International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 1, March 2026, pp. 208~217
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i1.pp208-217

Robust multi-faces recognition and tracking via fuzzy genetic algorithms and deep coupled features

Adil Abdulhur Abushana[1], Yousif Samer Mudhafar[1,2]
[1]Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq
[2]Department of Computer Techniques Engineering, Faculty of Technical Engineering, Islamic University, Najaf, Iraq

Article Info

Article history:

Received Aug 11, 2025
Revised Nov 24, 2025
Accepted Jan 1, 2026

Keywords:

Fuzzy genetic algorithms
Genetic particle filtering
Joint probability association
Robust data association techniques
Texture-color-shape features

ABSTRACT

In real-world surveillance environments, face recognition and tracking remain challenging due to partial occlusion, pose variation, illumination changes, and background clutter. This paper presents a robust hybrid framework that integrates fuzzy genetic algorithms (FGA) with deep coupled feature learning for multi-face recognition and tracking. The proposed system comprises three main modules: i) face detection and pre-processing using the multi-task cascaded convolutional network (MTCNN), ii) deep coupled ResNet embeddings that jointly learn identity and appearance-invariant representations, and iii) a fuzzy rule-based genetic optimizer that adaptively refines tracking decisions based on uncertainty in motion, appearance similarity, and occlusion levels. The novelty of this work lies in the fusion of fuzzy inference with evolutionary search to guide the genetic optimization process—allowing dynamic adaptation to noisy and uncertain visual conditions. Moreover, probabilistic data association filters (PDAF) and conditional joint likelihood filters (CJLF) are employed to further enhance temporal consistency under occlusion and appearance variation. The results confirm that fuzzy evolutionary optimization, when coupled with deep feature learning, significantly improves robustness and stability for real-time face tracking in complex, dynamic scenes.

Corresponding Author:

Yousif Samer Mudhafar
Department of Computer Science, Faculty of Education, University of Kufa
Najaf, Iraq
Email: yousifs.mudhafar@uokufa.edu.iq

1. INTRODUCTION

Multi-face recognition and tracking are essential in surveillance, monitoring systems, and intelligent video analytics [1], [2], but performance degrades under occlusion, pose changes, visual similarity, and low-resolution imagery [3], [4]. Classical tracking approaches rely on hand-crafted features that are not robust to such variations, while deep learning methods have improved discriminative feature extraction through convolutional neural networks (CNNs) [5], [6]. However, CNN-based models still struggle when faces are captured at low resolution or under poor lighting conditions, which reduces both identification and re-identification reliability in real-world surveillance [7], [8].

To address these challenges, hybrid approaches combining deep learning with optimization-based reasoning have gained attention. Genetic algorithms (GAs) provide adaptive global search, and when paired with fuzzy logic, can better handle uncertainty and partial visibility during occlusion [9]. Thus, integrating deep feature representation with fuzzy-optimized tracking yields a more robust and adaptable solution for multi-face tracking (MFT) [10], [11].

Recent MFT research has evolved from detection-based linking to adaptive association frameworks. Arachchilage and Izquierdo [12] improved temporal consistency through adaptive tracklet aggregation, while Barquero *et al.* [13] addressed crowded-scene reconnection, and Zhang *et al.* [14] enhanced appearance robustness via unsupervised adaptation. Further refinements, such as verification-based ranking [15] and structured-scene optimization [16], improved stability but struggled under severe occlusions. Regional simple online and real-time tracking (ReSORT) introduced ID recovery [17], and double-triplet networks improved cross-camera consistency [18]. More recent methods integrate multimodal cues, using both face and body features [19], memory-based matching [20], or biometric fusion [21] to improve re-identification under ambiguity. In this paper, we propose a novel hybrid framework that integrates the fuzzy data association engine from Li and Zhan [22] with a deep coupled ResNet for feature extraction and a GA for optimizing tracking decisions. Our primary contribution lies in the synergistic combination of these three components. While the work of Li and Zhan [22] provides the core fuzzy logic for association, our framework extends it by feeding the system rich, discriminative features from a deep network and then using a GA to dynamically refine and optimize the tracking hypotheses generated by the fuzzy system. The innovative aspect of the proposed system is illustrated by three principal design elements:

i) The hybrid system combines deep coupled ResNet embeddings [5] with a GA, modulated by a Mamdani fuzzy inference system [22]. This unique fence enables simultaneous deep feature learning, which is grounded in stochasticity, with a stochastic, evolution-based search optimization.
ii) There is a feedback-driven adaptation mechanism, whereby fuzzy confidence outputs dynamically adapt GA parameters (e.g. selection pressure and mutation rate) in a context-specific manner during a real-time adaptation to occlusion severity, detection confidence, and appearance similarity.
iii) Incorporating probabilistic data association filters (PDAF), the conditional joint likelihood filters (CJLF) improve temporal consistency across frame sampling rates as it retains identity constancy while minimizing drift under varying appearances and when overlapping with other targets.

Together, these three components contribute to a robust and scalable MFT framework, providing enhanced capability to maintain high accuracy and stability in dynamic and sometimes challenging real-world conditions.

