

Long-Range Biometric Identification in Real World Scenarios: A Comprehensive Evaluation Framework Based on Missions

Deniz Aykac, Joel Brogan, Nell Barber, Ryan Shivers, Bob Zhang, Dallas Sacca, Ryan Tipton, Gavin Jager, Austin Garret, Matthew Love, Jim Goddard, David Cornett III and David S. Bolme

Oak Ridge National Laboratory
1 Bethel Valley Road, Oak Ridge, TN 37831, USA
aykacdb@ornl.gov

Abstract

The considerable body of data available for evaluating biometric recognition systems in Research and Development (R&D) environments has contributed to the increasingly common problem of target performance mismatch. Biometric algorithms are frequently tested against data that may not reflect the real world applications they target. From a Testing and Evaluation (T&E) standpoint, this domain mismatch causes difficulty assessing when improvements in State-of-the-Art (SOTA) research actually translate to improved applied outcomes. This problem can be addressed with thoughtful preparation of data and experimental methods to reflect specific use-cases and scenarios.

To that end, this paper evaluates research solutions for identifying individuals at ranges and altitudes, which could support various application areas such as counterterrorism, protection of critical infrastructure facilities, military force protection, and border security. We address challenges including image quality issues and reliance on face recognition as the sole biometric modality. By fusing face and body features, we propose developing robust biometric systems for effective long-range identification from both the ground and steep pitch angles. Preliminary results show promising progress in whole-body recognition. This paper presents these early findings and discusses potential future directions for advancing long-range biometric identification systems based on mission-driven metrics.

1. Introduction

The increasing complexity of modern security scenarios necessitates the development of advanced biometric identification systems that can operate effectively over a wide range of distances, altitudes, and environmental conditions.



Figure 1. Example dataset showcases a collage with two distinct sections. The left portion contains gallery images and video frames sourced from a controlled collection. The right section comprises various video frame probes.

The Intelligence Community (IC) requires such technology for various critical applications, including counterterrorism, protection of vital infrastructure facilities, military force protection, and border security. However, effective implementations of current biometric systems are hindered by several major challenges.

Firstly, image quality issues arising from factors such as motion blur, atmospheric turbulence, and resolution limitations significantly degrade system performance. Secondly, most algorithms have been developed based on low-pitch angle views of people, which may not be effective in scenarios characterized by significant pitch angles, such as those encountered when using UAVs or cameras mounted on buildings. Lastly, the over-reliance on face recognition

as the primary biometric modality limits system efficiency and accuracy, especially in situations where faces are not visible or when other body features provide more reliable identification cues.

To support research that addresses these challenges, this paper evaluates performance by current research solutions on face, body, and gait imagery targeting operationally relevant mission areas. These mission areas include close- and long-range face recognition, long-range recognition using multiple fused modalities, recognition from UAV platforms and elevated cameras, atmospheric turbulence mitigation, gait recognition, and identification without any full views of the face. Each mission area introduces a unique set of experimental problems that algorithms must handle to ensure robust system performance and accuracy.

We select video recorded from ranges up to 500m, pitch angles up to 30 degrees and evaluate five software systems specifically designed to process short video clips and perform biometric matching on whatever features are available. Unlike face recognition systems, these are built to handle complex input comprising views of the person from any angle with potential occlusions while still addressing quality issues inherent to long-range and elevated imagery.

The solutions evaluated in this paper were developed by independent research teams whose goal is to develop an advanced biometric identification system that can operate effectively under a wide range of conditions to support the IC's critical mission requirements.

In the subsequent sections, we provide a comprehensive evaluation framework based on missions. We begin with discussing related work on biometric recognition under challenging conditions. Then, we describe our methods of data collection and preparation for this evaluation, provide an overview of the unified application programming interface (API) we use to evaluate algorithm performance, and detail the mission areas selected for this work. Finally, we explain our analysis design, discuss results, and lay out a path forward to continue research as algorithms continue to mature.

2. Related Work

The limitations of accurate human identification at stand-off distances has been of interest of both the IC and research community for years. Earlier works such as [29] acknowledge that a major limiting factor is the lack of available datasets on which to train and evaluate new biometric models. Authors in [16] and [20] demonstrate the need for systems in low-resolution and unconstrained collection environments. Research such as [18], [25], [19], and [27] explore the impact that the number of pixels on the face as well as the quality of the face video itself have on facial recognition systems, which has spurred entities such as IARPA and DARPA to invest in datasets that can help improve bio-

metric performance at long ranges and lower qualities [9, 8]. The utility for facial recognition systems that can perform accurately at long distances is also addressed in [4] and [21].

