A Comparative and Fusion-Based Analysis of Body Silhouette and Clothing Color for Real-Time Person Re-Identification

José Luis Carrillo-Medina[1][0000-0001-5222-4039], Ivone Arias-Almeida[1][0000-0001-7351-2013], Marcelo Alvarez V.[1][0000-0002-2781-2506], Edwin Quinatoa[1][0000-0001-9701-7463], Nancy Jacho Guanoluisa[1][0009-0004-9452-3810], and Marco Flores Calero[1][0000-0003-1597-3385]

Universidad de las Fuerzas Armadas ESPE, Sangolquí, Ecuador `{jlcarrillo, giarias, rmalvarez, eequinatoa3, jpjacho, mjflores}@espe.edu.ec`
http://www.espe.edu.ec

Abstract. Person re-identification (Re-ID) in non-overlapping multi-camera environments remains a challenging problem, particularly when facial information is unavailable or unreliable due to occlusions, viewpoint changes, and illumination variability. While recent advances in deep learning have achieved state-of-the-art performance, they typically require large-scale datasets and high computational resources, limiting their applicability in real-time and resource-constrained scenarios.

This paper proposes a holistic and interpretable architecture for real-time person identification and re-identification that integrates biometric and soft-biometric features within a unified pipeline. The system performs initial identification using facial features represented through Histogram of Oriented Gradients (HOG) and classified with Support Vector Machines (SVM), while subsequent re-identification relies on body silhouette and clothing color modeled using HOG descriptors and HSV color histograms. Person and face detection are conducted using YOLOv8n, enabling efficient operation under constrained computational conditions.

Experimental results demonstrate that the proposed approach achieves high performance, with an F1-score of 0.9835 in the re-identification stage when combining silhouette and color features. Comparative analysis shows that body silhouette provides the most stable discriminative representation, while color information acts as a complementary cue that improves classification balance when fused with structural features.

The main contribution of this work lies in the integration of biometric and soft-biometric modalities within a single interpretable framework that maintains identity continuity when facial evidence is unavailable. Unlike deep learning-based approaches, the proposed system offers a computationally efficient and explainable alternative suitable for real-time surveillance applications in constrained environments.

Keywords: Person re-identification · soft-biometrics · body silhouette · clothing color · YOLOv8n · HoG · SVM · multi-camera video surveillance.

1 Introduction

Person re-identification has become a central problem in video surveillance analytics [2] [3], particularly for maintaining identity continuity across multiple observation points without assuming visual continuity between scenes [1]. Recent studies indicate that the problem extends beyond appearance association and requires integrating detection, visual representation, and supervised decision-making under heterogeneous operational conditions [14].

This challenge is especially significant in multi-camera surveillance systems, where identity must be preserved despite changing and incomplete visual evidence [15]. The difficulty increases in non-overlapping camera environments [12], where illumination changes [4], perspective variations, partial occlusions, and inconsistent facial visibility complicate the re-identification process [16].

Under these conditions, person detection alone is insufficient to guarantee identity continuity. Effective solutions require computationally efficient pipelines capable of detecting people and faces while relying on complementary attributes when facial evidence becomes unreliable [17]. Consequently, YOLO-family detectors [7] and lightweight computational approaches have become relevant for near-real-time surveillance systems [8].

The literature differentiates biometric traits, such as facial information, from soft-biometric attributes, including body silhouette, texture, and clothing color, which have demonstrated relevance in multi-camera re-identification tasks [18]. Within this context, handcrafted descriptors remain valuable due to their interpretability and controlled computational cost [9].

In particular, Histogram of Oriented Gradients (HoG) has been widely applied to capture structural and contour information in recognition tasks [20], while Local Binary Patterns (LBP) and related texture descriptors have been used to model appearance features [10]. Similarly, color histograms continue to provide an effective representation of clothing chromatic distributions.

Supervised classifiers, especially Support Vector Machines (SVM), remain suitable for scenarios involving explicit descriptors and moderate-scale datasets [11] [19]. These approaches support interpretable and computationally feasible computer vision systems for visual recognition tasks.

Based on these foundations, this study proposes an integrated system for person identification and re-identification in video surveillance [21]. Initial identification is performed using facial features, while subsequent re-identification relies on body silhouette and clothing color within a holistic representation framework. The architecture combines YOLOv8n for person and face detection, HoG for structural representation, HSV color histograms for chromatic modeling, and SVM for final classification [22].