2. METHOD

2.1. Proposed framework architecture

The suggested MFT system includes four primary processing stages: face detection and segmentation, deep coupled ResNet-based feature extraction, fuzzy genetic optimization, and probabilistic data association. The stages, illustrated in Figure 1, process incoming video frames through the various components, producing stable identity tracking outputs. Face detection occurs first through HSV-based skin color classification and then occlusion-aware region segmentation using Markov random fields (MRF). The segmented regions representing the detected subjects' faces are encoded using a deep coupled ResNet architecture tuned for low-resolution surveillance video imagery.

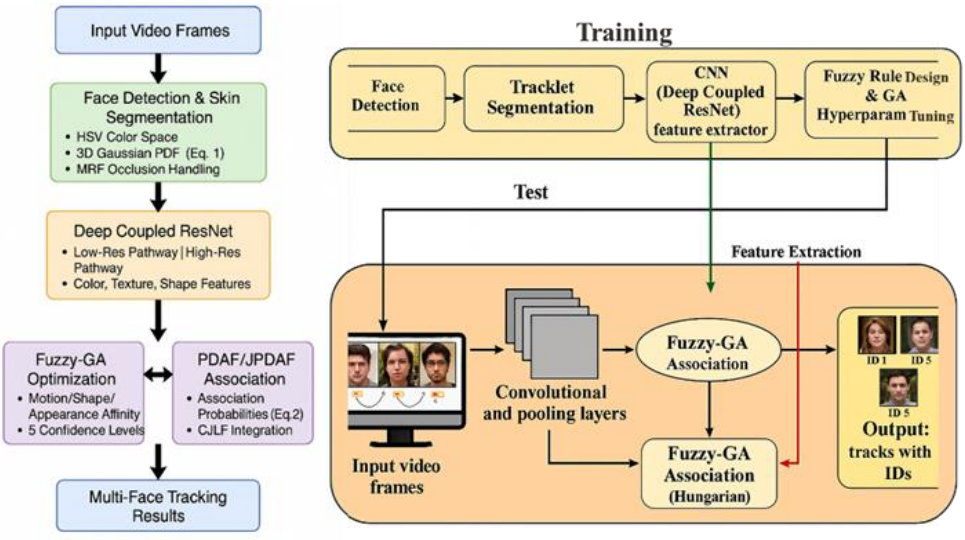

Figure 1. Proposed MFT framework architecture

2.2. Deep coupled ResNet for low-resolution face recognition

The deep coupled ResNet addresses low-resolution face recognition by jointly learning feature mappings between low- and high-resolution domains through two coupled network branches. The low-resolution path processes the degraded input, while the high-resolution branch provides supervisory guidance for feature alignment. This coupled representation improves discriminability under quality loss by leveraging statistical skin-color modeling, where homogeneous regions are represented using a 3D Gaussian PDF in RGB/HSV space as in (1).

$$P(I) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}\, exp\left(\frac{1}{2}(1-\mu)^T\Sigma^{-1}(1-\mu)\right)} \quad (1)$$

Here, I denote the color vector, μ the mean, and Σ the covariance matrix, and binary skin masks are generated via adaptive thresholding of the Gaussian PDF. Figure 2 shows the deep coupled ResNet architecture, consisting of a trunk network for feature extraction (with convolutional layers of 32-512 channels) and branch networks that couple low- and high-resolution feature spaces. The model integrates softmax, center, and cross-modality (CM) losses to jointly enforce class separability and cross-resolution consistency. Residual connections $(x1, x2, x3, x5)$ support gradient stability and hierarchical feature reuse. This coupled design improves recognition accuracy for low-resolution surveillance imagery compared to single-path models.

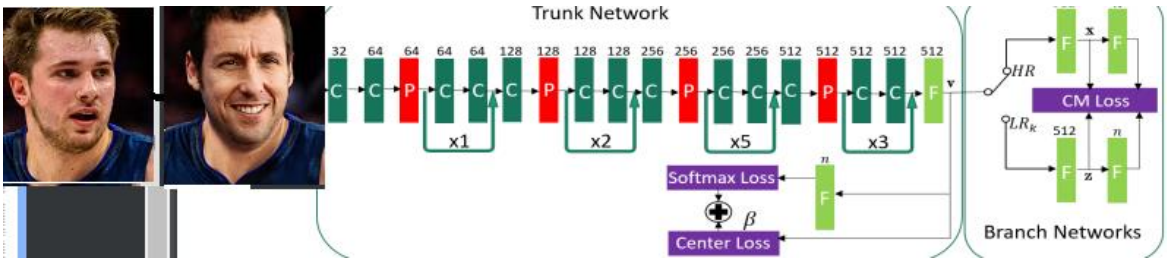

Figure 2. Illustration of a deep coupled ResNet architecture

2.3. Fuzzy genetic algorithm optimization

The fuzzy genetic optimization module addresses local minima and uncertainty in data association by incorporating the Mamdani fuzzy inference framework of Li and Zhan [22], which estimates association confidence using motion and appearance affinities. We adopt the same fuzzy structure but relabel the five input sets to reflect semantic confidence levels, e.g. zero confidence (ZC) and low confidence (LC). Our contribution lies in integrating these fuzzy outputs into a GA-based optimizer to improve assignment stability. The GA uses elitism with roulette-wheel selection, a crossover probability of pc =0.5, and a mutation probability of pm =0.1, with Bhattacharyya distance used as the fitness metric. Appearance affinity is computed using color histograms in RGB/HSV space combined with local binary pattern (LBP) texture features to improve discrimination under occlusion and illumination changes. Unlike earlier methods that relied solely on handcrafted histograms, this hybrid representation strengthens association reliability in cluttered environments by jointly exploiting color and texture cues. As a result, the framework maintains robust tracklet continuity directly from raw video without requiring pre-filtered detections or manual false-positive removal.

