Quantifying Operational Drivers of Multimodal Biometric Verification in Aerial Surveillance

Anonymous CVPR submission

Paper ID *****

Abstract

Multimodal biometric verification is increasingly applied across operational contexts ranging from close-range security cameras and building-mounted surveillance to long-range ground sensors and unmanned aerial system (UAS) imagery. Variations in acquisition conditions—such as image resolution, viewing geometry, and motion artifacts—pose significant challenges for cross-domain algorithmic generalization. This study evaluates two independent multimodal biometric verification systems developed under the Intelligence Advanced Research Projects Activity (IARPA) Biometric Recognition and Identification at Altitude and Range (BRIAR) program, comparing performance on close-range and aerial datasets. Close-range video served as a baseline to quantify the decline in verification performance on aerial footage. The dataset included six UAS platforms, spanning small quadcopters at 10m altitude to medium-sized fixed-wing aircraft at 360m. Mixed-effects logistic regression identified image resolution (head and body pixel counts), head height, sensor characteristics, and algorithm selection as primary determinants of verification success, whereas demographic attributes and mission gait were not significant predictors. Activity type and collection site influenced performance in close-range data but had negligible impact on UAS imagery. These results clarify modality-specific strengths and limitations and highlight opportunities to enhance cross-domain biometric verification.

1. Introduction

Biometric verification has become a critical component of modern security, surveillance, and intelligence operations. Traditional systems, optimized for close-range acquisition under controlled conditions, face substantial challenges in unconstrained environments. Unmanned aerial system (UAS) introduce unique constraints such as reduced facial resolution, variable viewpoints, motion-induced blur, and fluctuating lighting, all of which can degrade recognition performance relative to ground-based surveillance.

The Intelligence Advanced Research Projects Activity (IARPA) launched the Biometric Recognition and Identification at Altitude and Range (BRIAR; https://www.iarpa.gov/research-programs/briar) program to advance algorithms capable of robust operation across diverse acquisition scenarios. Phase 2 of BRIAR focused on developing multimodal systems that integrate facial, gait, and body appearance cues to verify identities across both aerial and close-range sensors [1, 2]. Multiple independent teams contributed end-to-end systems, enabling comparative evaluations of algorithmic performance under operationally relevant conditions.

Biometric verification has been extensively studied in both controlled and unconstrained settings, with foundational work demonstrating challenges of recognition under variable lighting, pose, and resolution [8]. Traditional face recognition research highlights issues such as viewpoint variation, head orientation, and illumination that adversely affect performance [8]. Multimodal systems that combine face, gait, and body-shape cues improve robustness, particularly at larger distances where faces are poorly resolved [3, 10]. Recent BRIAR-related studies, including FarSight, demonstrate that integrating multiple biometric cues improves recognition in UAS and high-altitude imagery [9], while covariate analyses identify image resolution, camera distance, and viewing geometry as key determinants of algorithm performance [4, 7]. Despite these advances, cross-domain verification—comparing aerial and close-range performance under typical operational conditions—remains underexplored.

This study analyzes two independent multimodal systems from BRIAR Phase 2: UAS/aerial imagery, characterized by large standoff distances, variable altitudes, and dynamic imaging conditions; and close-range surveillance data, offering higher-resolution facial imagery but greater scene complexity. Our goal is to quantify verification performance, identify modality-specific strengths and limitations, and evaluate the impact of sensor geometry, image quality, and fusion strategies on operational deployment.

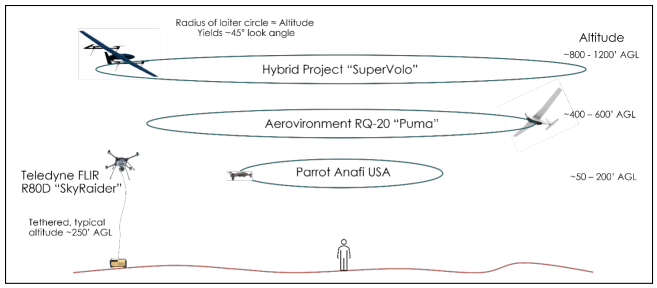


Figure 1. Overview of a typical daily collection plan and how the aircraft were positioned relative to the human subjects.

2. Methods

2.1. Aerial Collection Process

The aerial data collections for BRIAR were extensive and logistically complex, with up to four aircraft flown simultaneously at varying vertical and horizontal separations. These operations demanded careful planning and coordination to ensure safety and mission effectiveness. Safety was a top priority at every stage, given the operational challenges and regulatory demands. Scouting trips to potential field sites were crucial for assessing airspace, terrain, and logistics, ensuring site selection aligned with federal aviation regulations. The minimum 1-kilometer ground range requirement often limited ideal site options, requiring adaptations at more challenging locations, such as near airports.

Risk mitigation included issuing Notices to Airmen (NOTAMs), maintaining radio communication with manned aircraft, establishing emergency response plans, obtaining Certificates of Authority (COAs), and staffing multiple visual observers to monitor the skies continuously.

Accurate maps were critical for situational awareness and operational planning. Commercial satellite imagery often lacked sufficient resolution, so the Autonomous Systems group employed an eBee mapping drone to produce high-resolution orthomosaics (1cm ground sampling distance). These maps supported logistical planning, human subject positioning, and precise placement of ground-based cameras. Large-format "battle boards" printed from the maps further streamlined setup and coordination, enabling efficient and safe data collection.

A typical collection plan is illustrated in Figure 1. It generally involved the Anafi, Puma, and Supervolo circling the subject at various altitudes. Early collections maintained roughly constant radii to preserve a 45° look angle, while later flights incorporated offset and elliptical orbits to capture a broader range of distances and downward angles. The R80D was often stationed in a fixed location with a tether providing continuous power, allowing indefinite flight but restricting mobility to the area above its ground station. Supervolo and R80D could remain aloft most of the day, whereas smaller quadcopters and the Puma required frequent battery changes. Aircraft maintenance and refueling were scheduled during breaks or after hours when possible.