In recent years, the prevalence of high-quality, lower cost UAVs has lead to their widespread usage for various applications. Those systems with mid-to-high quality imaging sensors lend themselves especially to use cases involving human detection, identification and tracking. Research in [2], [3], [23], [26], and [22] investigate pairing facial detection and recognition systems with these types of UAVs. Research as seen in [10] even acknowledges the limitations of existing biometrics systems to address the challenges of uncontrolled lighting, distance and pitches demonstrate the need for new algorithms to overcome these challenges.

The current related body of research coupled with the growing needs of the IC demonstrate the need for the advances detailed in this paper.

3. Data Preparation

This section lays out the three components that comprise our methods to evaluate state-of-the-art solutions for identifying individuals at ranges and altitudes. The Data Collection section details our endeavor to produce a high-quality, statistically relevant dataset to provide both training and testing data for biometric recognition systems. The Data Curation and Partitioning section provides insights into how the data is structured and organized to evaluate the various mission categories. Finally, the Common API section summarizes the custom API used across all solutions so that a fair and non-biased evaluation can be performed.

3.1. Data Collection

It is necessary to build a unique biometric image and video dataset to evaluate algorithm performance. This dataset consists of planned biometric collection events held in different parts of the United States during different seasons, ultimately resulting in diversity of subject demographics, terrain, and weather conditions. The dataset includes images and video of subjects in a variety of controlled and uncontrolled situations using a variety of sensors.

Subjects participated in the experiments only for one day and were required to wear different clothing sets. In the field, for example, the probes had clothing set 1 and gallery had clothing set 2 for the same subject. This ensures that algorithms are performing recognition based on persistent biometric signals from the face, body, and patterns of movement and not simply recognizing their clothing.

The indoor phase consists of face and whole-body images captured from multiple pitch angles up to 50° and the full rotation of yaw angles at 45° increments. This phase also includes video captured by COTS sensors placed in a semicircle around the collection area to record subjects walking and using their cell phones from multiple angles.

Subjects repeat the indoor collection process in two different sets of clothing.

During the outdoor data collection, subjects are recorded while they are standing still facing different directions, walking along prescribed paths, and performing a random walk. Conditions and types of data available may vary considerably based on characteristics of each collection site. Outdoor data is collected without regard for the prevailing weather conditions. Data is collected using close-range video cameras at various pitch angles from sensors loaded on as many as three different UAVs and from several long-range sensors set at a range of distances from 100m to 1,000m. The long-range sensors comprise a selection of COTS and industrial USB cameras equipped with specialized optics. At most distances, there is a camera set to capture the whole-body (WB camera) and a second camera set to capture a closer view of the face, which may not include the entire body (Face camera). The media samples from both indoor and outdoor collection are shown in Fig. 1.

3.2. Data Curation and Partitioning

In this section, we present the steps to curate the data from all the sensor measurements and to formalize the probe and gallery compositions for evaluation.

Step 1: During data collection, researchers use a custom desktop application to record timestamps for when each subject begins and completes a recorded activity. In the first step of post-processing, individual images and video segments are associated with a particular subject based on these timestamps. Longer videos containing footage of multiple subjects are cut along these timestamps so that the resulting dataset is comprised of videos of one subject performing one activity in one clothing set. This is made possible via precise time synchronization of all sensors. The length of video segments varies depending on the activity being performed and, in the case of some activities, the subject's mobility and natural walking pace.

Step 2: Additional data is recorded alongside videos during the data collection which describe weather conditions, atmospheric conditions, subject demographics, sensor positioning, and sensor hardware. This information is compiled such that each video segment can be associated with the corresponding metadata. Weather and atmospheric metadata that is tracked includes: temperature, wind chill, heat index, relative humidity, wind speed, wind direction, barometric pressure, and solar loading. This data, along with turbulent fluctuation recordings measured by a scintillometer, are recorded every minute during the data collection. Video segments are associated with the weather and atmospheric data aligned with the minute of its starting timestamp. Detailed sensor information such as minimum / maximum focal length, serial number, model number, manufacturer name, and sensor location are also included in the

metadata describing each video segment.

Step 3: Once the data has been organized as described in Steps 1 and 2, it is partitioned into training and test sets. This is done in a manner designed to create consistency of demographic distributions between the two sets. Every subject is assigned to either the training set or the test set, and the data is organized into its final directory structure. An XML file is generated for each piece of media to store the metadata described in Step 2 and the annotations described in Steps 4 and 5. A schema definition is used to validate XML files after metadata has been added to detect improperly formatted or corrupted data.