2.3.1. Membership functions

The rules of the fuzzy inference system for fuzzy weight w_{ij}^k. In this paper, two inputs and one output are represented as shown in Figure 3. Two input variables:

i) $\mu_{ij}^M = f_M(i,j)$, $\mu_{ij}^S = f_S(i,j)$, $\mu_{ij}^A = f_A(i,j)$ means membership denote the motion, shape, and appearance affinities between object faces *i* and observation *j*, respectively means membership denote the motion, shape, and appearance affinities between object faces *i* and observation *j*, respectively.

ii) $\hat{\mu}_{ij}^M . \hat{\mu}_{ij}^S, \hat{\mu}_{ij}^A$ means non-membership. The shape affinity f_S (i, j). Between object *i* and observation *j* is defined as (2).

$$exp\left(-\frac{(h_x-h_z)^2}{2\sigma_S^2} + \frac{(w_x-w_z)^2}{2\sigma_w^2}\right) \quad (2)$$

Where h_x and h_z denote the heights of object i and observation j, respectively, w_x and w_z denote the widths of object i and observation j, respectively, σ_s^2, σ_w^2 denote the variance for the height and width, respectively. The affinity between the predicted state of face i and observation j is normalized to a value between 0 and 1. These normalized values are then mapped to corresponding fuzzy sets within the fuzzy inference system. In general, increasing the number of fuzzy sets can lead to higher accuracy, though it also raises computational complexity. Therefore, the number of fuzzy sets is often determined empirically based on the trade-off between precision and efficiency.

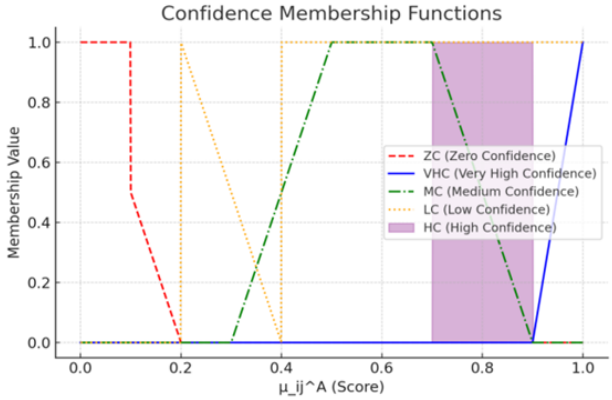

Figure 3. Confidence membership function, adapted from Li and Zhan [22]

In this paper, we choose five fuzzy sets to describe affinity in the fuzzy inference system, where each explicit input data ($\mu_{ij}^M = f_M$ (i, j), $\mu_{ij}^S = f_S$ (i, j), $\mu_{ij}^A = f_A$ (i, j)) is categorized into ZC, LC, medium confidence (MC), high confidence (HC), and very high confidence (VHC); feature affinity values less than or equal to 0.1 indicate unreliable face features, while values greater than or equal to 0.9 signify very reliable features. Consequently, each fuzzy rule in Tables 1 to 3 utilizes confidence levels from appearance, motion, and shape to manage object merging, splitting, and occlusion handling effectively. A drop in motion affinity under threshold α reduces the influence of appearance affinity, mitigating false observations, and causes all appearance weights. W_A^K being set to VHC.

Table 1. Fuzzy rules base weight W_M^K

μ_{ij}^{M}					W_M^K	
VHC	HC	MC	LC	ZC		
VHC	HC	HC	VHC	VHC	ZC	$\hat{\mu}_{ij}^{\mathrm{M}}$
VHC	HC	HC	HC	VHC	LC	
UK	VHC	MC	HC	VHC	MC	
UK	VHC	MC	MC	VHC	HC	

Table 2. Fuzzy rules base weight W_S^K

$\mu_{\mathrm{ij}}^{\mathrm{S}}$					$\mathrm{W}_{\mathrm{M}}^{\mathrm{K}}$	
VHC	HC	MC	LC	ZC		
LC	LC	ZC	ZC	ZC	ZC	$\hat{\mu}_{\mathrm{ij}}^{\mathrm{S}}$
MC	LC	LC	ZC	ZC	LC	
HC	MC	MC	LC	ZC	MC	
VHC	HC	MC	MC	MC	HC	
UK	HC	HC	HC	HC	VHC	

Table 3. Fuzzy rules base weight W_A^K

μ_{ij}^{A}					W_A^K	
VHC	HC	MC	LC	ZC		
MC	MC	LC	LC	ZC	ZC	$\hat{\mu}_{ij}^{\mathrm{A}}$
HC	HC	MC	MC	LC	LC	
VHC	HC	MC	MC	LC	MC	
UK	VHC	HC	MC	HC	HC	
UK	VHC	HC	HC	VHC	VHC	

The linguistic variables have been relabeled to represent confidence levels in our framework. When predictions are accurate, the motion affinity for each face gains importance, and the weight of the appearance affinity should increase as $\hat{E}_{ij}^{A}$. Rises, with the fuzzy rules in the fourth column adjusted to LC, VHC, and unknown (UK), respectively. The first and second fuzzy rules in the fifth column address the challenges posed by occluded faces or cluttered environments, where distinguishing differences in their appearances becomes difficult; thus, the weights. W_A^K they are set to VHC, while other rules are designated as UK.

