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

Gavilanes Puente, Pamela Michell

Departamento de Ciencias de la Computación

Carrera de Software

Artículo académico, previo a la obtención del título de Ingeniera en Software

Dr. Carrillo Medina, José Luis

19 de agosto de 2022

Latacunga

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

Pamela Gavilanes-Puente
Computer Science Department
Universidad de las Fuerzas Armadas - ESPE
Sangolquí, Ecuador
pmgavilanes1@espe.edu.ec

José Carrillo-Medina
Computer Science Department
Universidad de las Fuerzas Armadas - ESPE
Sangolquí, Ecuador
jlcarrillo@espe.edu.ec

Eddie E. Galarza
Electrical and Electronics Department
Universidad de las Fuerzas Armadas - ESPE
Sangolquí, Ecuador
eegalarza@espe.edu.ec

***Abstract*—Person re-identification is one of the most critical activities in the security area, specifically in video-surveillance since it has wide applications such as access control, people tracking and behavior detection. In this paper, a system of Re-Identification of people through 3 stages is proposed. The first one, detection and segmentation of people using Mask-RCNN method, the second, feature extraction with convolutional neural networks (CNN), and finally, the identification of people in different places with a multi-input neural network model and an output composed of a CNN. The model uses two types of descriptors based on soft-biometric appearance features, body silhouette and color in RGB space. These are treated and handled independently by deep learning techniques, which allows to generate as output the identification of persons. The experiments are carried out with a dataset created in a controlled environment by capturing videos with 2 counterposed cameras. Through a detailed comparison and the analysis of different models with different accuracy metrics, it can be indicated that the fusion of the silhouette and color features improve the solution robustness, than when treated individually. In terms of accuracy metrics, training time and validation, the multiple input model is the best evaluated in our experiments.**

Keywords— person re-identification, video-surveillance, soft-biometric, deep learning.

I. INTRODUCTION

Re-Identification of persons (Re-ID) is increasingly in demand in the field of video surveillance. It is becoming an essential part of security needs, in various infrastructures such airports, shopping malls, government buildings and train stations, etc. [1]. Re-ID allows to know the identity of a person by means of multiple cameras whose images do not overlap [2].This identification process focuses on the extraction of spatio-temporal information from a sequence of images [3]. Re-ID systems in addition to using biometric features (especially the face) for the identification of persons, use other types of patterns that allow identification, called soft-biometric features, which, despite not having a high discriminatory power than biometric features, prove to be of great help for the identification of persons [4] [5]. These features lack of sufficient stability and discriminatory level to distinguish one person from another, especially when used separately [6]. Some of these characteristics are: body silhouette, skin color, gait, scars, tattoos, clothing (color and texture), etc. [7], which are considered suitable for complex situations where a person cannot be identified when his or her face cannot be distinguished [8].

Re-ID systems based on soft-biometric features are related to the use of Computer Vision techniques and methods that combined with Machine and/or Deep Learning models and/or algorithms have made it possible to improve their effectiveness and matching accuracy [9]. Due to the various issues such as camera resolutions, viewing angle, background changes, illumination variation, occlusion, and person pose changes have made such systems face numerous technical challenges [10]. The Re-ID turns out to be of great importance for tracking people in a Video-Surveillance system [11], which makes it a current and priority research field.

In general, Re-ID systems establish 3 stages for their development: (1) Person detection and segmentation, which is used to identify the region of interest in an image, which contains at least one person inside [11], (2) Feature extraction, which is in charge of extracting the information of interest from a person and (3) person identification, in charge of distinguishing and identifying the identity of a person.

