction.

D. Feature Extraction

LBP characterises local texture by comparing each pixel to its circular neighbourhood, encoding the result as a rotation-sensitive binary pattern [15]. HOG encodes local edge orientation and is comparatively insensitive to uniform brightness change [14]. SURF is a scale- and in-plane-rotation-invariant detector/descriptor pair using a Hessian-matrix approximation and a 64-dimensional Haar-wavelet-based descriptor [19]. The Harris–Stephens detector locates corner-like interest points from the directional variation of a local autocorrelation function [22]. Fig. 2 summarises what each descriptor responds to and the dimensionality it produces prior to filtering.

TABLE I
COMPARISON OF FACE ANTI-SPOOFING METHOD CATEGORIES REPORTED IN THE REVIEWED LITERATURE

Method Category	Representative Cue	Strengths	Limitations
Motion-based [4], [2], [13]	Eye blink, head/mouth movement	Good generalisation across capture devices	Low robustness to replayed motion; slow response (>3 s); high computational cost
Texture-based [12], [15], [11]	LBP / DoG / HOG micro-texture	Fast response (<1 s); low computational cost	Sensitive to acquisition-condition changes; weaker cross-database generalisation
Other cues [17], [7], [6]	3D depth, IR, context, voice	High robustness within protocol	Needs additional sensing hardware; slow response with audio/3D cues
Image-quality analysis [16]	Recapture-induced quality loss	Good generalisation; fast response; low cost	Often needs different classifiers tuned per attack type

TABLE II
DATABASES USED FOR TRAINING AND TESTING

Database	Subjects	Img./Subj.	Total
FEI	200	14	2,800
DFD	409	6	2,460
NUAA	15	–	10,230

LBP — pixelwise circular binary pattern; captures local micro-texture

HOG — 8-neighbour gradient direction; captures edge/shape structure

SURF — Hessian-based keypoints; 64-D descriptor, scale/rotation invariant

Harris — autocorrelation-based corner keypoints

Fig. 2. Summary of the four feature-extraction descriptors used by the LPM, HPM, SM, and HM pipelines respectively (PM uses raw pixel intensity with no descriptor).

E. Feature Filtering

For LPM and HPM, PCA is applied to the corresponding feature vectors. Each training image is reshaped into a column vector x_i , and the mean face vector is

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1)$$

Each image is mean-centred, $a_i = x_i - \bar{x}$, and combined into $A = [a_1, \dots, a_n]$. Rather than diagonalising the large $MN \times MN$ covariance matrix $C = AA^T$, the eigenvectors of the smaller $n \times n$ matrix

$$L = A^T A \quad (2)$$

are computed via singular value decomposition, and the eigenvectors of C recovered from those of L . The eigenvectors with the largest eigenvalues (the “eigenfaces”) are retained. For SM and HM, PCA is not applied; the built-in strongest-point selection retains a fixed number of the most salient SURF or Harris keypoints instead.

TABLE III
COMBINED PIPELINES AND DECISION THRESHOLDS

Pipeline	Feature	Decision Basis	τ
PM	–	Min. Euclidean dist.	6.50×10^{17}
LPM	LBP	Min. Euclidean dist.	7.60×10^{17}
HPM	HOG	Min. Euclidean dist.	3.30×10^5
SM	SURF	Matched keypoints	20
HM	Harris	Matched keypoints	4

F. Classification

The MED classifier is used uniformly, with two decision rules. For PM, LPM, HPM, the minimum Euclidean distance between the test image’s projected weight vector w_{test} and each class mean w_i is

$$d_{\min} = \min_i \|w_{\text{test}} - w_i\|^2, \quad (3)$$

and the input is classified as fake if $d_{\min} > \tau$, otherwise matched to $\arg \min_i \|w_{\text{test}} - w_i\|^2$. For SM and HM, the number of matched keypoints m between the test image and the best-matching candidate is compared to τ : the input is matched if $m > \tau$, otherwise classified as fake. Table III lists the pipelines and their fixed thresholds.

IV. RESULTS

Table IV reports the recognition accuracy achieved by each pipeline across the six test conditions. HPM achieves the highest accuracy on four of six conditions – mixed-appearance disguise (94.59%), left-turned (91.6%), right-turned (93.2%), and dark illumination (86.0%) – and a competitive 91.67% on spoofing. PM achieves the single highest spoofing accuracy (96.67%) but the lowest dark-illumination accuracy (38.0%). LPM achieves the second-highest spoofing accuracy (93.2%) but falls to 56.4%/57.2% on left/right-turned faces. SM and HM are competitive on spoofing (86.67%/70.0%) and frontal faces (78.5%/94.5%) but degrade sharply under pose change and dark illumination.