Step 4: Automated annotations are generated using a chain of open-source and pretrained models. WB detection is done with YOLOv5 and a fine-tuned version is used on long-range and aerial videos [14]. 3D human mesh reconstruction with Meshtransformer and 2D keypoint estimation with DARK is performed on the WB detection results [30, 17]. Re-ID with DG-Net++ is then performed on the pose results to determine whether or not a WB detection is the intended subject or not [31]. The pose information helps narrow the gallery to reference images at a similar yaw angle to the detections. BoT-SORT is used for track generation, which leverages the Re-ID results for better track consistency [1]. Finally, various post-processing steps are performed such as estimating the head bounding box from the 3D mesh and 2D keypoints.

Step 5: Select video frames are sent out for manual annotation either for validation or correction of suspected automated annotation errors. Mainly, these consist of a couple frames per track to verify the correct subject was denoted, a non-subject made it into the video, or the detection was a non-person. These results are then merged with the automated annotations. The metadata information such as type, method of collection along with examples are summarized in Table 1.

Step 6: The test set is further partitioned into gallery and probe sets for evaluation. Gallery media is used to build a database of known identities and probe media is imagery of subjects whose identities are presumed unknown. Probe imagery is compared to gallery enrollments and the results are used to compute the metrics we use to evaluate system performance. The evaluation gallery sets contain data for distractor subjects who are not part of the probes to simulate a larger gallery. We perform additional data partitioning for each specific mission area. These more granular designations are described in detail in Section 5. Balancing the subjects for mission categories was applied during the probe selection process but variation in the number of subjects and the samples per mission still exist as shown in Table 2.

Step 7: The probe set for mission analysis is organized into two major partitions: Face Restricted and Face Included. In the Face Restricted probe set, all faces are either

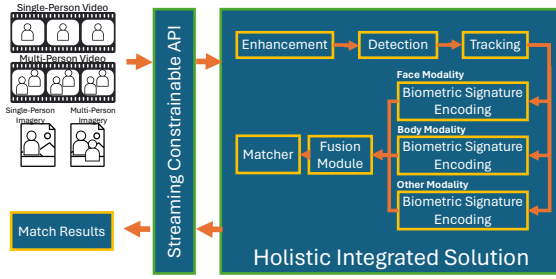


Figure 2. Diagram of an Integrated Holistic System (HS).

occluded, low resolution (<20 pixel head height), or not visible. The Face Included probe set contains data in which the face is visible and the head height is at least 20 pixels. The expectation is that there is at least one detectable face for every subject in every probe. Since width can vary more significantly with the changes of the yaw angle of the head, height is chosen as the basis for face detection. To establish whether the subject is facing the camera at any point during the 5-15 second video segment constituting a probe, we estimate the subject’s yaw angle relative to the camera. A subject is considered to be facing the camera if the yaw angle is estimated to fall between -110° and 110° . Two different techniques are utilized to achieve pose estimation: **Automatic:** The head portion of the 3D human mesh reconstruction is used to derive the yaw, pitch, and roll for person detected. Specifically, the Kabsch-Umeyama algorithm gives a rotation matrix from a reference 0 yaw, pitch, and roll head [15, 28]. **Activity timing:** Using RealSense cameras and elevated views, it is possible to track the subject’s position within the collection box. This identifies the different segments of the standing and structured walking activities. Directions during these segments are well known and can be used to label the expected yaw angle relative to each camera position with accuracy.

Step 8: Face Included and Face Restricted partitions are further split into treatment and control sets for evaluation. Control set consists of cameras positioned on the ground and at close range of less than 75m. For some evaluations, these cameras serve to set expectations for traditional performance. Most of these close-range cameras will have a wide field of view to capture the entire walking area and therefore may be at the low end of resolution needed for accurate face recognition. The treatment set consists of cameras from long-range, elevated, and UAV platforms. The treatment probes are given more weight in the probe selection process compared to the control probes as they are the focus of the project.