Some studies address the issue of person detection and segmentation by using the Mask R-CNN framework, which is a conceptually simple, flexible, and a general framework for segmenting instances of objects or classes [12]. Another technique for segmentation is the use of a fully convolutional network (FCN) where pixel-by-pixel prediction is performed from supervised pre-training, thus achieving training efficiency and drastically improving the state of the art, while simplifying and accelerating segmentation learning and inference [13]. Regarding feature extraction, there are different Computer Vision techniques (descriptors), among the best known are LBP (Local Binary Pattern) and its variations U(LBP), H(LBP). LBP is one of the most widely used texture descriptors today [14], histograms, that allow extracting image features, e.g., the color of a person's clothing [15]. HOG (Histogram of Oriented Gradients) is a dense

(texture) descriptor computed in overlapping blocks along a grid of cells over regions of interest (ROI) [16]. For the person identification stage, Machine Learning models (classifiers) are used such as SVM (Support Vector Machine) which is a linear binary classifier that has been extended to nonlinear data using Kernels [17] and lately the frequent use of CNNs (Convolutional Neural Networks), which is a technology that combines Artificial Neural Networks with modern Deep Learning strategies [18].

Standard datasets from the surveillance community are used to train and validate this type of Re-ID systems, including PETS 2006 [19], PETS 2009 [20], SAIVT-SoftBio [21], CAVIAR4REID [22] and Market 1501 [23]. In addition, there is a new fully realistic database, Multi-camera Aeropuerto Internacional de Barajas (MUBA) [24]. Datasets containing images captured by high-definition cameras and processed by tracking objects in different weather conditions, light variation and period. They have been used to perform simulation tests and to adjust and validate the results of Re-ID systems, which turn out to be good references to determine the level of effectiveness achieved by these types of systems. Previous studies indicate the feasibility of such systems, for example, by using soft-biometric features such as body silhouette, clothing color or color information with CNN classifiers that have yielded accuracies above 40% [25] [26] [27].

However, the use of models with good performance and low computational cost are still under investigation. In this work we propose to implement a Re-ID system by analyzing two soft-biometric features such as body silhouette and clothing color. To achieve acceptable Re-ID accuracy, two models are used, first, a model for person detection and segmentation based on the Mask R-CNN framework, which allows to obtain masks and regions of interest (ROI) from the detections, secondly, a model based on a multi-layer multi-input network architecture. This model has as input two types of data, a feature map formed by the body silhouette and a color descriptor in RGB space, the color image with background subtraction. Each type of data is processed by two independent branches of the network, each of which is composed of a CNN-based structure capable of extracting the features of the images using their first hidden capabilities, and then converging on a branch combining the two features, to finally perform a classification of the person. For the training of the model and the experiments of the Re-ID system, a proprietary dataset was created in a controlled environment.

II. Materials and Methods

In the field of video surveillance, the re-identification of people plays an important role in tracking, controlling, and monitoring between cameras [28]. In this research work, the aim is to know the identity of a person, who are not occluded, through a sequence of images captured by different cameras, in different locations. To identify it, two patterns were used, the soft-biometric characteristics body silhouette and a complementary characteristic such as the color of the clothing. To predict it, Computer Vision and Machine Learning techniques and/or models were used.

A. System construction

The proposed system was developed with the Python programming language, using the Tensor Flow and OpenCV libraries. Fig. 1 shows the functional scheme of Re-ID, in which the implementation of three modules is proposed: (a) detection and segmentation of persons through the R-CNN Mask Network that allows generating a binary mask of the person [29], (b) soft-biometric feature extraction, in charge of analyzing and obtaining a descriptor of body silhouette and color of the person's clothing [30]. (c) Person identification, which allows classifying a person through a class, in a sequence of images, according to a degree of accuracy.