The main contribution of this work is the integration of a biometric facial identification stage with a soft-biometric re-identification stage based on body silhouette and clothing color for non-overlapping multi-camera environments [5] [6]. Experimental validation in a real capture environment demonstrates that the system achieves operational viability without depending on highly complex deep-learning architectures [24]. Overall, the work contributes to the intersection of person re-identification, holistic visual representation, and interpretable computer vision and machine learning systems for intelligent surveillance [23] [25].

2 Materials and Methods

The methodology integrates the identification and re-identification of individuals in video surveillance through a sequential and interpretable architecture. The first level uses facial evidence for initial identification; when the face is unavailable or unreliable, the second level maintains identity continuity using soft-biometric attributes such as body silhouette and clothing color. The workflow does not operate as isolated modules, but as a single chain of detection, feature extraction, descriptive combination, and classification.

2.1 System Architecture

The system tracks individuals in a camera network through person detection, facial detection, biometric feature extraction, and soft-biometric attribute extraction. YOLOv8n locates faces and bodies in real time; HoG represents the face and silhouette; HSV describes clothing color; and SVM makes the final identity decision. The architecture dynamically shifts the weight of the evidence: face for direct identification and silhouette-color for re-identification when facial information is partial, rear-view, or absent; (see Fig. 1).

Figure 1 shows the operational sequence: a) a person is detected, b) the face is detected, and c) facial features are extracted, d) a person is identified using a previously built model, followed by e) body detection, f) extraction of soft-biometric attributes (silhouette and clothing color) to train a re-identification model, and f) a person is re-identified using this model when the face is not available.

2.2 Materials, Environment, and Tools

The implementation was developed in Python due to its integration with computer vision and machine learning. OpenCV was used for image manipulation and preprocessing; scikit-learn for SVM training and feature vector management; PyTorch for data augmentation; NumPy for numerical computation; Matplotlib for visualization; and YOLOv8n for face and person detection. The design accounts for execution in a constrained computational environment without a dedicated GPU, so low-complexity descriptors and efficient classifiers are prioritized.

2.3 Face and Person Detection

Regions of interest are localized using YOLOv8n, selected for its efficiency and ability to detect multiple objects under varying lighting conditions and viewpoints. Facial detection triggers the initial identification; body detection feeds

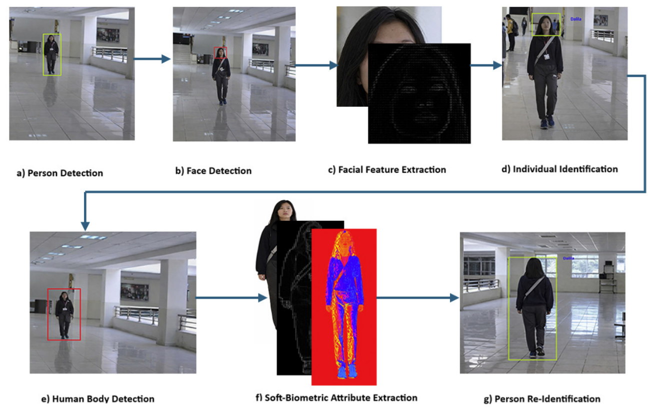

Fig. 1. Functional scheme of the re-identification system

the re-identification module when the face does not provide sufficient evidence. This stage determines the quality of the entire pipeline, as it defines the regions to which preprocessing, feature extraction, and classification are applied.

2.4 Dataset and Data Augmentation

The dataset contains five classes, each associated with an identity. For facial identification, 50 images per class were initially used, for a total of 250 images; subsequently, through data augmentation, this was expanded to 100 images per class, reaching 500 samples (see Fig. 2). For re-identification, 100 full-body images per class were used, with 500 initial images; after augmentation, the dataset reached 200 images per class and 1,000 training samples, (see Fig. 3). Transformations included rotations and brightness adjustments to increase variability, strengthen generalization, and simulate changes in orientation and lighting without altering the number of identities.

Fig. 2 shows the faces of 5 people (classes) used to train and validate biometric identification using the HoG descriptor and the SVM classifier; Fig. 3 shows the body images used for re-identification based on silhouette and clothing color using SVM. (For the identification and re-identification of individuals, see the sections: Feature Extraction and Integrated Representation and Classification).