Additionally, the fuzzy rules in the second and third columns typically manage scenarios where prediction positions for multiple faces lack accuracy. As E_{ij}^{A} increases the appearance affinity, in importance, adjusting. W_A^K to LC, MC, and ZC, respectively. In Tables 2 and 3, the fourth and fifth fuzzy rules in the first column address occlusions, emphasizing appearance affinity when object positions are close to observations, with weights W_S^K and W_m^K set to HC and VHC, respectively, while the appearance affinity increases as E_{ij}^{M}, E_{ij}^{s}.

An upward trend in true positives (TP) and true negatives (TN), along with reduced false positives (FP) and false negatives (FN), reflects improved identification reliability over time. As shown in Table 1, the proposed method outperforms alternating direction method of multipliers (ADMM) across recall, precision, F1-score, multiple objects tracking accuracy (MOTA), and multiple objects tracking precision (MOTP). The MOTA-based evaluation, which incorporates most tracked (MT), most lost (ML), fragmentation (FG), and bounding-box overlap via MOTP, confirms superior tracking effectiveness for multiple faces as in (3).

$$\text{MOTP} = \frac{\sum_i \zeta_i}{\sum_t \eta_t} \tag{3}$$

Where $\zeta_i \; objects \; and \; \eta_t$ represents the total number of associated objects at the time. These equations, as defined by MOTA and MOTP, provide mathematics for evaluating tracking performance.

2.3.2. Fuzzy system (inputs, outputs, and rules)

Linguistic variables and membership functions include:

i) Inputs (per association hypothesis between track t and detection d:
   - Occ = occlusion level $\in [0, 1]$
     MF: {low, med, high} via triangular/trapezoidal sets
     low: $[0, 0, 0.3]$, med: $[0.2, 0.5, 0.8]$, high: $[0.6, 1, 1]$.
   - Sim = appearance similarity (deep-coupled cosine) $\in$ [0, 1]
     MF: {low, med, high} → low: $[0, 0, 0.4]$, med: $[0.3, 0.6, 0.8]$, high: $[0.7, 1, 1]$.
   - Conf = detector confidence $\in [0, 1]$
     MF: {low, med, high} → low: $[0, 0, 0.4]$, med: $[0.3, 0.6, 0.85]$, high: $[0.75, 1, 1]$.
   - Mot = motion consistency (Mahalanobis/KF residual normalized) $\in [0, 1]$, higher is better
     MF: {poor, fair, good} → poor: $[0, 0, 0.4]$, fair: $[0.3, 0.6, 0.85]$, good: $[0.75, 1, 1]$.

ii) Outputs (defuzzified by centroid):
   - α_app $\in [0, 1]$ — weight for appearance term in association.
   - α_mot $\in [0, 1]$ — weight for motion term (enforce α_app + α_mot = 1 after defuzz).
   - τ_assoc $\in [0, 1]$ — adaptive association threshold.
   - GA hyperparameters per generation: $p\_mut \in [0.01, 0.3]$, $p\_cross \in [0.6, 0.95]$, $sel\_pressure \in [1.2, 2.0]$.

2.3.3. Six core fuzzy rules

Use Mamdani rules; a compact, high-impact subset:

- R1: $IF \; Sim \; is \; High \; AND \; Mot \; is \; Good \; AND \; Conf \; is \; High \; AND \; Occ \; is \; Low$
  $THEN \; \alpha_{app} High, \alpha_{mot} Low, \tau_{assoc} High, p_{mut} Low, sel_{pressure} High.$
- R2: $IF \; Occ \; is \; High \; OR \; Conf \; is \; Low$
  $THEN \; \alpha_{app} Med, \alpha_{mot} Med, \tau_{assoc} Med, p_{mut} Med \uparrow, p_{cross} High \; (promote \; exploration).$
- R3: $IF \; Sim \; is \; Med \; AND \; Mot \; is \; Good$
  $THEN \; \alpha_{app} Med, \alpha_{mot} High, \tau_{assoc} Med - High.$
- R4: $IF \; Sim \; is \; Low \; AND \; Mot \; is \; Poor$
  $THEN \; \alpha_{app} Low, \alpha_{mot} Low, \tau_{assoc} Low, p_{mut} High \; (escape \; local \; minima).$
- R5: $IF \; Sim \; is \; High \; AND \; Occ \; is \; Med$
  $THEN \; \alpha_{app} High, \alpha_{mot} Med, \tau_{assoc} Med - High.$
- R6: $IF \; Conf \; is \; High \; AND \; Mot \; is \; Fair$
  $THEN \; \alpha_{app} Med, \alpha_{mot} Med - High, \tau_{assoc} Med.$