Weather conditions significantly affected operations, particularly for small UAS, which are sensitive to environmental factors. High humidity and low temperatures risked ice accretion on airfoils and pitot tubes, while rain and dust threatened functionality. Wind was a common constraint, as small UAS generally operate safely only below 20–25mph; high winds often grounded flights, notably during BGC-6 in Fort Collins. Extreme conditions also reduced battery efficiency and impacted human performance. The tethered FLIR system, with higher wind tolerance, occasionally provided limited capability during adverse weather.

A variety of aircraft were selected based on sensor type, operational capabilities, and field-proven use. Planned mid-project upgrades and new procurements did not occur due to funding constraints, but overall the systems performed as expected, with few issues.

- **Hybrid Project Supervolo**: Large hybrid VTOL aircraft with a 9' wingspan, combining vertical takeoff via electric motors and horizontal flight powered by a gasoline 2-stroke engine (100 mph cruise, 7-hour endurance). Equipped with **Trillium HD25** and later upgraded to **HD40** camera payload. Ideal for long-duration, high-altitude operations.
- **FLIR/Aeryon R80D Skyraider**: Mid-sized, military-spec quadcopter with fully electric operation and 25-minute flight time (extended to unlimited with an Elistair Safe T2 tether system). Equipped with **SR-EO/IR Mk II thermal camera** (no zoom) and **HDZoom30** optical sensor for high-resolution face and body data collection. Ideal for stationary, persistent biometric collection in localized areas.
- **Parrot Anafi USA**: Small government grade quadcopter with 30-minute flight time per battery, U.S.-made and NDAA-compliant (DIU Blue Listed). Equipped with an integrated gimbal featuring a combined EO and thermal camera package, ideal for reliable and straightforward biometric collection. Also used in early collections was the commercial grade Parrot Anafi which had a simpler and lower resolution camera system.
- **Autel Evo II**: Small quadcopter with 30-minute flight time, equipped with a non-swappable payload featuring a 1" CMOS EO camera and thermal sensor. Reliable and used in a similar role to the Anafi USA to diversify the biometric data collection.
- **Aerovironment RQ-20 Puma**: Medium-sized fixed-wing drone with 2–3 hours of flight time, launched by hand or pneumatic catapult, and utilizing a unique deep stall landing system. Equipped with a swappable **Mantis i45 EO/IR/thermal camera** package, ideal for ISR operations and extended biometric collection. Reliable overall, with occasional issues related to icing, landing

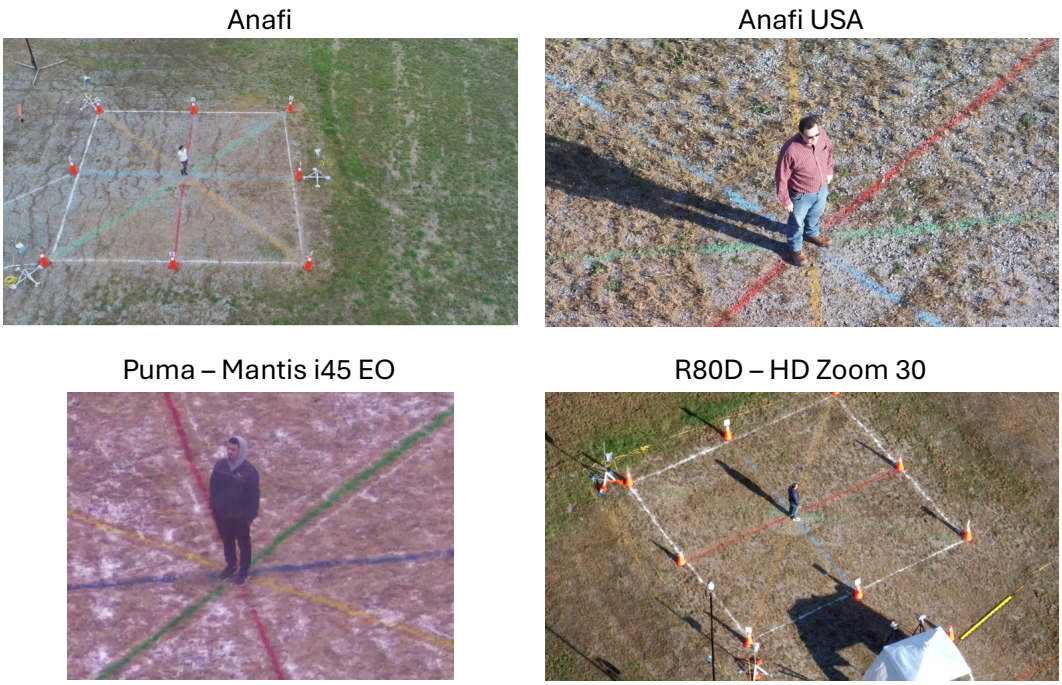


Figure 2. Sample imagery from the various aircraft and camera systems flown for this study. Note that the weather, zoom level, and aircraft vibration can significantly change the quality of the imagery. *Permission granted by subjects for use of imagery in publications.*

accuracy, and structural integrity during high winds.

Figure 2 shows image samples from some of the platforms used for the collection. The image quality can vary significantly between platform and also based on how the aircraft are configured and operated. The small quadcopters often produced very nice imagery being positioned very close to the subjects with high quality HD cameras. The R80D, Puma, and Supervolo were operated at much larger distances but also with more capable camera payloads which helped compensate for the distance.

2.2. Data

To investigate performance differences between aerial and close-range biometric verification, we conducted a systematic comparative analysis using the UAS and close-range partitions of the BRIAR datasets [6, 7], restricting to full-body captures and excluding probes with missing verification metrics or unspecified sex labels (16 probes removed from the aerial dataset and 17 from the close-range dataset). Metadata were organized into four categories: (1) *Algorithm and identification*: algorithm; collection site; sensor model; (2) *Video-derived characteristics*: head pixel height, head count, body count; (3) *Mission characteristics*: activity type, mission gait; (4) *Demographics*: age, sex, Hispanic origin, race, height (inches), weight (lbs).