2.5 Preprocessing

Preprocessing standardizes the visual input and stabilizes the descriptive representation. Facial images are resized to 80×100 pixels, and body images to 64×160 pixels. For HoG, the image is converted to grayscale, while for color analysis, it is transformed into HSV color space. Additionally, filtering and smoothing are applied to reduce irrelevant variations, improve the quality of the descriptors, and provide consistent vectors to the classifier.

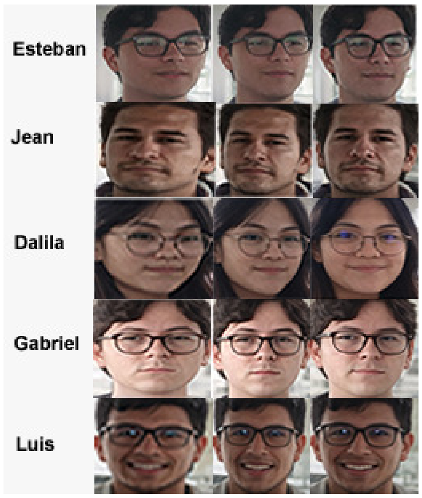

Fig. 2. Face dataset used in the facial identification stage.

Fig. 3. Body dataset used in the re-identification stage.

2.6 Feature Extraction

Facial identification uses Histogram of Oriented Gradients (HoG), a descriptor that summarizes the local distribution of gradient orientations and preserves structural information about facial contours and shapes. The image is divided into cells; gradient direction histograms are calculated and grouped and normalized by blocks to increase robustness against changes in contrast and lighting.

In re-identification, the body silhouette is also modeled using HoG to capture edges, contours, and the individual's overall configuration. The HoG descriptor configuration reports nine orientations and 2×2 cell blocks; for the face, a cell size of 13×13 pixels is used, and for the full body, a differentiated configuration determined by the structural template used is employed. The logic, in both cases, is to represent visual geometry using normalized gradient histograms.

On the other hand, the color of the clothing is described using HSV histograms; this color space separates hue, saturation, and value, as it provides a more stable response than RGB to certain lighting variations. The parameters used for the Hue, Saturation, and Value histograms are eight bins per channel, with ranges of 0–180 for Hue and 0–256 for Saturation and Value.

2.7 Integrated Representation and Classification

The system is based on a global visual representation, defined as a complementary combination of the face, body shape, and color within a single stream. For re-identification, the HoG (body silhouette) and HSV (clothing color) descriptors are concatenated into a single representation (vector); this allows the classifier to distinguish classes based on body structure and chromatic appearance. This integration forms the core of the person Re-ID module, as it provides continuity in the person's identity when the face is not visible.

Classification is performed using Support Vector Machines (SVM), for both facial identification and re-identification. The model learns decision boundaries between identities based on the extracted vectors. The primary methodological configuration specifies a linear kernel and regularization C=0.1, a value retained as an explicit reference since it is derived from the base methodological document.

2.8 Methodological Validation

The evaluation incorporates metrics suitable for scenarios with potential imbalances between matches and non-matches; the metrics Accuracy, Precision, Recall, and F1-score are applied to characterize confusion errors and identity recovery capability. The balance between correct detections and incorrect matches is determined using the Precision-Recall curve. Additionally, in identification and re-identification tasks, Mean Average Precision (mAP) is used to measure the quality of the retrieval ranking as a central metric in multi-camera scenarios, since performance is not limited solely to the classifier's isolated accuracy.

3 Results

This section presents the experimental results. First, the behavior of the facial identification module as the primary biometric reference is examined. Next, re-identification is analyzed as a complementary mechanism to sustain identity continuity when facial information is unavailable. In this second stage, three modalities are compared: clothing color, body silhouette, and silhouette + color fusion.

3.1 Facial identification results

The facial identification module was evaluated with 936 images distributed across five classes and captured at different locations within the camera network. This design introduces significant illumination variations, including scenes with high brightness, pronounced shadows, and intermediate conditions. Consequently, the observed performance reflects not only classification capability but also stability against capture heterogeneity. The overall metrics obtained were Accuracy = 0.9989, Precision = 0.9986, Recall = 0.9992, F1-score = 0.9989, and mAP = 0.9978 (see Table 1).