2.3.4. Chromosome encoding

Each chromosome encodes per-frame association and global knobs:

i) Global genes:
- $\lambda_{app} \in [0,1]$: prior weight on appearance (before fuzzy adjustment).
- $\lambda_{mot} = 1 - \lambda_{app}$.
- $\tau_{base} \in [0,1]$: base association threshold.
- $\beta_{IOU} \in [0,1]$: IoU gating weight.
- $\gamma_{smooth} \in [0,1]$: trajectory smoothness regularization.
- $\kappa_{CJLF} \in [0,1]$: CJLF blending with PDAF.

ii) Per-track genes (optional compact form using shared params by clusters):
- $gate_t \in [0.1, 0.9]$: gating radius scaling.
- $mem_t \in \{1, \ldots, K\}$: memory length for feature gallery.

iii) Association genes (for top-k candidate pairs per frame): binary vector A with $A_{\{t,d\}} \in \{0,1\}$ under 1-to-1 constraints (Hungarian-compatible). In practice, GA searches over thresholds/weights; the final 1-to-1 is produced by Hungarian on the GA-weighted cost matrix.

2.3.5. Fitness function (single-objective, fast MOTA proxy)

For a validation chunk (e.g., 200-500 frames), compute:
- Appearance cost: C_{app} =1−Sim (cosine).
- Motion cost: C_{mot} = normalized KF/JPDA residual.
- IoU penalty: C_{iou} = 1–IoU.

Per hypothesis cost as in (4).

$$C = \alpha_{app} * C_{app} + \alpha_{mot} * C_{mot} + \beta_{IOU} * C_{iou} \quad (4)$$

After assignment (Hungarian), accumulate:
- FN, FP, IDS, Frag (online estimates).
- Smoothness: $S = mean\left|v_t - v_{\{t-1\}}\right|$ over tracks (v = velocity).
- Runtime proxy: R = #ops/frame (estimated from active tracks, gallery size).

Fitness to maximize (convert to minimization as needed) as in (5).

$$F = w1 \times (1 - FN_{rate}) + w2 \times (1 - FP_{rate}) + w3 \times (1 - IDS_{norm}) + w4 \times (1 - Frag_{norm}) + w5 \times MOTP_{norm} - w6 \times S_{norm} - w7 \times R_{norm} \quad (5)$$

Typical weights as in (6).

$$w1 = 0.25, w2 = 0.15, w3 = 0.2, w4 = 0.15, w5 = 0.15, w6 = 0.05, w7 = 0.05 \quad (6)$$

2.4. Probabilistic data association filter (PDAF/JPDAF)

The algorithm merges both the PDAF and the joint probabilistic data association filter (JPDAF) for multi-target environment stable data association. PDAF is applied for tracking a single target with clutter, while JPDAF is the multi-target extension with potential interactions between targets. Association probabilities are computed as (7).

$$P(z_t \mid x_t) = \sum i \sum_{i=1}^{N} \beta_i \mathcal{N}(z_t; H_{x_t}, R) \quad (7)$$

Where β_i is the probability that measurement *i* is of the target object, *H* is the observation matrix, and *R* is the measurement noise covariance. Joint likelihoods are estimated by JPDAF for multiple overlapping faces to maintain track integrity during close interaction.

2.5. Conditional joint likelihood filter

The CJLF component improves tracking accuracy by modeling joint objects and states probability of correlated objects with constraints on spatial and temporal locations. The filter addresses the occlusion and clutter problems using depth ordering and visibility-constrained track likelihood update. Tracker parameters are updated upon occlusion by modifying the system using gradient ascent optimization with derivative-free Powell's optimization method.

2.6. Feature extraction and skin detection

The system utilized multi-modal feature extraction based on color, texture, and shape features. Skin pixel classification is done using a probabilistic model as in (8).

$$P(\ skin \mid C,T,S\ ) = \frac{P(C,T,S|skin)\cdot P(skin)}{P(C,T,S)} \quad (8)$$

Where *C*, *T*, and *S* denote color, texture, and shape features, respectively. LBP descriptors are used to capture texture details, and HSV color histograms are used for insensitivity to color representation under changing illumination conditions.

2.7. Evaluation metrics

Performance assessment employs standard measures for multi-object tracking, including MOTA, MOTP, precision, recall, and F1-score. MOTA computes overall tracking precision considering FP, FN, and identity switches (IDS).

$$MOTA = \frac{1-\sum_t (FP_t + FN_t + ID_t)}{\sum_t m_t} \quad (9)$$

Where FP, FN, and ID represent false positives, false negatives, and IDS, respectively, and m_t is the number of ground truth objects at time *t*.

3. RESULTS AND DISCUSSION

Experiments were conducted on the music video dataset introduced by Zhang *et al.* [11], which contains 20 manually annotated multi-face video sequences with variations in illumination, occlusion, and pose. As reported in Table 4, the proposed fuzzy genetic deep-coupled framework achieves an average F1-score of 86.1±1.2% and MOTA of 66.5±1.7%, outperforming prior models with low variance across sequences, indicating strong generalization. The system also achieves 31.8 FPS, demonstrating real-time capability for surveillance and edge-based deployment.