The close-range surveillance dataset included 3,221 probes from 423 subjects, providing higher face resolution but greater scene complexity. The UAS dataset comprised 2,020 probes from 186 subjects, largely a subset of the ground-camera participants, collected concurrently at the same locations. Table 1 summarizes subject and probe characteristics.

Both datasets were processed by two algorithms, A and B, with roughly equal representation. UAS data were collected across five sites with six sensor models, and close-range data across six sites with six ground-based sensors[1]. UAS availability was primarily determined by weather, maintenance, or battery swaps.

Head pixel height distributions reflect imaging distance: UAS data contain more small heads (<40 pixels) and show greater variability in head and body counts, while close-range data include more large heads (>90 pixels) and are more constrained. Both datasets include standing, walking, and transitional activities, with gait annotations (movement state) available for roughly half of UAS probes and slightly fewer close-range probes.

Demographics are similar across datasets. UAS subjects are slightly older, close-range subjects slightly younger, with both predominantly White, non-Hispanic, and near-even male/female ratios. Height and weight distributions are comparable. These similarities ensure that differences

[1]Collection sites and dates included: BGC1 (Oak Ridge, TN; May–June 2021), BGC1.1 (Oak Ridge, TN; November 2021), BGC2 (Perry, GA; March–April 2022), BGC3 (Oak Ridge, TN; August–September 2022; close-range only), BGC4 (Chicago, IL; January 2023), and BGC5 (Bryan, TX; February 2024).

(a) UAS data.

Category	Subjects		Probes	
	N	% or Mean(SD)	N	% or Mean(SD)
Algorithm and Identification				
Algorithm				
… A	186	50	1007	49.9
… B	186	50	1013	50.1
Collection Site				
… BGC1	36	19.4	338	16.7
… BGC1.1	6	3.2	78	3.9
… BGC2	35	18.8	338	16.7
… BGC4	63	33.9	740	36.6
… BGC5	46	24.7	526	26
Sensor Model				
… Anafi	54	17.9	372	18.4
… Anafi USA	19	6.3	190	9.4
… EVO-II-Dual	4	1.3	14	0.7
… HD40-XV (Supervolo)	22	7.3	64	3.2
… HDZoom-30 (R80)	101	33.6	764	37.8
… Mantis-i45-EO (Puma)	101	33.6	616	30.5
Video-derived characteristics				
Head Height				
… Restricted	177	34.8	735	36.4
… <30 pix	81	15.9	328	16.2
… 30-40 pix	101	19.8	471	23.3
… 40-50 pix	49	9.6	140	6.9
… 50-60 pix	51	10	152	7.5
… 60-90 pix	28	5.5	110	5.4
… >90 pix	22	4.3	84	4.2
Head Count	186	*NA*	2020	145.92(125.4)
Body Count	186	*NA*	2020	264.21(117.87)
Mission characteristics				
Activity				
… Standing Static	116	32.3	621	30.7
… Walking Dynamic	172	47.9	1137	56.3
… Transitional / Other	71	19.8	262	13
Gait included				
… False	143	45.7	929	46
… True	170	54.3	1091	54
Demographics				
Age				
… <30 years old	41	22	444	22
… 30-50 years old	38	20.4	418	20.7
… 50-65 years old	30	16.1	288	14.3
… >65 years old	77	41.4	870	43.1
Sex				
… Female	104	55.9	1123	55.6
… Male	82	44.1	897	44.4
Hispanic				
… Hispanic or Latino	23	12.4	266	13.2
… Not Hispanic or Latino	163	87.6	1754	86.8
Race				
… White	144	77.4	1579	78.2
… Black or African American	24	12.9	241	11.9
… Some other race	18	9.7	200	9.9
Height (inches)	186	67.27(4.33)	2020	*NA*
Weight (lbs)	186	188.48(47.31)	2020	*NA*

(b) Close-range data

Category	Subjects		Probes	
	N	% or Mean(SD)	N	% or Mean(SD)
Algorithm and Identification				
Algorithm				
… A	423	50	1608	49.9
… B	423	50	1613	50.1
Collection Site				
… BGC1	73	17.3	560	17.4
… BGC1.1	15	3.5	134	4.2
… BGC2	95	22.5	714	22.2
… BGC3	82	19.4	424	13.2
… BGC4	98	23.2	841	26.1
… BGC5	60	14.2	548	17
Sensor Model				
… DWC-MPTZ336XW	25	2.8	56	1.7
… HDZP252DI	329	36.8	1342	41.7
… M3057-PLVE	42	4.7	150	4.7
… PNP-9200RH	156	17.4	570	17.7
… Q6215-LE	245	27.4	829	25.7
… QNP-6230H	98	10.9	274	8.5
Video-derived characteristics				
Head Height				
… Restricted	396	32.4	1031	32
… <30 pix	30	2.5	70	2.2
… 30-40 pix	91	7.5	210	6.5
… 40-50 pix	130	10.6	296	9.2
… 50-60 pix	106	8.7	244	7.6
… 60-90 pix	201	16.5	524	16.3
… >90 pix	267	21.9	846	26.3
Head Count	423	*NA*	3221	124.99(98.64)
Body Count	423	*NA*	3221	211.26(95.68)
Mission characteristics				
Activity				
… Standing Static	307	38.8	1118	34.7
… Walking Dynamic	374	47.2	1827	56.7
… Transitional / Other	111	14	276	8.6
Gait included				
… False	403	55	1923	59.7
… True	330	45	1298	40.3
Demographics				
Age				
… <30 years old	101	23.9	837	26
… 30-50 years old	138	32.6	998	31
… 50-65 years old	99	23.4	737	22.9
… >65 years old	85	20.1	649	20.1
Sex				
… Female	231	54.6	1754	54.5
… Male	192	45.4	1467	45.5
Hispanic				
… Hispanic or Latino	36	8.5	302	9.4
… Not Hispanic or Latino	387	91.5	2919	90.6
Race				
… White	341	80.6	2585	80.3
… Black or African American	43	10.2	336	10.4
… Some other race	39	9.2	300	9.3
Height (inches)	423	67.12(4.18)	3221	*NA*
Weight (lbs)	423	188.95(45.92)	3221	*NA*

Table 1. Summary of subject and probe characteristics for UAS and close-range datasets, including algorithm, collection site, sensor model, video-derived characteristics, activity, gait, and demographics.

in verification performance primarily reflect capture conditions rather than subject characteristics.