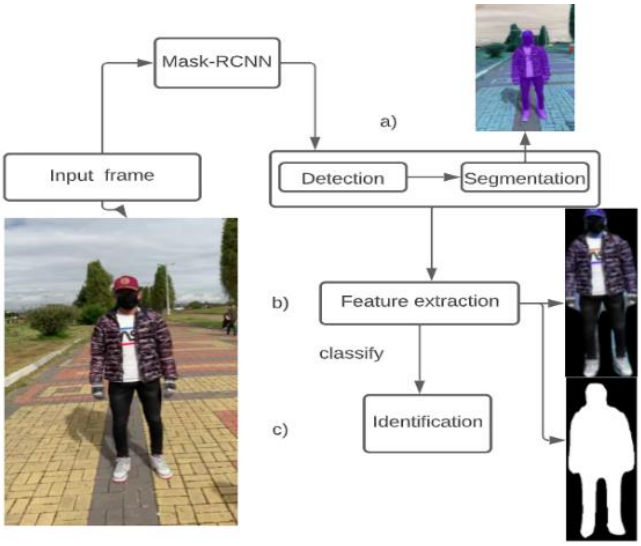

Fig. 1. Proposed process for the reidentification of persons

B. People detection and segmentation

This module is responsible for analyzing the images of a video sequence to locate objects within a scene (person detection) and to group homogeneous regions forming the object of interest (person segmentation). For detection, there are several factors that are considered, such as body appearance and/or environment background, which are usually complex [31]. In this work, it is assumed that the only thing moving is people. The Mask R-CNN model is used, imported from the Pixellib library that allows person detection, point of interest detection, and person segmentation [31] [32]. With this model, pixel-level positioning accuracy is achieved. The binary mask encodes the spatial distribution of the image using a fully convolutional neural network to mask each region of interest to the target [12]. Thus, pixel-level segmentation of the target, person-background, can be performed.

Fig. 2, shows flowchart of the Mask R-CNN network, where: (a) network input, in which an image of any size is input, (b) the general features of the image are extracted through the combination of the residual neural network (ResNet) and the feature pyramid network (FPN), which allows generate the feature map of different sizes of the image, (c) the regions of interest are predicted through the region proposal network (RPN), thereby creating the candidate window area of each image and the score of the detected persons at each location. (d) ROIs are aligned, to correct the pixels corresponding to the feature map of the input image to further improve the target detection accuracy. Once the candidate region is combined with the feature map, the system can achieve (e) target class masking (segmentation) using a fully connected convolutional network, which enables (f) person detection using the bounding window coordinates of the region of interest and (g) target classification [33] [32].

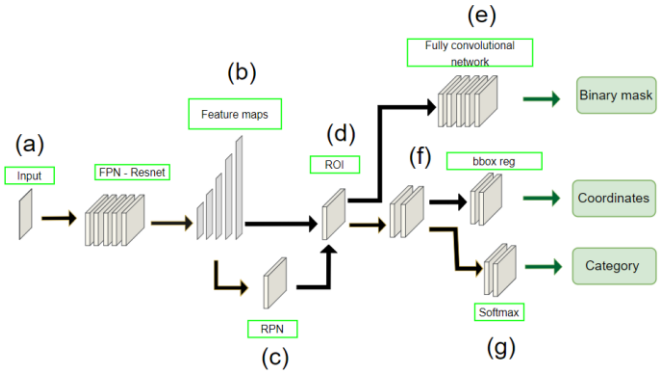

Fig. 2. Flowchart of the R-CNN mask for person segmentation

C. Clothing color

One of the most powerful and important features for object distinction and recognition is color. However, defining a robust descriptor with color can be a difficult task [34]. To simplify the analysis, a feature representation method is proposed which is based on learning, i.e., on RGB images of the person. The process to follow consists of 3 steps: (a) Apply the detection model of the target class (person) with the help of the PixelLib library, with the function segmentFrame, which returns a binary mask, (b) subtract the original image from this mask to remove the background and extract the appearance of the person. Equation (1) represents the region of interest without background, BI, where M is the binary mask of the image, and I is the color image (in RGB space). Finally, (c) the image is cropped according to the coordinates of the regions of interest, as shown in Fig. 3. This process returns an image of a person with the soft-biometric feature of color (complementary feature of clothing) as the main appearance cue [34]. Then, images are analyzed based on pixel patterns with the help of a convolutional neural network which automatically learns the optimal features for the Re-ID task from color data [2]. This method achieves better recognition performance than hand-made representation methods because the information can be better exploited [35].