The confusion matrix confirms this trend. The main diagonal concentrates all correct classifications for Dalila, Gabriel, Jean, and Luis, while only a single confusion is observed for the Esteban class, classified once as Jean (see Table 2). From a discriminative perspective, this behavior indicates that the facial representation based on HoG and classification using SVM maintain highly separable decision boundaries between identities, even under lighting variations.

Table 1. Global metrics of the facial identification model.

Metrics	Value
Accuracy	0.9989
Precision	0.9986
Recall	0.9992
F1-score	0.9989
mAP	0.9978

Table 2. Confusion matrix of the facial identification model.

Real class / predicted	Dalila	Esteban	Gabriel	Jean	Luis
Dalila	180	0	0	0	0
Esteban	0	249	0	1	0
Gabriel	0	0	145	0	0
Jean	0	0	0	146	0
Luis	0	0	0	0	215

3.2 Re-identification results

After testing the face module, re-identification was evaluated using soft-biometric cues when faces were unavailable. Three methods were compared: clothing color, body silhouette, and their fusion. Color alone performed worst (Accuracy = 0.7331, F1 = 0.7410). Silhouette performed much better (Accuracy = 0.9744, F1 = 0.9751). Fusion achieved top results (Accuracy = 0.9831, F1 = 0.9835, mAP = 1.00) (see Table 3). Silhouette is the most reliable shape cue; color is only complementary.

Table 3. Comparative global metrics of re-identification.

Modality	Accuracy	Precision	Recall	F1-score	mAP
Color	0.7331	0.7433	0.7506	0.7410	0.81
Silhouette	0.9744	0.9779	0.9732	0.9751	0.99
Silhouette + color	0.9831	0.9856	0.9818	0.9835	1.00

At the matrix level, color alone shows clear under and overestimation between classes. Esteban and Jean are underestimated (-202 and -316), while Luis (+281), Gabriel (+146), and Dalila (+91) are overestimated. Precision is lowest for Luis (0.4175), and recall is lowest for Jean (0.5187). Main confusions: Jean→Luis (362), Esteban→Gabriel (162), Esteban→Luis (102). Thus, color alone yields unstable boundaries under clothing or lighting variations.

Silhouette greatly reduces class reassignment bias, with frequency differences shrinking to a range of -36 to +34. Dalila and Luis retain nearly all diagonal mass (recall 0.9981 and 1.0000). Esteban and Jean concentrate residual error: Jean shows very high precision (0.9959) but recall drops to 0.9275, shifting samples to Gabriel (20) and Luis (15). Gabriel reaches 0.9878 recall but 0.9484 precision

Table 4. Confusion matrix of color-based re-identification.

Real class / predicted	Dalila	Esteban	Gabriel	Jean	Luis
Dalila	402	10	24	0	5
Esteban	50	459	162	2	102
Gabriel	69	26	568	7	36
Jean	9	12	4	417	362
Luis	2	66	94	62	362

due to external samples from Esteban and Jean. Thus, silhouette is morphologically stable but slightly overestimates classes with similar body geometry (see Table 5).

Table 5. Confusion matrix of silhouette-based re-identification.

Real class / predicted	Dalila	Esteban	Gabriel	Jean	Luis
Dalila	530	0	0	1	0
Esteban	0	566	24	1	3
Gabriel	0	6	808	0	4
Jean	2	1	20	486	15
Luis	0	0	0	0	545

The silhouette + color fusion delivers the most balanced behavior of the entire system and further attenuates underestimation and overestimation by class. The discrepancies between actual and predicted frequency are confined between -26 and +26 samples, indicating a calibration much closer to the observed distribution. Dalila is classified without false negatives, and Jean achieves a precision of 1.0000, meaning that whenever the system predicts this class, it does so correctly; however, its recall remains at 0.9494 because deviations toward Gabriel (12) and Luis (11) still exist. Gabriel retains the largest volume of false positives among the main classes, with 31 additional samples, suggesting that it continues to function as a recipient class for borderline cases. Even so, its recall rises to 0.9939 and its precision to 0.9636, confirming that the incorporation of color does not destabilize the structure provided by the silhouette but rather refines it and corrects part of the residual errors (see Table 6).

Table 6. Confusion matrix of silhouette + color-based re-identification.