2.3. Outcome Verification

For each algorithm, verification attempts compared query (probe) videos against a reference gallery, the *simple gallery*, containing face and whole-body images and videos captured indoors under controlled conditions. Outdoor probe clips lasted 5–15 seconds, with subjects standing in various poses or walking along a straight path.

All attempts involved matched identity pairs under the prosecution hypothesis (H_0), with evidence evaluated relative to the alternative (H_1), where probe and gallery images come from different individuals. This framework allows systematic assessment of covariate effects on verification outcomes under realistic operational conditions.

Raw match scores were normalized to a common scale using the log-transformed false accept rate ($\log_{10}(\text{FAR})$) [4], facilitating comparisons across algorithms and inter-

pretation relative to raw similarity scores [11]. Verification was evaluated at three operational FAR thresholds: 1/100, 1/1,000, and 1/10,000. UAS data achieved 85.7%, 70.2%, and 54.2%, while close-range data performed higher at 96.3%, 90.4%, and 83.2% across the same thresholds.

2.4. Statistical Analysis

To quantify the effects of subject and acquisition factors on verification performance, we fit mixed-effects logistic regression models separately for the UAS and close-range datasets using identical structures. The outcome indicates whether probe j of subject i was correctly verified. Fixed effects included mission gait, head height category, sensor model, activity group, algorithm, collection ID, FAR, demographics (age, sex, race, ethnicity, height, weight), and scene-level covariates (head count, body count), while subject ID was modeled as a random intercept to account for intra-subject correlation [5, 12]. The log-odds of successful verification are:

$$\text{logit}(P(\text{verified}_{ij} = 1)) = \beta_0 + \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} + u_i, \qquad (1)$$

where $\mathbf{x}_{ij}$ is the predictor vector for probe j of subject i, and $u_i \sim \mathcal{N}(0, \sigma_u^2)$ is the subject-specific random intercept.

Model fit and discrimination were assessed using marginal and conditional R^2, AUC, and AIC, and a fixed-effects only GLM was also fit to quantify the contribution of random effects.

For UAS versus close-range comparisons, fixed-effect coefficients capture the conditional impact of each predictor, while adjusted predicted probabilities summarize average influence across the empirical covariate distribution[2].

3. Results

3.1. Model Fit and Comparison

In the UAV dataset, a fixed-effects model achieved AUC=0.812, marginal/full R^2=0.232, and AIC=5,749. Adding a random intercept for subjects improved AUC to 0.943, marginal/full R^2 to 0.402/0.783, and reduced AIC to 4,034. For the close-range dataset, the fixed-effects model yielded AUC=0.873, marginal/full R^2=0.289, and AIC=4,535; inclusion of subject-level random effects increased AUC to 0.971, marginal/full R^2 to 0.476/0.822, and lowered AIC to 3,421. Subject identity accounts for substantial variation, and mixed-effects models with random intercepts outperform fixed-effects models, highlighting the importance of modeling intra-subject correlation.

[2]Predicted probabilities were computed by marginalizing over dataset-specific covariates (e.g., sensor model, collection ID) while explicitly varying only covariates with identical support (e.g., activity, gait). This approach ensures that differences reflect modality-specific effects rather than unmatched sensor or site factors.

3.2. Drivers of Verification Performance

Mixed-effects logistic regression results (Table 2) highlight the operational and contextual factors associated with verification success in UAS and close-range imagery.

Head height, a proxy for image resolution, emerged as the strongest predictor in both datasets. In UAS imagery, subjects with head heights above 90 pixels exhibited an odds ratio of 60 (log-odds = 4.10, $p < 0.001$) relative to the reference category, whereas in close-range data the corresponding odds ratio exceeded 90 (log-odds = 4.51, $p < 0.001$). Effects were graded across categories, with even moderate increases in head pixel height substantially increasing verification odds. This underscores the critical role of image resolution, particularly for UAS-acquired faces that occupy fewer pixels.

Sensor model had a strong impact on verification outcomes. For UAV imagery, the HD40-XV sensor reduced verification odds by 98% (log-odds = -4.04, $p < 0.001$), while HDZoom-30 increased odds by 158% (log-odds = 0.95, $p < 0.001$). In close-range imagery, high-quality cameras such as PNP-9200RH and HDZP252DI yielded odds ratios of 8.2 (log-odds = 2.10, $p < 0.001$) and 4.05 (log-odds = 1.40, $p < 0.001$), respectively. These effects highlight the importance of sensor selection in operational verification performance.

Algorithm choice consistently influenced outcomes across both modalities. In UAS imagery, Algorithm B decreased verification odds (log-odds = -1.19, $p < 0.001$), while in close-range imagery the effect was slightly smaller but still significant (log-odds = -1.07, $p < 0.001$), reflecting meaningful differences in feature extraction, modeling, or fusion approaches.

Activity and motion characteristics played a modest role. Transitional or dynamic activities in close-range data were associated with higher verification odds (log-odds = 1.17, $p < 0.001$), whereas activity effects in UAV data were smaller and non-significant.

Demographic factors, including age, ethnicity, and race, did not have a significant impact in either dataset; however, male subjects in close-range imagery exhibited modestly higher verification odds (log-odds = 1.04, $p = 0.009$).

Figures 3–6 show modality-specific conditional effects under typical operating conditions, derived from model-based predicted probabilities for each covariate averaged over each dataset's distribution.