$$BI = M * I \quad (1)$$

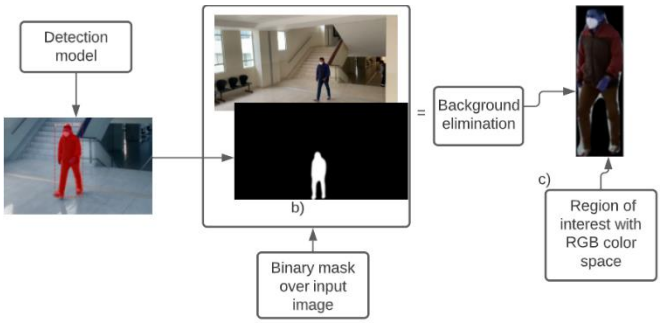

Fig. 3. Process for extracting a person in the RGB color space of an image.

D. Person silhouette

The silhouette of a person is a soft-biometric characteristic that identifies people by the shape of their body. It allows to capture and identify a person from a certain distance. For these reasons, the silhouette is a promising soft-biometric feature for scenarios where the face cannot be observed with sufficient resolution for recognition. To obtain this feature, the proposed process is much more direct after detection, since the Mask R-CNN segmentation method provides a binary mask. In this work this mask is resized to 40x40 pixels to reduce computational costs, within the window that delimits the silhouette, returning the silhouette image ready for a network or model to work on the image features. Fig. 4 shows an example of the silhouettes generated by this process.

Fig. 4. Example of five different silhouettes of a person

E. Re-Identification Model

This model uses deep learning based on one of the best-known architectures Convolutional Neural Networks (CNN), which are applied in mature identity recognition systems [25] because it allows color discrimination in the image and body silhouette domain [36]. For Re-ID using silhouette and color descriptors, a combination of CNNs was developed for both feature extraction and the use of a fusion network to generate a classification. A multi-layer multi-input multi-output perceptron-like network architecture is obtained, capable of efficiently exploiting the useful information of the descriptors, since these data types are processed independently.

The proposed architecture is shown in Fig. 5, which consists of two branches. The first branch receives as input the binary mask image of the silhouette, which is processed by a 16-filter convolution layer, a two-step max pooling (MaxPooling2D) 2x2. A second convolution layer of 32 filters, with MaxPooling2D 2x2 and a third convolution layer of 64 filters followed by a MaxPooling2D 2x2. The second branch receives as input the color image of the person, this branch is composed of only two convolution layers of 32 and 64 filters correspondingly, also with their respective 3x3 MaxPooling2D. Each branch also contains three fully connected network layers with a ReLU-like activation function [37], consisting of 128 neurons. The outputs of the three hidden layers are concatenated by a dense layer and a 256-neuron output layer, followed by a softmax activation function that uses the outputs of the feature extractors to perform a classification and display the person class number. As a result, the proposed architecture offers more flexibility for a low development cost which means that the computational effort of the training process is reduced.