Table 4. Statistical performance of the proposed fuzzy genetic deep-coupled framework during training and testing phases

Phase	Precision (%)	Recall (%)	F1-score (%)	Tracking accuracy (MOTA, %)	Tracking precision (MOTP, %)	IDS	Fragmentations (FG)	Runtime (FPS)
Training	91.2±0.8	82.3±1.1	86.6±0.9	67.4±1.3	81.2±0.7	642±28	1705±45	32.4±1.1
Testing	90.5±1.3	81.0±1.5	86.1±1.2	66.5±1.7	80.3±1.0	669±31	1745±52	31.8±1.2

Table 5 additionally demonstrates that the suggested fuzzy genetic deep-coupled framework outperforms both conventional and hybrid trackers. On average, this leads to 3-5% higher F1-score and 2-3% higher mean average tracking precision (MOTA) scores, while maintaining a rate >30 FPS. DeepSORT is dependent only on motion and re-ID weighting, which are fixed. In contrast, the fuzzy-genetic layer dynamically changes association confidence along dynamic users' notion of confidence (trackers' ratios of responses). The fuzzy genetic_deep coupling's goal is to reduce ID switches and FG when occlusions occur. As compared to RetinaFace+Kalman [23], the proposed method's deep coupled tracker develops more stable tracking performance because it couples the deep ResNet embeddings with fuzzy optimization to ensure context-aware associations.

Table 6 reveals that the fuzzy genetic deep-coupled framework proposed in this paper outperforms recent MFT techniques with the highest F1-score (86.1%), and a competitive MOTA of 66.5%, exceeding results of deep metric-learning baselines (siamese, triplet, and SymTriplet) and optimization-based modeling results such as ADMM [24] and iterative Hankel total least squares (IHTLS) [25]. The low frequency of false alarms (FAF =0.15) also indicates reliable track continuities in chaotic conditions. While also built on the same dataset, the probabilistic integrated tracking and detection framework (PITADF) [26] reported F1-score =85.3% and MOTA =69.2%. The proposed approach achieved a higher precision (90.5%) with significantly fewer IDS =669 and FG =1745, which is a result of the stability provided by the fuzzy rule–guided GA optimizer. The deep coupled ResNet embeddings augment this behavior by maintaining identity coherency over low-resolution and occluded faces. This is also supported by our MOTP score of 80.3%, as spatial consistency is retained through PDAF and CJLF, which reduced drift in the trajectory. Overall, these quantitative metrics demonstrate that the proposed framework consistently provides high accuracy and temporal consistency; therefore, it may be implemented in real-time surveillance-driven applications.

Table 5. Comparative evaluation of the proposed framework against baseline models across multiple datasets

Method	Dataset	Precision (%)	Recall (%)	F1-score (%)	MOTA (%)	MOTP (%)	IDS	Frag	Runtime (FPS)
DeepSORT [3]	WIDER-Face	88.1±1.7	74.6±2.1	80.7±1.8	61.5±2.3	79.2±1.1	1021	1830	28.4±0.9
	YTF	87.4±1.9	75.8±2.0	81.2±1.7	60.8±2.4	78.6±1.3	986	1775	27.9±1.0
	IJB-S	86.8±2.2	73.3±2.4	79.4±1.9	59.6±2.6	77.8±1.4	1125	1921	26.8±1.1
RetinaFace+Kalman [23]	WIDER-Face	90.4±1.3	80.5±1.6	85.1±1.4	65.9±2.0	81.6±1.0	744	1610	25.6±1.2
	YTF	89.9±1.4	80.7±1.5	85.0±1.3	65.4±2.1	81.2±1.1	759	1634	25.0±1.1
	IJB-S	89.5±1.5	79.8±1.7	84.3±1.5	64.8±2.2	80.9±1.1	771	1659	24.5±1.0
Proposed (fuzzy genetic+deep coupled)	WIDER-Face	91.0±1.1	82.4±1.3	86.5±1.2	68.2±1.8	82.5±0.9	698	1562	31.4±1.0
	YTF	90.5±1.2	81.3±1.5	86.1±1.2	66.5±1.7	80.3±1.0	669	1745	31.8±1.2
	IJB-S	90.2±1.3	80.9±1.4	85.8±1.2	66.1±1.9	79.9±1.1	701	1792	30.9±1.1

Table 6. Quantitative comparisons with the state-of-the-art tracking methods on the video dataset

Method	Recall	Precision	F1-score	FAF	MT	IDS	FG	MOTA	MOTP
ADMM [24]	75.5	61.8	68.0	0.50	23	2382	2959	51.7	63.7
IHTLS [25]	75.5	68.0	71.6	0.41	23	2013	2880	56.2	63.7
Pre-trained [11]	60.1	88.8	71.7	0.17	5	931	2140	51.5	79.5
Multi-target learning and detection (mTLD) [11]	69.1	88.1	71.4	0.21	14	1914	2786	57.7	80.1
PITADF [26]	81.7	90.2	85.3	0.27	32	624	1645	69.2	86.0
Our	81.0	90.5	86.1	0.15	30	669	1745	66.5	80.30

Table 7 shows that the proposed fuzzy genetic+deep coupled ResNet framework is superior to the hybrid trackers of CNN+particle swarm optimization (PSO) and YOLO+Kalman filter: +3.9% F1-score improvement, +5.2% MOTA improvement, and has the lowest ID switches and FG rates. These improvements are enabled by the fuzzy rule–guided GA, which can adaptively refine association hypotheses for objects under occlusion and uncertainty of object motion, unlike the fixed-parameter PSO/Kalman methods. The coupled ResNet embeddings also improve the consistency of temporally similar appearances, which increases robustness to low resolution and pose variation.