Figure 3 shows the effect of algorithm, collection site, and sensor model on verification probability. Algorithm B consistently yields lower predicted verification probabilities in both the UAS and close-range models. Sensor heterogeneity is evident for UAS imagery, with certain platforms (e.g., HD40-XV) exhibiting reduced predicted performance, whereas all close-range sensors show uniformly high predicted verification success. At the same site, UAS

(a) UAS coefficients.

Variable	Estimate	95% CI	Odds Ratio	95% CI	p-value	
Algorithm and Identification						
FAR						
… 1/100	—	—	—	—		
… 1/1,000	-1.933	-2.2, -1.7	0.14	0.11, 0.18	<0.001	***
… 1/10,000	-3.471	-3.7, -3.2	0.03	0.02, 0.04	<0.001	***
Algorithm						
… A	—	—	—	—		
… B	-1.191	-1.4, -1.0	0.3	0.26, 0.36	<0.001	***
Collection Site						
… BGC1	—	—	—	—		
… BGC1.1	0.076	-2.3, 2.4	1.08	0.11, 11.1	0.949	
… BGC2	-0.067	-1.5, 1.3	0.94	0.23, 3.78	0.925	
… BGC4	-1.666	-3.4, 0.04	0.19	0.03, 1.04	0.055	.
… BGC5	-0.792	-2.2, 0.60	0.45	0.11, 1.83	0.266	
Sensor Model						
… Anafi	—	—	—	—		
… Anafi USA	0.284	-0.84, 1.4	1.33	0.43, 4.09	0.620	
… EVO-II-Dual	-0.169	-1.8, 1.4	0.84	0.17, 4.18	0.836	
… HD40-XV	-4.043	-4.9, -3.2	0.02	0.01, 0.04	<0.001	***
… HDZoom-30	0.948	0.45, 1.4	2.58	1.57, 4.24	<0.001	***
… Mantis-i45-EO	-0.946	-1.4, -0.46	0.39	0.24, 0.63	<0.001	***
Video-derived characteristics						
Head Height						
… Restricted	—	—	—	—		
… <30 pix	1.079	0.75, 1.4	2.94	2.11, 4.10	<0.001	***
… 30-40 pix	1.257	0.95, 1.6	3.52	2.59, 4.77	<0.001	***
… 40-50 pix	1.053	0.63, 1.5	2.87	1.89, 4.36	<0.001	***
… 50-60 pix	1.387	0.93, 1.8	4	2.53, 6.33	<0.001	***
… 60-90 pix	1.982	1.4, 2.5	7.26	4.24, 12.4	<0.001	***
… >90 pix	4.102	3.1, 5.1	60.5	23.3, 157	<0.001	***
Head Count	0.002	0.00, 0.01	1	1.00, 1.00	<0.001	***
Body Count	0.002	0.00, 0.01	1	1.00, 1.00	<0.001	***
Mission characteristics						
Activity						
… Standing Static	—	—	—	—		
… Walking Dynamic	0.367	-0.32, 1.1	1.44	0.73, 2.87	0.294	
… Transitional / Other	0.094	-0.24, 0.42	1.1	0.79, 1.53	0.579	
Gait included						
… False	—	—	—	—		
… True	0.250	-0.42, 0.92	1.28	0.66, 2.51	0.464	
Demographics						
Age						
… <30 years old	—	—	—	—		
… 30-50 years old	-0.714	-2.1, 0.62	0.49	0.13, 1.86	0.295	
… 50-65 years old	-0.450	-1.9, 0.96	0.64	0.16, 2.61	0.531	
… >65 years old	-0.061	-1.8, 1.7	0.94	0.16, 5.42	0.945	
Sex						
… Female	—	—	—	—		
… Male	0.710	-0.28, 1.7	2.03	0.75, 5.50	0.161	
Hispanic Origin						
… Hispanic or Latino	—	—	—	—		
… Not Hispanic or Latino	0.278	-0.97, 1.5	1.32	0.38, 4.61	0.662	
Race						
… White	—	—	—	—		
… Black or African American	-0.373	-1.5, 0.76	0.69	0.22, 2.14	0.520	
… Some other race	1.007	-0.40, 2.4	2.74	0.67, 11.2	0.161	
Height (inches)	0.050	-0.07, 0.17	1.05	0.93, 1.18	0.402	
Weight (lbs)	0.007	0.00, 0.02	1.01	1.00, 1.02	0.148	

Baseline log-odds at reference levels: Intercept = -1.90, SE = 3.94, p = 0.63.
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

(b) Close-range coefficients.