Fig. 3 visualises these results, making the spoof-versus-disguise trade-off discussed in Section V directly visible: the blue and orange bars (PM, LPM) peak on the rightmost group (spoofing) but dip lowest on the dark-illumination group, while the green bars (HPM) stay comparatively high across every group except a small gap on spoofing.

TABLE IV
RECOGNITION ACCURACY (%) BY TEST CONDITION

Pipeline	Mixed (n=37)	Front (n=200)	Left (n=250)	Right (n=250)	Dark (n=50)	Spoof (n=60)
PM (PCA+MED)	70.27	76.50	84.40	77.60	38.00	96.67
LPM (LBP+PCA+MED)	75.68	66.50	56.40	57.20	72.00	93.20
HPM (HOG+PCA+MED)	94.59	81.50	91.60	93.20	86.00	91.67
SM (SURF+MED)	81.08	78.50	36.40	37.20	26.00	86.67
HM (Harris+MED)	81.08	94.50	64.80	64.80	68.00	70.00

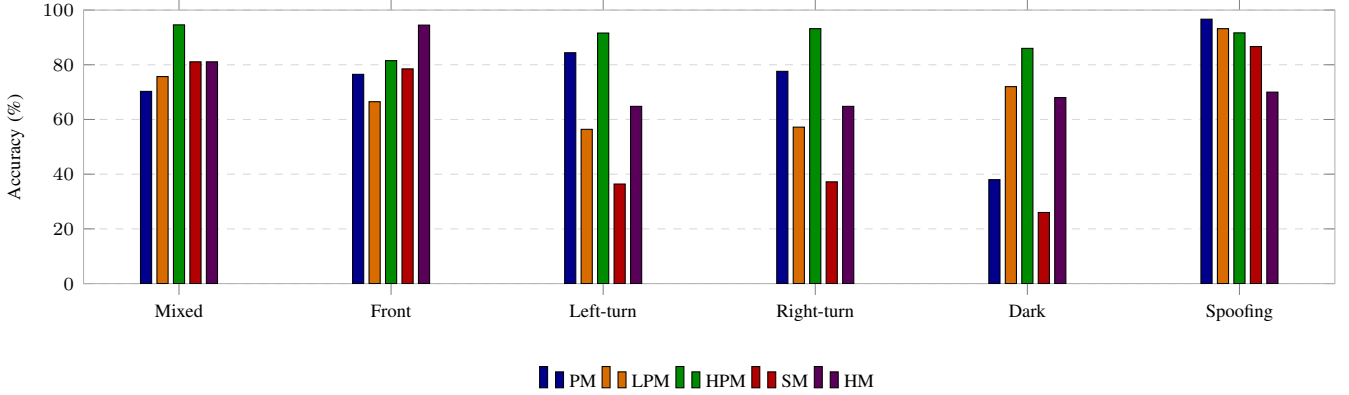


Fig. 3. Recognition accuracy (%) of the five pipelines across the six test conditions (data from Table IV). HPM (green) is the most consistently high performer outside spoofing, where PM (blue) and LPM (orange) lead.

V. DISCUSSION

Two patterns emerge from Table IV. First, HPM is the most consistently strong pipeline overall, plausibly because HOG's gradient-orientation encoding is comparatively insensitive to uniform illumination change, explaining its advantage on the dark-illumination and pose-variation subsets relative to the intensity-sensitive PM and keypoint-based (SM, HM) pipelines. Second, no pipeline that performs well on spoofing is simultaneously the strongest on pose variation: PM and LPM, both applying PCA to a global or texture-global representation, are better suited to detecting a flat photographic replay, while the gradient-based HPM is better suited to the fine-grained appearance changes of disguise. The keypoint-based pipelines underperform on pose and illumination variation because local keypoint detectors are sensitive to the appearance change introduced by out-of-plane rotation and low contrast.

These accuracy figures should be read as a same-protocol relative comparison rather than as directly comparable to headline metrics reported in the anti-spoofing literature, such as HTER or EER on Idiap REPLAY-ATTACK or CASIA-FASD, which use different databases, protocols, and error definitions.

A. Limitations and Threats to Validity

Several limitations constrain the generality of the reported results. First, all decision thresholds were fixed from the training set rather than calibrated on a held-out validation set. Second, training and testing were performed within the same combined databases rather than across independent databases, so results should be read as within-protocol performance.

Third, the evaluated attacks are limited to 2D photographic spoofing; video-replay and 3D-mask attacks are outside the current scope. Fourth, all five pipelines use hand-crafted, classical descriptors and were not compared against convolutional-neural-network or transformer-based baselines. Fifth, the training population (115 subjects) is modest by contemporary standards.