4. Common API

The API leveraged in this work took inspiration from previous works in [13], [12], [5], and [6, 7]. It was built for

this evaluation program and was guided by the same principles of maintaining operational relevance. To that end, the API has 3 main purposes: 1) To provide an algorithm harness with a front-end common set of callable functions that can abstract away complex back-end SOTA biometric algorithms, 2) to enforce strict standards on research code that help guarantee stable and predictable execution patterns, and 3) to operate as a data-passing middle-man, which allows the API to simulate the types of data and usage behaviors that would reflect those of a deployed system. Figure 2 shows a general overview of the API that is meant to interface with both streaming data and SOTA integrated biometric solutions to provide a realistic operational scenario. These end-to-end biometric solutions incorporate many internal algorithms across multiple biometric modalities, along with database-level storage and caching capabilities, to ultimately produce a single set of match scores for verification and search requests. For this reason, we refer to these SOTA solutions as Holistic integrated Solutions (HS). To evaluate an HS, the API ingests signature sets (sig-sets) similar to those introduced in [11] and used in numerous biometric challenges [24], which define and provide locations to subsets of video and imagery pertaining to the missions outlined in Section 5. The API then iterates over these sig-set-defined media subsets, loads the media, and subsequently streams the media to the HS. Because the API is constrainable, it can be configured to strip out, modify, or reformat media and its metadata to better simulate real-world operating scenarios. For instance, some scenarios may allow for metadata that provides camera information such as make, model, frame rate, operating conditions, etc., while other scenarios may need to simulate that metadata being unavailable to an HS at ingestion time.

5. Mission Areas

In this section, we explain the mission areas selected for this analysis and provide examples of the kinds of use cases that may be captured by each of them. Table 2 provides a brief overview of the mission areas discussed in this paper. Based on use case, some applications may focus on face recognition performance, whereas others might require the fusion of multiple modalities. The modality may also be determined by the biometrics available in galleries and watchlists. Currently, the barrier to entry for face recognition is relatively low due to the high availability of both frontal face imagery and highly accurate software systems. Other modalities such as body and gait are as yet less sophisticated and require significant performance advancement before they will be widely adopted. There is also a much smaller corpus of ground truth data for these modalities. Evaluating performance for these specific mission areas may also inform areas in which additional research is needed to achieve suitable accuracy for deployment in a real

Table 1. Types of metadata collected

Metadata Type	Method of Collection	Examples
Weather	At collection	Temp, humidity, wind speed
Demographics	At collection	Age, gender, height
Timing	At collection	Video start, end time
Camera	At collection	Make, model, resolution, focal length
UAV	At collection	Alt., GPS, camera telemetry
Face Bounding Boxes	Video processing	Location, size, tracks
Whole Body Bounding Boxes	Video processing	Location, size, tracks
Subject Identity	Annotation	Confirmation of identity
Additional Bounding Boxes	Annotation	Other objects and bounding boxes
Occlusion Information	Annotation	Info on occlusion of face/body
Clothing Information	Annotation	Type of clothing, shoes, accessories

world application.

Experimental Control: We use imagery captured by high resolution surveillance cameras positioned at eye-level. This relatively "easy" data is used to establish a performance baseline.

Close Range Face: This area focuses on imagery with high resolution, frontal profile views of the face, simulating environments like building entrances, choke points, and security checkpoints. Typically, these areas are well-suited to face recognition because there is a high degree of certainty that people will pass through a contained space and face a certain direction. We use close range video from cameras positioned at eye-level or slightly elevated.

Close Range Body: This mission area is a super set of Close Range Face that includes all views of the face along with body and gait signatures. These videos are taken from high resolution surveillance cameras at close-range positioned at eye-level or slightly elevated. This mission can be used as a baseline for comparing the effects of distance in biometric performance.

Long Range Face: This category consists of data captured using long range cameras on the ground that are configured for facial imagery at a resolution suitable, though perhaps not ideal, for recognition. Long range face recognition is an active research area and encompasses surveillance use cases requiring that cameras be positioned at significant distances from their intended subjects. We are particularly interested in any recognition improvements that may be observed when leveraging body and gait imagery.

Long Range Body (Fusion): This area includes all ground sensors at long range that are poorly configured for biometric recognition. This group is a super set of the long range face mission area which also includes body and gait, highlighting the improvements achieved by fusing these additional modalities along with the face.

UAV: Imagery captured from UAV platforms introduces unique challenges relating to distance, pitch angle, and platform size and weight. Consistent, accurate performance us-

ing this data would constitute a large step forward in the realm of biometric recognition and would support many surveillance and security use cases in which UAV platforms are already used or static camera equipment on the ground would be intractable.

Turbulence: Atmospheric conditions can degrade outdoor image quality significantly and lead to poor recognition performance. Turbulence visibly distorts imagery, particularly at medium and long ranges, and may challenge any biometric application which relies on data captured outside from any nontrivial distance. This problem varies based on environmental conditions like climate, weather, and terrain.

Elevated Cameras: This area focuses on close range surveillance cameras positioned to look down from rooftops or masts at pitch angles exceeding 12° . In many instances, this is the most suitable way to place cameras due to space constraints or concern over damage to surveillance equipment. Extreme pitch angles are challenging since they do not capture the frontal or eye-level face imagery that typically produces the most accurate recognition.