Variable	Estimate	95% CI	Odds Ratio	95% CI	p-value	
Algorithm and Identification						
FAR						
… 1/100	—	—	—	—		
… 1/1,000	-1.804	-2.1, -1.5	0.16	0.12, 0.22	<0.001	***
… 1/10,000	-3.175	-3.5, -2.9	0.04	0.03, 0.06	<0.001	***
Algorithm						
… A	—	—	—	—		
… B	-1.066	-1.3, -0.86	0.34	0.28, 0.42	<0.001	***
Collection Site						
… BGC1	—	—	—	—		
… BGC1.1	0.161	-1.7, 2.1	1.17	0.18, 7.88	0.868	
… BGC2	-2.271	-3.3, -1.2	0.1	0.04, 0.30	<0.001	***
… BGC3	-2.604	-3.7, -1.5	0.07	0.02, 0.22	<0.001	***
… BGC4	-3.011	-4.1, -1.9	0.05	0.02, 0.14	<0.001	***
… BGC5	-1.454	-2.7, -0.20	0.23	0.07, 0.82	0.023	*
Sensor Model						
… DWC-MPTZ336XW	—	—	—	—		
… HDZP252DI	1.400	0.69, 2.1	4.05	2.00, 8.21	<0.001	***
… M3057-PLVE	0.183	-0.79, 1.2	1.2	0.45, 3.17	0.712	
… PNP-9200RH	2.104	1.2, 3.0	8.2	3.32, 20.2	<0.001	***
… Q6215-LE	1.226	0.46, 2.0	3.41	1.58, 7.33	0.002	**
… QNP-6230H	1.587	0.81, 2.4	4.89	2.24, 10.7	<0.001	***
Video-derived characteristics						
Head Height						
… Restricted	—	—	—	—		
… <30 pix	1.503	0.83, 2.2	4.5	2.28, 8.85	<0.001	***
… 30-40 pix	2.348	1.8, 2.9	10.5	6.04, 18.1	<0.001	***
… 40-50 pix	2.488	2.0, 3.0	12	7.12, 20.4	<0.001	***
… 50-60 pix	3.629	2.9, 4.3	37.7	18.6, 76.3	<0.001	***
… 60-90 pix	4.403	3.8, 5.0	81.7	43.4, 154	<0.001	***
… >90 pix	4.509	4.0, 5.0	90.8	54.3, 152	<0.001	***
Head Count	0.001	0.00, 0.01	1	1.00, 1.00	0.177	
Body Count	0.004	0.00, 0.01	1	1.00, 1.01	<0.001	***
Mission characteristics						
Activity						
… Standing Static	—	—	—	—		
… Walking Dynamic	1.065	0.62, 1.5	2.9	1.86, 4.52	<0.001	***
… Transitional / Other	1.169	0.65, 1.7	3.22	1.91, 5.42	<0.001	***
Gait included						
… False	—	—	—	—		
… True	-0.312	-0.75, 0.12	0.73	0.47, 1.13	0.161	
Demographics						
Age						
… <30 years old	—	—	—	—		
… 30-50 years old	-0.706	-1.6, 0.17	0.49	0.21, 1.19	0.115	
… 50-65 years old	-0.436	-1.4, 0.51	0.65	0.25, 1.66	0.365	
… >65 years old	-0.671	-1.7, 0.36	0.51	0.18, 1.43	0.200	
Sex						
… Female	—	—	—	—		
… Male	1.035	0.26, 1.8	2.81	1.29, 6.12	0.009	**
Hispanic Origin						
… Hispanic or Latino	—	—	—	—		
… Not Hispanic or Latino	0.072	-1.1, 1.3	1.07	0.32, 3.58	0.907	
Race						
… White	—	—	—	—		
… Black or African American	0.391	-0.61, 1.4	1.48	0.54, 4.02	0.444	
… Some other race	0.878	-0.32, 2.1	2.41	0.72, 8.01	0.152	
Height (inches)	-0.086	-0.18, 0.01	0.92	0.83, 1.01	0.080	.
Weight (lbs)	0.007	0.00, 0.01	1.01	1.00, 1.01	0.080	.

Baseline log-odds at reference levels: Intercept = 8.50, SE = 3.16, p = 0.007.
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 2. Fixed-effect coefficients quantify changes in the log-odds of verification success.

predictions are consistently lower than close-range.

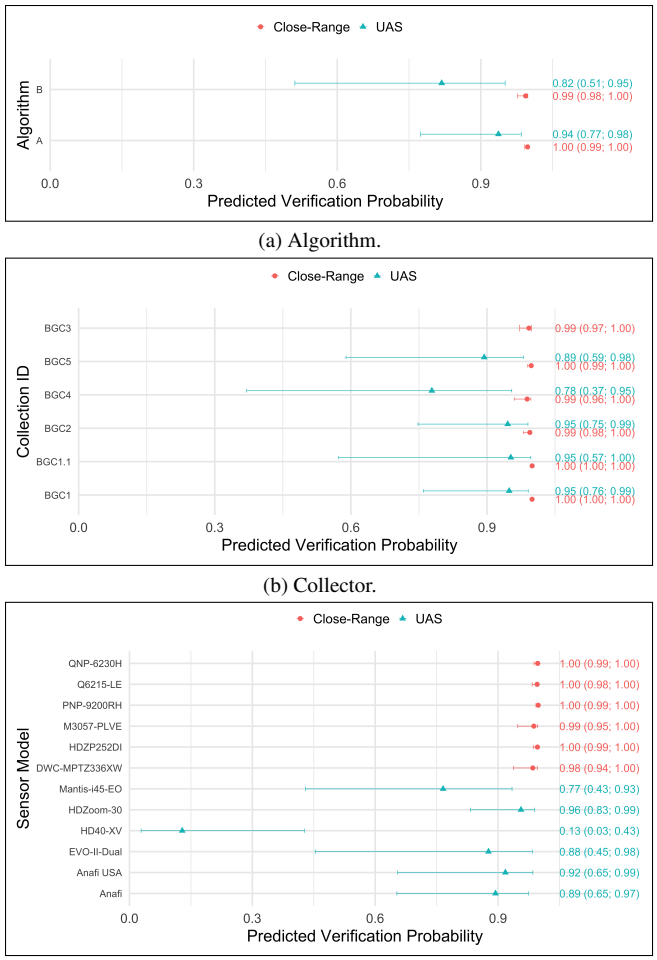
Figure 4 shows the dominant effect of head pixel height: higher resolution yields substantially higher predicted verification probabilities. Head and body counts also predict higher success for close-range imagery than UAS data.

Figure 5 shows that dynamic and transitional activities increase predicted verification probability in close-range data, whereas UAS predictions are slightly lower, reflecting reduced influence of subject pose in aerial captures.

Figure 6 shows minimal variation across demographic groups, with predicted probabilities consistently higher for close-range data, indicating limited demographic impact on verification success.