Real class / predicted	Dalila	Esteban	Gabriel	Jean	Luis
Dalila	530	0	0	0	0
Esteban	0	469	17	0	1
Gabriel	0	3	821	0	2
Jean	2	1	12	488	11
Luis	0	0	2	0	553

Modalities differ mainly in error distribution, not just total correct classifications. Color causes asymmetric, massive sample reassignment: some identities are strongly overestimated, others underestimated. Silhouette fixes most of this via consistent structural organization. Fusion further improves interclass calibration, reduces reassignment bias, and localizes residual errors. Thus, identity continuity relies primarily on body shape; appearance or texture works only as fine-tuning, not standalone evidence.

3.3 Comparative re-identification figures

The figures not only illustrate an overall improvement in performance but are also consistent with the error structure shown by the confusion matrices.

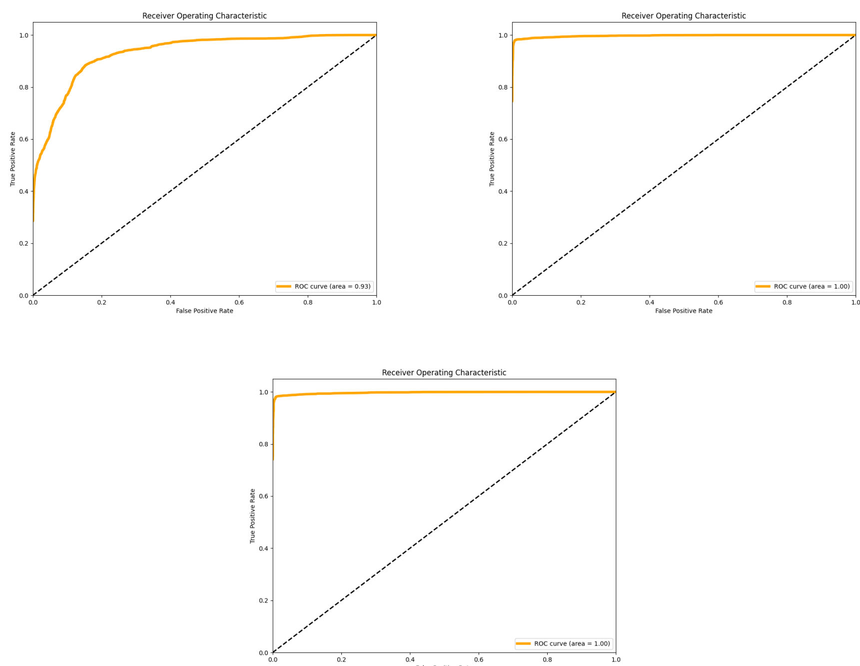

Fig. 4. Comparison of ROC curves for color, silhouette, and silhouette + color.

In Fig. 4, color alone shows an ROC curve far from the top-left corner (AUC = 0.81), reflecting the error dispersion seen in its confusion matrix: massive sample redistribution toward Gabriel and Luis, and poor recovery of Esteban and Jean. Thus, color fails to separate positives from negatives stably under clothing or lighting variations. In contrast, silhouette and fusion stay near the top edge (AUC = 0.99 and 1.00), consistent with diagonal-dominated matrices and few interclass transfers. Hence, the ROC gain is not decorative—it reflects real reductions in false positives and negatives