Gait: Gait is a relatively nascent biometric modality when compared to those that are more established and widely adopted like the face. However, advancing gait recognition performance will serve many applications not only on its own but in conjunction with modalities like the face and body. Even as an auxiliary signal, gait patterns may provide more accurate recognition capabilities in challenging uncontrolled imaging conditions.

Face Restricted: This area targets data from long-range cameras, elevated cameras, and cameras mounted to UAV platforms. Faces are totally or partially occluded, observed from extreme pitch angles, or are of lower resolution than 20 pixels in height. Improving recognition in the absence of clear facial imagery would be advantageous for a wide range of use cases in which subjects are captured in an uncontrolled environment not suitable for close range, ground level equipment installation. While previous research has shown that incorporation of degraded facial imagery can

improve identification performance [16], recognition under these conditions still requires that algorithms not rely on facial imagery but instead make accurate predictions using modalities such as body and gait.

No biometric modality or system is suitable for all applications. Stakeholders may care about one or two particular mission areas and will focus on results for those use cases. For example, many organizations are interested in recognition from cameras mounted on building rooftops while others may be better served by long-range systems. We selected the mission areas detailed above because they encompass many recognition applications and use cases for which there is need by government stakeholders.

6. Analysis Design

Evaluating biometric systems based on specific mission areas informs development of applications targeting use cases that fall within these areas. It is useful to consider factors that may improve performance in one domain while hindering it in another. We design experiments targeting these mission areas in greater isolation than would be captured by a more general performance evaluation. Additionally, this analysis design is based on the idea that, under challenging conditions, systems will perform better when they can leverage imagery of both the face and the body than facial imagery alone. To this end, these experiments measure the performance of the selected algorithms on face, body, and/or WB fusion (face, body, and gait) recognition tasks using a dataset that consists of two distinct probe sets, Face Included and Face Restricted. For the purposes of evaluation we use probe videos from field data and gallery enrollments from controlled data. Probe videos contain long-range or elevated views for a single subject. After processing a video, a labeled database entry for the subject is made. It is assumed that each entry could have multiple tracklets and each tracklet would be associated with a template matching an entry in the gallery. Gallery enrollments are a collection of templates associated with subjects. An enrollment produces a single template entity that can be searched against and matched. And the template encapsulates features extracted from one or more pieces of input imagery from a subject. The imagery could be face and/or WB images and videos. Evaluations are implemented by creating probe and gallery databases. API commands are provided to verify against these databases. A ROC curve, which plots the true accept rate (TAR) vs the false accept rate (FAR) over the range of thresholds, is generated for each mission area. The experiments can be run in different modes such as face only, body only, gait only and fusion to highlight the strengths and weaknesses of systems as well as to show improvements, if any, when fusion is selected. As mentioned in Section 5, the main focus of this evaluation analysis is to use mission areas to better utilize the data to

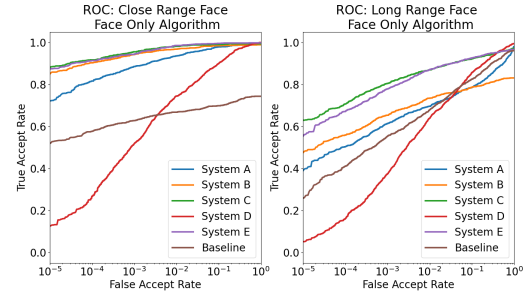


Figure 3. Close range, long range - face performance results

analyze recognition performances with experimental control as the baseline for comparison. In addition, investigating recognition performance between two data treatments, such as close range and long range, highlights the effects of certain factors such as distance.

7. Results and Discussion

Face Recognition: The Close Range Face mission serves as a basis for comparison and consists of high-resolution imagery representative of high-quality video from current surveillance camera deployments. Face recognition algorithms should be able to perform well with existing technology. The Long Range Face mission presents additional challenges related to image quality, allowing us to test both the basic functionality of the algorithms and understand how new developments are improving accuracy in more challenging scenarios. ROC results are compared in Figure 3, which shows promising performance.

Whole Body Matching: One of the primary goals of this research is to develop techniques for recognizing individuals based on their whole-body appearance. Combining face recognition with body geometry, gait, and other features should improve matching performance. This is particularly important given the image quality challenges in our dataset. Figure 4 compares results in this area and shows how performance decreases under different levels of difficulty. Close range shows very high accuracy, while longer ranges show systems performing well despite being poorly configured for recognition. Face restricted performance is surprisingly good given that these videos were selected in a way that well-established face recognition algorithms lacked usable faces, requiring the algorithms to rely heavily on body and gait features.