4. Discussion

Our analysis shows that multimodal biometric verification is highly context-dependent, with clear differences between UAS/aerial and close-range data. Close-range systems consistently outperformed UAS systems across all FARs, reflecting higher-resolution imagery, more controlled conditions, and reduced motion blur. UAS verification was more sensitive to head height, sensor type, and algorithm choice, highlighting the operational challenges of aerial acquisition. These findings align with prior work emphasizing the critical role of resolution and sensor quality in cross-domain recognition [4, 9]. *Image resolution*, measured via head height, was the strongest determinant of success in both modalities, with effects particularly pronounced for close-

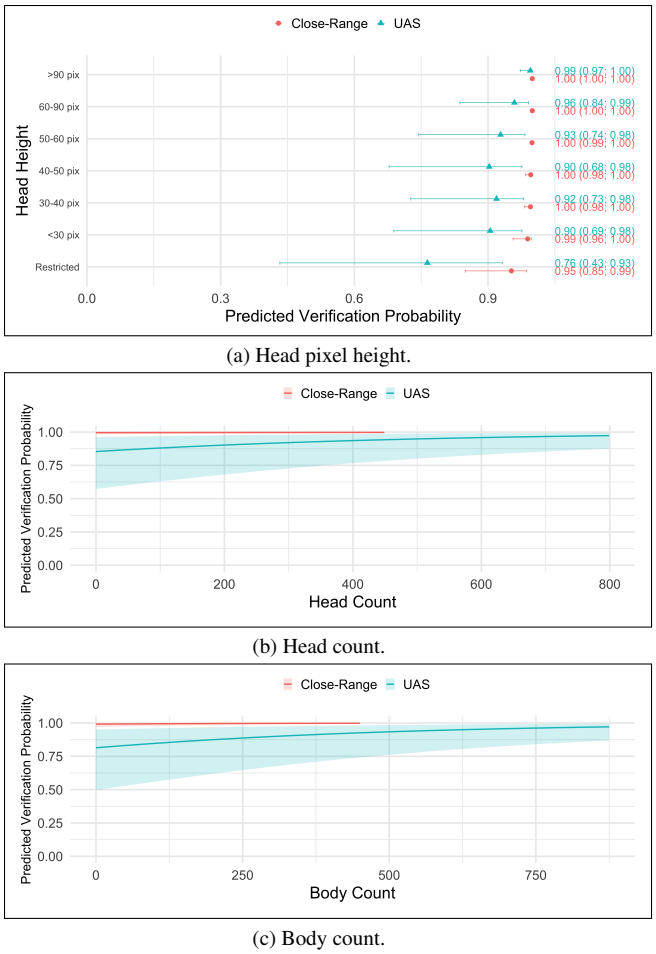


(a) Algorithm.

(b) Collector.

(c) Sensor model.

Figure 3. Model-based predicted probabilities of verification success across algorithm and identification characteristics.



(a) Head pixel height.

(b) Head count.

(c) Body count.

Figure 4. Model-based predicted probabilities of verification success across video-derived characteristics.

range data, where higher pixel density enabled extraction of discriminative facial and gait features. Sensor model also influenced performance, as some UAS sensors produced substantially lower verification odds, likely due to optical limitations, stabilization, and compression. Algorithm choice further contributed to variability, emphasizing the importance of feature extraction and multimodal fusion strategies. *Activity and motion* had limited impact on UAV verification but were meaningful for close-range imagery, with dynamic or transitional activities often increasing verification odds. *Demographic* factors, including age, ethnicity, and race, had minimal impact in both datasets.

4.1. Implications for System Design

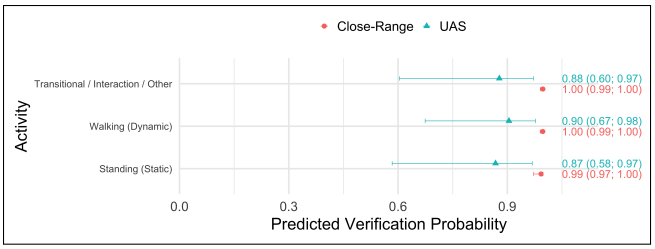
Our findings highlight key considerations for operational deployment of multimodal biometric systems. UAS-based systems require maximizing image resolution, selecting optimal sensors, and tailoring algorithms for low-resolution, motion-degraded imagery. Close-range systems benefit from leveraging dynamic cues such as gait and activity patterns. Across both modalities, accounting for intra-subject variability improves performance, as shown by mixed-effects models outperforming fixed-effects-only models.

4.2. Limitations

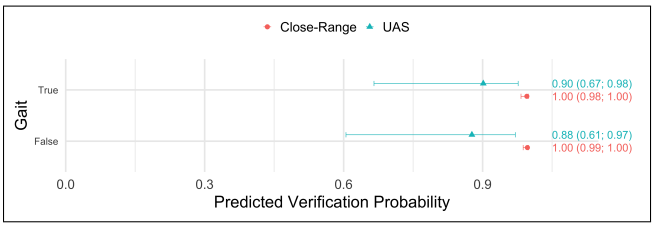
The findings may not generalize to all algorithms or sensor configurations. Environmental conditions, background complexity, and subject behavior could introduce unmeasured confounds. While head height was used as a proxy for image resolution, factors like illumination, lens distortion, or other environmental effects were not explicitly modeled.

4.3. Future Work

Future studies should expand cross-domain evaluation to additional UAS platforms, sensor types, and environmental conditions. Algorithmic improvements could explore domain adaptation, super-resolution, and motion compensation to enhance UAS verification. Dynamic multimodal fusion strategies that weight cues by image quality and context may further improve performance. Longitudinal



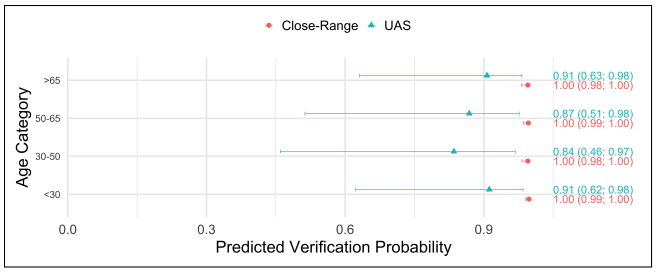
(a) Head pixel height.



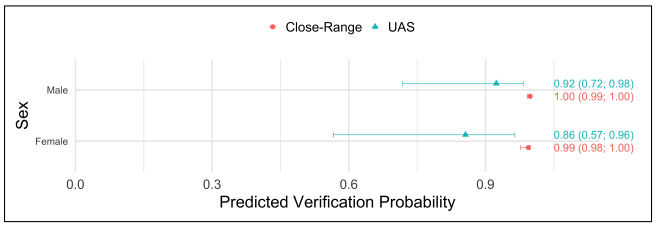
(b) Head count.