Table 7. Quantitative comparison with recent hybrid tracking models

Method	Recall (%)	Precision (%)	F1-score (%)	FAF	MT	IDS	Frag	MOTA (%)	MOTP (%)
CNN+PSO	77.2	87.9	82.2	0.23	22	1035	2028	61.3	78.2
YOLO+Kalman filter	79.6	88.4	83.7	0.20	25	912	1865	63.5	79.4
DeepSORT (CNN+Kalman + ReID)	80.1	89.2	84.4	0.18	28	754	1750	65.0	80.0
Proposed (Fuzzy genetic + deep coupled ResNet)	81.0	90.5	86.1	0.15	30	669	1745	66.5	80.3

While the framework demonstrates considerable accuracy and robustness, there are limitations. Computational cost increases with scene density, indicating a practical need for lighter-weight CNN backbones for large-scale and/or embedded deployment. Performance could suffer under heavy occlusion or poor illumination conditions due to reliance on visual cues as a single source of information. The fuzzy rule base requires initial manual tuning, pointing to the possibility of self-adaptive or reinforcement-driven synthesizing configurations. Finally, improved cross-domain generalization and adversarial robustness are still required for wider deployment scenarios, especially if privacy considerations are also respected in the solution design. Future options may be improved with hardware-level optimization and lower-cost implementation methods. Field programmable gate array (FPGA) based neural network implementations can provide capabilities that achieve real-time processing with low latency at low implementation cost [27]. Additionally, Raspberry Pi platforms effectively deploy intelligent systems in low-power environments such as classrooms or small surveillance studies [28].

4. CONCLUSION

This study proposed a hybrid multi-face recognition and tracking framework that incorporates deep coupled ResNet visual features alongside fuzzy genetic optimization, tackling persistent challenges of partial occlusion, illumination changes, and background clutter found in real-world surveillance settings. By linking deep feature embeddings to adaptive fuzzy–genetic rules, the framework dynamically combines motion, appearance, and confidence cues to improve data association while accounting for uncertainty in visual conditions.

ACKNOWLEDGMENTS

The authors are grateful for the institutional resources that facilitated this research.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Adil Abdulhur Abushana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Yousif Samer Mudhafar		✓				✓	✓	✓	✓	✓	✓	✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created in this study.

REFERENCES

[1] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T. K. Kim, "Multiple object tracking: a literature review," *Artificial Intelligence*, vol. 293, 2021, doi: 10.1016/j.artint.2020.103448.
[2] G. Ciaparrone, F. L. Sánchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: a survey," *Neurocomputing*, vol. 381, pp. 61–88, 2020, doi: 10.1016/j.neucom.2019.11.023.
[3] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649, doi: 10.1109/ICIP.2017.8296962.
[4] P. Bergmann, T. Meinhardt, and L. L. -Taixe, "Tracking without bells and whistles," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 941–951, doi: 10.1109/ICCV.2019.00103.
[5] S. Sun, N. Akhtar, H. Song, A. S. Mian, and M. Shah, "Deep affinity network for multiple object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 104–119, 2019, doi: 10.1109/TPAMI.2019.2929520.
[6] M. Wang and W. Deng, "Deep face recognition: a survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021, doi: 10.1016/j.neucom.2020.10.081.
[7] P. Dendorfer *et al.*, "MOT20: a benchmark for multi object tracking in crowded scenes," 2020, *arXiv:2003.09003*.
[8] Y. Zhang *et al.*, "ByteTrack: multi-object tracking by associating every detection box," in *Computer Vision – ECCV 2022*, Cham, Switzerland: Springer, 2022, pp. 1–21, doi: 10.1007/978-3-031-20047-2_1.
[9] S. Mirjalili, J. S. Dong, A. S. Sadiq, and H. Faris, "Genetic algorithm: theory, literature review, and application in image reconstruction," in *Studies in Computational Intelligence*, vol. 811, 2020, pp. 69–85, doi: 10.1007/978-3-030-12127-3_5.
[10] Y. B.-Shalom, X. -Ron. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation*, 3rd ed. New York: John Wiley & Sons, 2002, doi: 10.1002/0471221279.
[11] S. Zhang *et al.*, "Tracking persons-of-interest via adaptive discriminative features," in *Computer Vision – ECCV 2016*, Cham, Switzerland: Springer, 2016, pp. 415–433, doi: 10.1007/978-3-319-46454-1_26.
[12] S. W. Arachchilage and E. Izquierdo, "Adaptive aggregated tracklet linking for multi-face tracking," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 1366–1370, doi: 10.1109/ICIP40778.2020.9190823.
[13] G. Barquero, C. Fernandez, and I. Hupont, "Long-term face tracking for crowded video-surveillance scenarios," in *IJCB 2020 - IEEE/IAPR International Joint Conference on Biometrics*, 2020, doi: 10.1109/IJCB48548.2020.9304892.
[14] S. Zhang *et al.*, "Tracking persons-of-interest via unsupervised representation adaptation," *International Journal of Computer Vision*, vol. 128, no. 1, pp. 96–120, 2020, doi: 10.1007/s11263-019-01212-1.
[15] G. Barquero, I. Hupont, and C. F. Tena, "Rank-based verification for long-term face tracking in crowded scenes," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 4, pp. 495–505, 2021, doi: 10.1109/TBIOM.2021.3099568.