Improvements from Fusion: Another critical area of interest is how the fusion of different biometric modalities improves recognition accuracy. Figure 5 illustrates how face only, body only, and gait only recognition is combined to produce improved fusion accuracy. The body only case outperforms the face only case for most of the systems and fusion of all modalities help the overall recognition performance. Based on this analysis, the challenges that come from identifying individuals at long range would be greatly

Table 2. Each category’s combination of working distance/angle and biometric modality is reflective of a specific application of interest.

Mission	Description	Distance	Modality	Subjects/ Samples
Experimental Control	<i>High- resolution surveillance cameras at eye-level</i>	3.8m-17.2m	Face, Body, Gait	256/2282
Close Range Face	<i>High-resolution surveillance cameras on the ground or slightly elevated, frontal and profile angles of the face</i>	3.8m	Face	245/1199
Close Range Body	<i>High-resolution surveillance cameras on the ground or slightly elevated</i>	3.8m	Face, Body, Gait	250/1719
Long Range Face	<i>Cameras configured for face recognition on the ground</i>	>250m	Face	226/1351
Long Range Body (Fusion)	<i>Cameras designed for long range but poorly configured for biometric recognition</i>	>250m	Face, Body, Gait	258/3166
UAV	<i>Video from UAV platforms at varied elevation, pitch angle, movement speed</i>	-	Face, Body, Gait	44/238
Turbulence	<i>Long range video reflecting medium to high turbulence</i>	>250m	Face, Body, Gait	232/3272
Elevated cameras	<i>Close range cameras at severe pitch angles</i>	5.8m-12.9m	Face, Body, Gait	242/1532
Gait	<i>Video of walking sequences from all angles</i>	>3.8m	Gait	260/3744
Face Restricted	<i>Video lacking sufficient facial quality for face recognition</i>	>5.8m	Body, Gait	260/2078

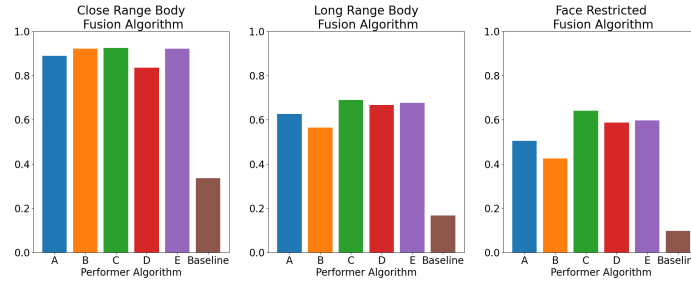


Figure 4. Close range body, long range body, face restricted - fusion results compared for all systems with the performance metric of TAR @ 1% FAR

reduced when all modalities are utilized. By analyzing individual modalities as well as fused modalities across a variety of missions, we can better understand the contributions of each modality as well as the value of fusion.

Elevated and UAV: At this early stage of the program, elevated and UAV performance have not yet been a key focus. However, providing tests in these areas early on is intended to guide future research in this field.

Current results offer a snapshot of performance, showing that developed algorithms significantly outperform the baseline, which has not been updated to handle high pitch angle data. Comparing the Experimental Control (Close-range Ground) test to the Elevated Mission (Close-range Elevated) illustrates the challenges posed by these scenarios as seen in Figure 6. The UAV mission is currently intended to focus development on some of the most difficult challenges in this research program. We anticipate seeing significant improvements in these areas in the next year of research.

Turbulence and Gait: The turbulence and gait missions are used to focus algorithm development on two specific problems: gait recognition and turbulence mitigation. The

gait mission provides a test set where all media contains people walking, providing a good way to test gait recognition in isolation on a challenging dataset. Likewise, the turbulence mission focuses on data known to be at long distances and with significant amounts of turbulence. By splitting these problems out, it allows us to focus research and development in these specific areas, benchmark progress over time, and compare different solutions. Results are shown in the ROC plots in Figure 7 which demonstrate current performance on these two problems. It should be noted that the Baseline algorithm uses Body and Gait recognition algorithms that predate the data collection and are known to be inadequate for the level of difficulty of this data set. System A also did not include a gait algorithm, but provided body features that still perform competitively with its peers.