VI. CONCLUSION

This paper evaluated five combined feature-extraction and classification pipelines – PM, LPM, HPM, SM, and HM – for the joint problem of disguise-robust recognition and photo-spoofing detection in 2D face recognition. Under an identical protocol across three public databases and six explicit test conditions, HPM provided the most consistent performance across pose, illumination, and disguise variation (81.5%–94.59%) while remaining competitive on spoofing (91.67%), whereas PM and LPM achieved higher peak spoofing accuracy (96.67% and 93.2%) at the cost of reduced robustness to pose and illumination change.

VII. FUTURE WORK

Four directions follow from the limitations in Section V-A: (i) cross-database evaluation; (ii) extension to video-replay and 3D-mask spoofing with temporal liveness cues; (iii) direct comparison against CNN- or transformer-based baselines under the same protocol; and (iv) explicit modelling of long-term appearance change (ageing).

ACKNOWLEDGMENT

The author would like to thank Prof. N. D. Kodikara for his supervision and guidance during the original MCS dissertation research at the University of Colombo School of Computing, upon which this work is based.

DECLARATIONS

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. **Conflict of Interest:** The author declares no known competing financial interests or personal relationships that could have appeared to influence this work. **Data Availability:** The datasets used (FEI, DFD, NUAA) are publicly available from their respective sources, cited in the References. **Ethical Approval:** This study used only publicly available, previously de-identified face image databases; no new human-subject data were collected.

REFERENCES

- [1] N. Kose, "Spoofing and disguise variations in face recognition," Ph.D. dissertation, Télécom ParisTech, 2014.
- [2] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcam," in *Proc. IEEE 11th Int. Conf. Computer Vision (ICCV)*, 2007.
- [3] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Proc. Eur. Conf. Computer Vision (ECCV)*, 2010, pp. 504–517.
- [4] K. Kollreider, H. Fronthaler, and J. Bigun, "Real-time face detection and motion analysis with application in 'liveness' assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, 2007.
- [5] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, 2015.
- [6] L.-P. Morency, P. Sundberg, and T. Darrell, "Pose estimation using 3D view-based eigenspaces," in *Proc. IEEE Int. Workshop Analysis and Modeling of Faces and Gestures (AMFG)*, 2003.
- [7] M. Hayat, M. Bennamoun, and A. A. El-Sallam, "An RGB-D based image set classification for robust face recognition from Kinect data," *Neurocomputing*, vol. 171, pp. 889–900, 2016.
- [8] I. Chingovska *et al.*, "The 2nd competition on countermeasures to 2D face spoofing attacks," in *Proc. IAPR Int. Conf. Biometrics (ICB)*, 2013.
- [9] N. Ramanathan, A. Chowdhury, and R. Chellappa, "Facial similarity across age, disguise, illumination and pose," in *Proc. Int. Conf. Image Processing (ICIP)*, 2004, pp. 1999–2002.
- [10] A. M. Martinez and R. Benavente, "The AR face database," CVC Technical Report No. 24, 1998.
- [11] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," in *Proc. SPIE*, 2004, pp. 296–303.
- [12] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognition*, vol. 43, no. 3, pp. 706–719, 2010.
- [13] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *Proc. Int. Conf. Biometrics (ICB)*, 2007, pp. 252–260.
- [14] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. Int. Conf. Image Analysis and Signal Processing (IASP)*, 2009, pp. 233–236.
- [15] J. Yang and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proc. Int. Conf. Biometrics (ICB)*, 2013, pp. 1–6.
- [16] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, 2014.
- [17] J. Komulainen, A. Hadid, and M. Pietikäinen, "Context based face anti-spoofing," in *Proc. IEEE Int. Conf. Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.
- [18] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, 2001.
- [19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Computer Vision (ICCV)*, 2011, pp. 2564–2571.
- [20] Priyanka, R. Kaur, and A. Kaur, "A survey on feature extraction techniques in face recognition," *Int. J. Comput. Sci. Mobile Computing*, vol. 4, no. 5, pp. 166–174, 2015.
- [21] "FEI Face Database," fei.edu.br. [Online]. Available: <https://fei.edu.br/~cet/facedatabase.html>
- [22] "Harris Corner Detector," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Harris_Corner_Detector
- [23] K. Verma, A. S. Thoke, and P. Singh, "Neural network based approach for face detection cum face recognition," *Int. J. Electrical and Computer Engineering*, vol. 6, no. 3, 2012.
- [24] S. Pararajasingham, "Disguise and Spoofing Detection in Face Recognition," M.C.S. dissertation, University of Colombo School of Computing (UCSC), University of Colombo, Sri Lanka, 2019. [Online]. Available: <https://dl.ucsc.cmb.ac.lk/jspui/handle/123456789/4375>