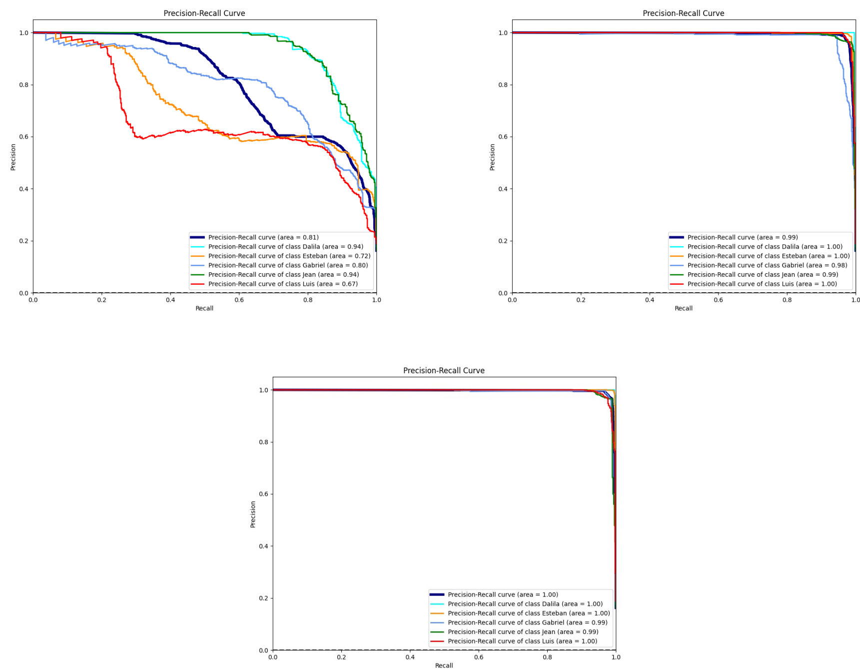

Fig. 5. Precision-Recall curves per class for color, silhouette, and silhouette + color.

Fig. 5 reinforces this via precision-coverage. Color's PR curve declines steadily as recall increases, meaning more correct detections require accepting more false positives—consistent with overestimating Luis (+281) and Gabriel (+146), and underestimating Jean (-316) and Esteban (-202). Silhouette stays near precision 1.0 across most of the range, dropping only at the extreme end, matching localized residual error and stable interclass assignment. At class level: color shows heterogeneous curves with early declines and large separations between identities (unbalanced boundaries). Silhouette clusters near the upper limit, though Gabriel and Jean show minor degradations. Fusion keeps almost all curves near ideal, reflecting a more concentrated diagonal, reduced bias, and homogeneous errors. Thus, figures align with confusion matrices: the hybrid modality corrects error geometry and stabilizes operational separation.

4 Discussion

Overall, the results support the feasibility of the proposed system as an integrated identification and re-identification architecture in non-overlapping cameras. Its main contribution does not lie in solving both tasks separately but in articulating them within a single operational flow, where facial identification acts as the initial biometric anchor and re-identification based on body silhouette and clothing color sustains identity continuity when the face is no longer usable. This functional transition is consistent with the problem posed in the manuscript, since the real difficulty consists not only of recognizing a person in an isolated scene but of maintaining their traceability under changes in view, lighting, and capture conditions.

From a methodological standpoint, the observed behavior can be explained by how the architecture distributes visual evidence according to its availability. YOLOv8n not only fulfills the detection function but also enables the transition between modules: when the face is visible, the system operates on biometric evidence; when this evidence becomes unreliable, recognition relies on soft-biometric attributes. In this context, HoG plays a dual role, as it describes both facial structure and the global geometry of the body silhouette, while HSV introduces chromatic information about clothing. Classification with SVM, moreover, maintains an interpretable and computationally controlled decision framework, which is especially relevant in a constrained, near-real-time environment.

The facial identification stage is clearly the most stable component of the system. Metrics close to perfection and a practically diagonal confusion matrix indicate that HoG + SVM generates well-separated decision boundaries when facial evidence is available. More importantly, this stage establishes the initial identity with very low uncertainty, which reduces the ambiguity burden that re-identification must subsequently absorb. The presence of a single isolated confusion suggests that the facial module is not completely immune to error but is sufficiently robust to function as the biometric foundation of the pipeline. In technical terms, this confirms that the structural representation of the face provides a more specific and stable discriminative basis than the body signals used in the second stage.

Re-identification, on the other hand, becomes relevant when the system can no longer rely on the face and must sustain identity through body signals. Here, the comparison between color, silhouette, and silhouette + color fusion is decisive [12] [13]. The chromatic descriptor alone shows the weakest performance, suggesting that clothing appearance is an informative source but highly sensitive to similarities between individuals and lighting variations. The body silhouette provides a much more stable basis, as it encodes the global shape of the individual and considerably reduces the ambiguity observed with color. However, the fusion of HoG and HSV achieves the best balance, indicating that body structure constitutes the primary support for separability, while color acts as a complementary signal that refines decision-making in cases where shape alone does not capture all visual variability.

The analysis of underestimation, overestimation, and interclass transfer reinforces this interpretation. In the color-only modality, some identities lose a significant portion of their actual samples, while others absorb excessive external examples, revealing an unbalanced decision pattern and strong discriminative fragility. In other words, isolated color not only performs worse in global terms but also introduces prediction biases that favor certain classes as recipients of ambiguity. The silhouette corrects much of this problem, as it reduces the discrepancy between actual and predicted frequency, although it still maintains small exchanges between physically close classes. The fused modality further improves this behavior: overestimations and underestimations are reduced, erroneous transfers become more localized, and interclass balance becomes more consistent. This is technically relevant because it shows that the advantage of HoG + HSV is not limited to increasing aggregated metrics but improves the quality of the classifier's decision pattern.