8. Conclusion and Future Work

The design of thoughtful experiments that help move the needle on SOTA biometric algorithms requires a tight marriage between physical data collection processes and the detailed curation and partitioning of imagery and video

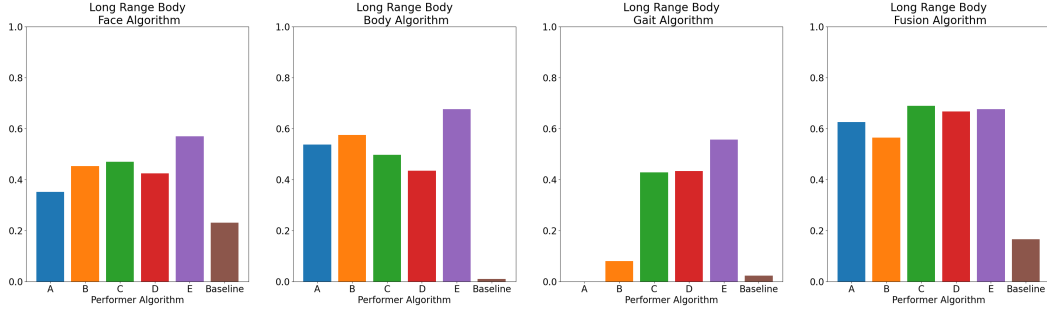


Figure 5. Results from the system with individual modalities isolated vs the fusion. All algorithms are run on the long range body mission. The performance metric compared is TAR @ 1% FAR

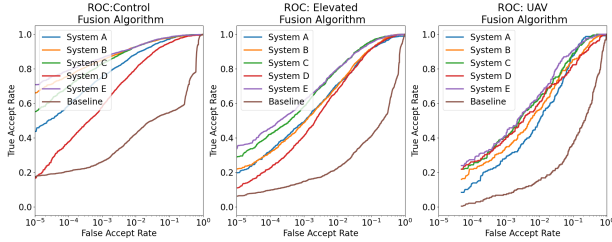


Figure 6. Control, Elevated, UAV - fusion performance results

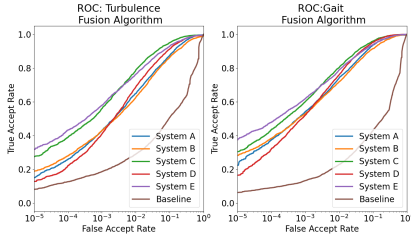


Figure 7. Turbulence, Gait - fusion performance results

therein. We hope this paper acts as a thought-provoking case study into that process, along with providing a rich analysis of anonymous SOTA algorithms tested against our experimental protocols.

Our evaluation methods addresses challenges related to image quality and focuses on reducing the dependence on face recognition as the sole modality by providing data and tests focused on whole body matching. The evaluation utilized a diverse range of hardware solutions including COTS cameras, custom long-range cameras, and UAVs. Preliminary results indicate promising progress in WB recognition.

In future work, we intend to further address challenges related to image quality and dependence on face recognition as the primary modality in long-range biometric identification systems. We aim to promote development of robust systems that are less susceptible to these drawbacks. We also plan to focus on video recordings featuring larger groups of individuals as well as more intricate and dynamic scenarios. By doing so, we hope to continue encouraging development of more accurate and reliable methodolo-

gies for WB recognition technology. We are confident that these advanced systems will have significant implications for various mission areas, including law enforcement and national security. Further leveraging WB recognition capabilities will potentially improve situational awareness, public safety, and operational efficiency.

Acknowledgements

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via D20202007300010. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

This research used resources from the Knowledge Discovery Infrastructure at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725

Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

References

- [1] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking, 2022.
- [2] S. Aruna Kumar, E. Yaghoubi, A. Das, B. Harish, and H. Proença. The p-destre: A fully annotated dataset for pedestrian detection, tracking, re-identification and search from aerial devices. *arXiv e-prints*, pages arXiv–2004, 2020.
- [3] K. Bálint. Uavs with biometric facial recognition capabilities in the combat against terrorism. In *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, pages 000185–000190. IEEE, 2018.
- [4] D. S. Bolme, H. J. Santos-villalobos, and D. C. Cornett III. Rifle-like camera for long distance face recognition, us patent 11574501, 2023. US Patent 11574501.
- [5] D. S. Bolme, N. Srinivas, J. Brogan, and D. Cornett. Face recognition oak ridge (faro): A framework for distributed and scalable biometrics applications. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–8. IEEE, 2020.
- [6] J. Brogan, N. Barber, D. Cornett, and D. Bolme. Vdisc: An open source framework for distributed smart city vision and biometric surveillance networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 148–154, 2023.
- [7] J. Brogan, B. Zhang, D. Bolme, and M. Ngan. Ornl/briar-api. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2022.
- [8] P. Calhoun. Darpa emerging technologies. *Strategic Studies Quarterly*, 10(3):91–113, 2016.
- [9] D. Cornett, J. Brogan, N. Barber, D. Aykac, S. Baird, N. Burchfield, C. Dukes, A. Duncan, R. Ferrell, J. 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.
- [10] J. Diez-Tomillo, J. M. Alcaraz-Calero, and Q. Wang. Dynamic-distance-based thresholding for uav-based face verification algorithms. *Sensors*, 23(24):9909, 2023.
- [11] P. Flanagan. Ice 2006 frequently asked questions. <https://www.nist.gov/itl/iad/image-group/ice-2006-frequently-asked-questions>, 2011.
- [12] M. Golec, S. S. Gill, R. Bahsoon, and O. Rana. Biosec: A biometric authentication framework for secure and private communication among edge devices in iot and industry 4.0. *IEEE Consumer Electronics Magazine*, 11(2):51–56, 2020.
- [13] M. Günther, R. Wallace, and S. Marcel. An open source framework for standardized comparisons of face recognition algorithms. In *Computer Vision—ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7–13, 2012, Proceedings, Part III 12*, pages 547–556. Springer, 2012.
- [14] G. Jocher, Ayush Chaurasia, A. Stoken, J. Borovec, NanoCode012, Yonghye Kwon, TaoXie, Kalen Michael, Jiacong Fang, Imyhxy, , Lorna, C. Wong, (Zeng Yifu), Abhiram V, D. Montes, Zhiqiang Wang, C. Fati, Jebastin Nadar, Laughing, UnglvKitDe, Tkianai, YxNONG, P. Skalski, A. Hogan, M. Strobel, M. Jain, L. Mammana, and Xylieong. ultralytics/yolov5: v6.2 - yolov5 classification models, apple m1, reproducibility, clearml and deci.ai integrations, 2022.
- [15] W. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 34(5):827–828, 1978.
- [16] P. Li, J. Brogan, and P. J. Flynn. Toward facial re-identification: Experiments with data from an operational surveillance camera plant. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2016.
- [17] K. Lin, L. Wang, and Z. Liu. End-to-end human pose and mesh reconstruction with transformers. *CoRR*, abs/2012.09760, 2020.
- [18] L. S. Luevano, L. Chang, H. Méndez-Vázquez, Y. Martínez-Díaz, and M. González-Mendoza. A study on the performance of unconstrained very low resolution face recognition: Analyzing current trends and new research directions. *IEEE Access*, 9:75470–75493, 2021.
- [19] Q. Meng, S. Zhao, Z. Huang, and F. Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14225–14234, 2021.
- [20] J. Neves, F. Narducci, S. Barra, and H. Proença. Biometric recognition in surveillance scenarios: a survey. *Artificial Intelligence Review*, 46:515–541, 2016.
- [21] U. Park, H.-C. Choi, A. K. Jain, and S.-W. Lee. Face tracking and recognition at a distance: A coaxial and concentric ptz camera system. *IEEE transactions on information forensics and security*, 8(10):1665–1677, 2013.
- [22] M. Rostami, A. Farajollahi, and H. Parvin. Deep learning-based face detection and recognition on drones. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–15, 2022.
- [23] R. N. Sarath, J. T. Varghese, and F. Pandya. Unmanned aerial vehicle for human tracking using face recognition system. In *2019 Advances in Science and Engineering Technology International Conferences (ASET)*, pages 1–5. IEEE, 2019.
- [24] W. J. Scheirer, P. J. Flynn, C. Ding, G. Guo, V. Struc, M. Al Jazaery, K. Grm, S. Dobrizek, D. Tao, Y. Zhu, et al. Report on the btas 2016 video person recognition evaluation. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2016.
- [25] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch. Face image quality assessment: A literature survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–49, 2022.
- [26] K. Shanthi, P. Sivalakshmi, et al. Smart drone with real time face recognition. *Materials Today: Proceedings*, 80:3212–3215, 2023.
- [27] P. Terhöst, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper. Pixel-level face image quality assessment for explainable face recognition. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2023.
- [28] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(04):376–380, 1991.
- [29] Y. Yao, B. R. Abidi, N. D. Kalka, N. A. Schmid, and M. A. Abidi. Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Computer Vision and Image Understanding*, 111(2):111–125, 2008.
- [30] F. Zhang, X. Zhu, H. Dai, M. Ye, and C. Zhu. Distribution-aware coordinate representation for human pose estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [31] Y. Zou, X. Yang, Z. Yu, B. V. Kumar, and J. Kautz. Joint disentangling and adaptation for cross-domain person re-identification. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 87–104. Springer, 2020.