Figure 5. Model-based predicted probabilities of verification success across mission characteristics.

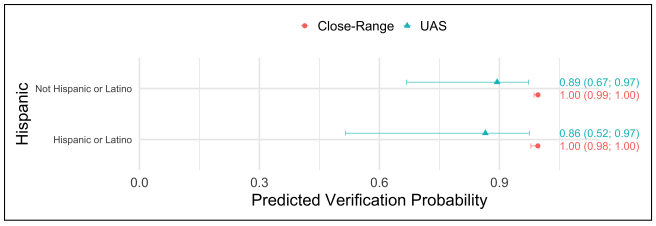
studies tracking verification over time and repeated deployments would provide insights into system reliability and robustness.



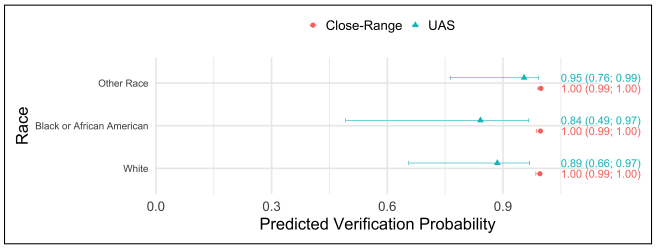
(a) Age.



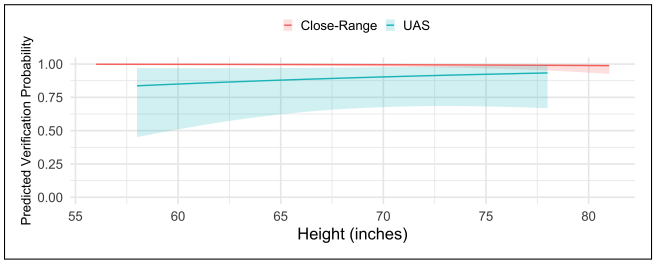
(b) Sex.



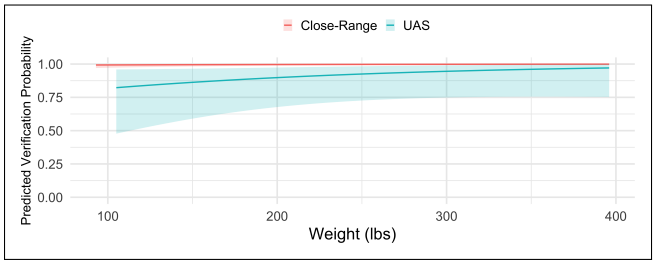
(c) Hispanic origin.



(d) Race.



(e) Height.



(f) Weight.

Figure 6. Model-based predicted probabilities of verification success across demographics characteristics.

References

[1] J Ross Beveridge, Geof H Givens, P Jonathon Phillips, Bruce A Draper, and Yui Man Lui. Focus on quality, predicting frvt 2006 performance. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–8. IEEE, 2008. 1

[2] J Ross Beveridge, Geof H Givens, P Jonathon Phillips, Bruce A Draper, David S Bolme, and Yui Man Lui. Frvt 2006: Quo vadis face quality. *image and Vision Computing*, 28(5):732–743, 2010. 1

[3] J Bhuvana, Amit Barve, Shah Pradeep Kumar, and Sukanya Dikshit. Image sensor fusion for multimodal biometric recognition in mobile devices. *Measurement: Sensors*, 36: 101309, 2024. 1

[4] David S Bolme, Deniz Aykac, Ryan Shivers, Joel Brogan, Nell Barber, Bob Zhang, Laura Davies, and David Cornett. From data to insights: A covariate analysis of the iarpa briar dataset for multimodal biometric recognition algorithms at altitude and range. In *2024 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–9. IEEE, 2024. 1, 4, 6

[5] Norman E Breslow and David G Clayton. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25, 1993. 5

[6] David Cornett, Joel Brogan, Nell Barber, Deniz Aykac, Seth Baird, Nicholas Burchfield, Carl Dukes, Andrew Duncan, Regina Ferrell, Jim Goddard, et al. Expanding accurate person recognition to new altitudes and ranges: The briar dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 593–602, 2023. 3

[7] Gavin Jager, David Cornett, Gavin Glenn, Deniz Aykac, Christi Johnson, Robert Zhang, Ryan Shivers, David Bolme, Laura Davies, Scott Dolvin, et al. Expanding on the briar dataset: A comprehensive whole body biometric recognition resource at extreme distances and real-world scenarios (collections 1-4). In *2025 IEEE 19th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–9. IEEE, 2025. 1, 3

[8] Yassin Kortli, Maher Jridi, Ayman Al Falou, and Mohamed Atri. Face recognition systems: A survey. *Sensors*, 20(2): 342, 2020. 1

[9] Feng Liu, Nicholas Chimitt, Lanqing Guo, Jitesh Jain, Aditya Kane, Minchul Kim, Wes Robbins, Yiyang Su, Dingqiang Ye, Xingguang Zhang, et al. Person recognition at altitude and range: Fusion of face, body shape and gait. *arXiv preprint arXiv:2505.04616*, 2025. 1, 6

[10] Blake A Myers, Lucas Jaggernauth, Thomas M Metz, Matthew Q Hill, Veda Nandan Gandi, Carlos D Castillo, and Alice J O'Toole. Recognizing people by body shape using deep networks of images and words. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2023. 1

[11] P Jonathon Phillips, W Todd Scruggs, Alice J O'toole, Patrick J Flynn, Kevin W Bowyer, Cathy L Schott, and Matthew Sharpe. Frvt 2006 and ice 2006 large-scale experimental results. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):831–846, 2009. 5

[12] Russ Wolfinger and Michael O'connell. Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation and Simulation*, 48(3-4):233–243, 1993. 5