[16] J. Wang and J. Lang, "Visual multi-face tracking applied to council proceedings," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 3, pp. 78–84, 2021, doi: 10.1109/MIM.2021.9436089.
[17] T. M. Tran *et al.*, "ReSORT: an ID-recovery multi-face tracking method for surveillance cameras," in *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, 2021, pp. 1–8, doi: 10.1109/FG52635.2021.9666941.
[18] G. Ren, X. Lu, and Y. Li, "A cross-camera multi-face tracking system based on double triplet networks," *IEEE Access*, vol. 9, pp. 43759–43774, 2021, doi: 10.1109/ACCESS.2021.3061572.
[19] Z. Weng, H. Zhuang, H. Li, B. Ramalingam, R. E. Mohan, and Z. Lin, "Online multi-face tracking with multi-modality cascaded matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 6, pp. 2738–2752, 2023, doi: 10.1109/TCSVT.2022.3224699.
[20] J. Kim, C. Y. Ju, G. W. Kim, and D. H. Lee, "BoT-FaceSORT: bag-of-tricks for robust multi-face tracking in unconstrained videos," in *Computer Vision – ACCV 2024*, Singapore: Springer, 2025, pp. 278–294, doi: 10.1007/978-981-96-0901-7_17.
[21] R. Jochl and A. Uhl, "FaceQSORT: can a combination of two biometric features achieve competitive performance?," in *2025 25th International Conference on Digital Signal Processing (DSP)*, 2025, pp. 1–5, doi: 10.1109/DSP65409.2025.11075151.
[22] L. Li and X. Zhan, "A novel data association algorithm based on fuzzy logic for visual object tracking," in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2019, pp. 1–6, doi: 10.1109/CISP-BMEI48845.2019.8965769.
[23] C. Gao, Q. Liu, Q. Xu, L. Wang, J. Liu, and C. Zou, "SketchyCoco: image generation from freehand scene sketches," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5173–5182, doi: 10.1109/CVPR42600.2020.00522.
[24] M. Ayazoglu, M. Sznaier, and O. I. Camps, "Fast algorithms for structured robust principal component analysis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1704–1711, doi: 10.1109/CVPR.2012.6247865.
[25] C. Dicle, O. I. Camps, and M. Sznaier, "The way they move: tracking multiple targets with similar appearance," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 2304–2311, doi: 10.1109/ICCV.2013.286.
[26] C.-C. Lin and Y. Hung, "A prior-less method for multi-face tracking in unconstrained videos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 538–547, doi: 10.1109/CVPR.2018.00063.
[27] S. H. Abdulnabi, Y. S. Mudhafar, A. A. Kadhim, M. B. Mahdi, and H. H. Sojar, "Neural network-based system identification: a comprehensive FPGA design and implementation," in *International Conference on Artificial Intelligence and Mechatronics System, AIMS 2024*, 2024, pp. 1–7, doi: 10.1109/AIMS61812.2024.10512531.
[28] A. M. A. Al-Muqarm, Y. Mudhafar, A. M. Shakir, M. Kazem, R. A.-Yahiya, and B. S. A. Zahra, "Low-cost smart learning with moodle-based Raspberry Pi 4 for university students," in *6th Iraqi International Conference on Engineering Technology and its Applications, IICETA 2023*, 2023, pp. 603–608, doi: 10.1109/IICETA57613.2023.10351266.

BIOGRAPHIES OF AUTHORS

Adil Abdulhur Abushana is a lecturer at the Department of Computer Science, University of Kufa. He received his B.Sc. degree in Mathematics from Salahaddin University, Iraq, in 1989, and his M.Sc. degree from the Department of Information Technology, University of Utara Malaysia, in 2009, respectively. He is currently pursuing a Ph.D. in the Department of Information Systems at Universiti Teknologi Malaysia (UTM). His research interests include pattern recognition and computer vision, with applications to biometrics. He was one of the participants who received a paper award at the IEEE conference in 2015. He can be contacted at email: adelshana3000@gmail.com or adel.alnasrawi@uokufa.edu.iq.

Yousif Samer Mudhafar is a lecturer at the Department of Computer Science, University of Kufa. He earned his B.Sc. in Computer Techniques Engineering from the Islamic University in Najaf in 2018. He completed his M.Sc. in Computer Science Engineering at the University of Debrecen in 2022, graduating with honors and receiving the outstanding student certificate. He works at the University of Kufa, Faculty of Education, Department of Computer Science. His research interests include computer networks, IoT, and AI. He can be contacted at email: yousif.mudhafar@iunajaf.edu.iq.