The ROC and Precision-Recall curves are consistent with this interpretation. The color modality exhibits a geometry farther from the ideal edge, corresponding to its greater instability in separating positives and negatives. The silhouette, in contrast, shows an organization closer to ideal behavior, confirming that body structure provides a more solid discrimination basis. The fusion consolidates this trend and brings the curves even closer to the upper limit, which coincides with a confusion matrix more concentrated on the diagonal and with more homogeneous residual error. Therefore, the hybrid modality not only achieves more correct classifications but also makes errors in a less biased and more controlled manner, which is especially important in re-identification systems where the stability of decision-making between classes is as relevant as overall performance.

In summary, the evidence from the file supports the main contributions of the manuscript: an integrated and explainable architecture that combines biometric facial identification with soft-biometric re-identification based on body silhouette and clothing color, maintaining operational viability under computational constraints. The most relevant technical discussion shows that the real strength of the system lies not in color alone but in the complementarity between HoG applied to the silhouette and HSV applied to clothing. At the same time, the results also delimit the scope of the study: the observed robustness depends on the evaluated network, the number of identities, and the stability of visual appearance. Therefore, the work provides solid evidence of feasibility and interpretability in the analyzed scenario, but its future generalization should be tested under broader populations, greater lighting variability, and more severe changes in clothing appearance.

5 Conclusions

The study supports the feasibility of an integrated person identification and re-identification system in non-overlapping cameras, in which detection with YOLOv8n articulates initial identification based on the face and subsequent re-identification supported by body silhouette and clothing color within a holistic and interpretable architecture. The findings obtained are consistent with the logic of the implemented system and show that the integration between HoG, HSV, and SVM constitutes a consistent methodological basis for sustaining identity continuity when facial evidence is no longer available. In this framework, the contribution of the work lies in the articulation of facial biometrics and soft-biometrics as parts of the same operationally viable pipeline, rather than in the isolated use of each module. Although the scope of the results remains delimited by the conditions of the evaluated environment, the study leaves as a reasonable projection the validation of the approach in scenarios with greater variability of subjects, lighting, and appearance.

Acknowledgments The research project entitled "Development of video surveillance models and/or algorithms for Re-Identification of persons based on soft biometric features in a closed-circuit camera using computer vision and machine learning tech-niques" is the focus of this study. This project was supported by the Research Group on Intelligent Systems, Data Ana-lytics, Robotics, Cybersecurity and Software (SIARCyS) for their valuable assistance in the implementation of these models. In addition, we extend our thanks to the Universidad de las Fuerzas Armadas ESPE for their support during the realization of this research effort.

References

1. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proc. IEEE Int. Conf. Comput. Vis. (ICCV), pp. 1116–1124 (2015). https://doi.org/10.1109/ICCV.2015.133
2. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.: Deep learning for person re-identification: A survey and outlook. IEEE Trans. Pattern Anal. Mach. Intell. **44**(6), 2872–2893 (2021). https://doi.org/10.1109/TPAMI.2021.3054775
3. Fu, Y., Wei, Y., Zhou, Y., Shi, H., Huang, G., Wang, X., Yao, Z., Huang, T.: Horizontal pyramid matching for person re-identification. In: Proc. AAAI Conf. Artif. Intell. vol. 33(01), pp. 8295–8302 (2019). https://doi.org/10.1609/aaai.v33i01.33018295
4. Khan, S.U., Khan, N., Hussain, T., Muhammad, K., Hijji, M., Del Ser, J., Baik, S.W.: Visual appearance and soft biometrics fusion for person re-identification using deep learning. IEEE J. Sel. Top. Signal Process. **17**(3), 575–586 (2023). https://doi.org/10.1109/JSTSP.2023.3241514
5. Gavilanes, P., Carrillo, J., Galarza, E.: Person Re-identification System in a Controlled Environment Based on Soft Biometric Features: Clothing Color and Body Silhouette Collected on Short Video Sequences Using Computer Vision and Machine Learning Algorithms. In: Proc. 3rd Int. Conf. Inf. Syst. Softw. Technol. (ICI2ST), pp. 14–20 (2022). https://doi.org/10.1109/ICI2ST57357.2022.00011
6. Carrillo-Medina, J., Chango-Caisabanda, D., Cuyo-Chiluisa, V., Galarza-Medina, E.: Person Re-identification using soft-biometric features: body silhouette and clothing texture in a multi-camera video surveillance environment. In: Proc. 3rd Int. Conf. Inf. Syst. Softw. Technol. (ICI2ST), pp. 1–8 (2022). https://doi.org/10.1109/ICI2ST57357.2022.00009

7. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 779–788 (2016). https://doi.org/10.1109/CVPR.2016.91
8. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
9. Zdravevski, E.: Unsupervised weed detection in spinach seedling organic farms. In: Proc. Int. Conf. ICT Innovations, pp. 1–12. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86249-7_3
10. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002). https://doi.org/10.1109/TPAMI.2002.1017623
11. Bishop, C.M.: Pattern Recognition and Machine Learning. Information Science and Statistics, Springer, New York (2006)
12. Peng, Y.X., Li, Y., Zheng, W.S.: Revisiting person re-identification by camera selection. IEEE Trans. Pattern Anal. Mach. Intell. **46**(5), 2692–2708 (2023). https://doi.org/10.1109/TPAMI.2023.3331454
13. Bi, Y., Wang, R., Zhou, Q., Lin, R., Wang, M.: Enhancing person re-identification through attention-driven global features and angular loss optimization. Entropy **26**(6), 436 (2024). https://doi.org/10.3390/e26060436
14. Wang, X., Zhao, R.: Person re-identification: System design and evaluation overview. In: Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.) Person Re-Identification, pp. 351–370. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6296-4_17
15. Senelle, M., Garcia-Diez, S., Mantrach, A., Shimbo, M., Saerens, M., Fouss, F.: The sum-over-forests density index: Identifying dense regions in a graph. IEEE Trans. Pattern Anal. Mach. Intell. **36**(6), 1268–1274 (2013). https://doi.org/10.1109/TPAMI.2013.195
16. Shang, C., Barnes, D.: Fuzzy-rough feature selection aided support vector machines for mars image classification. Comput. Vis. Image Underst. **117**(3), 202–213 (2013). https://doi.org/10.1016/j.cviu.2012.10.006
17. Possegger, H., Mauthner, T., Bischof, H.: In defense of color-based model-free tracking. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2113–2120 (2015). https://doi.org/10.1109/CVPR.2015.7298823
18. Czyz, J., Sadeghi, M., Kittler, J., Vandendorpe, L.: Decision fusion for face authentication. In: Zhang, D., Jain, A.K. (eds.) Biometric Authentication. ICBA 2004. LNCS, vol. 3072, pp. 686–693. Springer, Berlin, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25948-0_94
19. Zhao, G., Hassan, S., Zou, Y., Truong, D., Corbin, T.: Predicting performance anomalies in software systems at run-time. ACM Trans. Softw. Eng. Methodol. **30**(3), 1–33 (2021). https://doi.org/10.1145/3440757
20. Déniz, O., Bueno, G., Salido, J., De la Torre, F.: Face recognition using histograms of oriented gradients. Pattern Recognit. Lett. **32**(12), 1598–1603 (2011). https://doi.org/10.1016/j.patrec.2011.01.004
21. Jemni, S.K., Kessentini, Y., Kanoun, S.: Out of vocabulary word detection and recovery in Arabic handwritten text recognition. Pattern Recognit. **93**, 507–520 (2019). https://doi.org/10.1016/j.patcog.2019.05.005
22. Diwan, T., Anirudh, G., Tembhurne, J.V.: Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimedia Tools Appl. **82**(6), 9243–9275 (2023). https://doi.org/10.1007/s11042-022-13628-6

23. Tjoa, E., Guan, C.: A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE Trans. Neural Netw. Learn. Syst. **32**(11), 4793–4813 (2020). https://doi.org/10.1109/TNNLS.2020.3027314
24. Tan, M., Le, Q.: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proc. 36th Int. Conf. Mach. Learn. (ICML). PMLR, vol. 97, pp. 6105–6114 (2019)
25. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., Gao, J.: Deep learning–based text classification: a comprehensive review. ACM Comput. Surv. **54**(3), 1–40 (2021). https://doi.org/10.1145/3439726