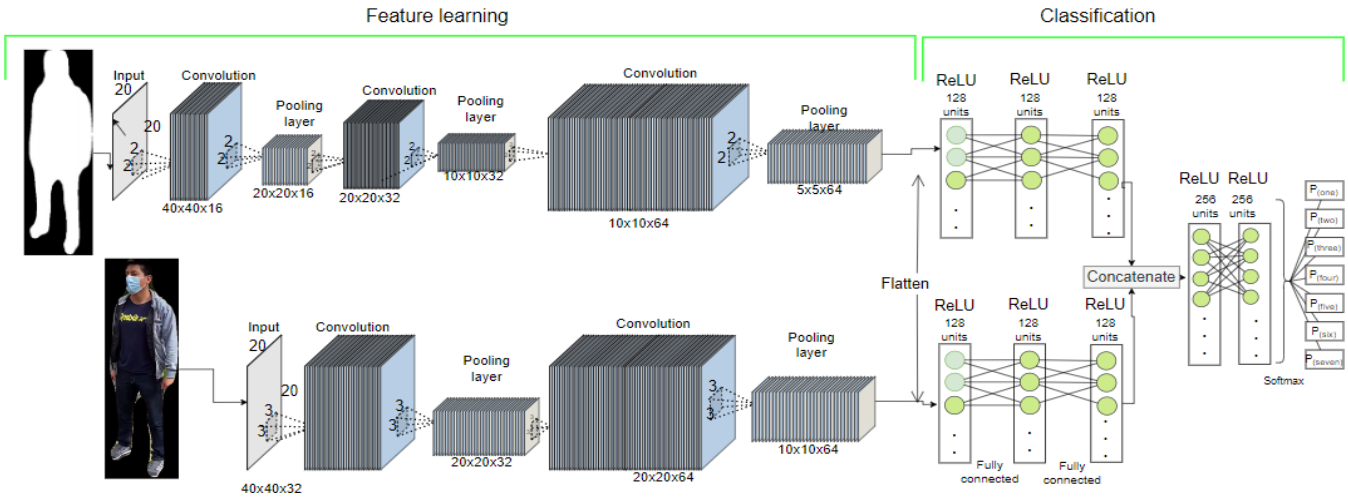

Fig. 5. Proposed architecture of a multi-input neural network and an output for classification

F. Dataset

For training and validation of the Re-ID system, a dataset was created capturing a sequence of images with different perspectives of the person, within a controlled environment, 7 different classes (persons) were taken. Initially, the dataset consists of 105 images (see Fig. 6). To avoid overfitting the models during training, improve the performance and results of their inference, the technique of data augmentation was used on these base images with the help of the ImageDataGenerator class class, which allows modifying an original image by zooming, scaling or rotation. In this case we rotated the sample images by 10 degrees and enlarged them within a 0.1 zoom range, they were also shifted by 10 % both horizontally and vertically, to further increase the size of our samples [38]. After this process we had a data set consisting of approximately 1862 training images and 481 test images.

Fig. 6. Sample datasets of 3 classes for training and testing

III. Results

In this section, the performance of the multi-input neural network is evaluated and compared with 2 models based on the branches that make up the proposed network, that is, a model for body silhouette and another model for clothing color. In addition, a comparison of the prediction effectiveness of the proposed model is performed using the in-house dataset with 12,295 images of 7 identities. The experiments are conducted on a MSI GL62M 7RDX computer running Windows 10, a 4 GB GDDR5 Nvidia GTX 1050 video card.

A. Training of the re-identification model

For compilating the model, the loss function "Categorical Crossentropy", an optimization function "Adam" and a learning degree of 0.0025 were defined. Additionally, to avoid the overfitting problem the network is trained and evaluated using the K-Fold cross-validation technique, where the training and validation data is used, and they are divided into *k* non-overlapping subsets (or folds) which are used to construct *k* training sets: the *i -th* training set is obtained as the difference between the full data set and the *i-th* subset. Each training set is overlapped with the others by (*k* - 2) folds [39]. In this case, the data were divided into 3 subsets, where each fold in turn plays the role of testing the model induced from the other *k* - 1 folds. From *k* models trained, the one that generated reliable accuracy in learning and evaluating results was kept.

Three models were trained for Re-ID. The first model was trained with only the binary mask descriptors, i.e., body silhouette. The second one used RGB color images without background of the person. And in the third one, the combination of the two models mentioned above (multi-input neural network). Table 1 describes the accuracy of each model after being trained with its own data set. Thus, the use of color in the image, showed a better result compared to the model where was applied only to the silhouette with 39.78% higher accuracy and 54 seconds longer training time, while the multi-input neural network (Silhouette + RGB) generates an accuracy of 80.21%, i.e. a higher accuracy level of 2.4% and a longer training time of 1 minute and 34 seconds with respect to the color model (RGB).

TABLE I. ACCURACY OF THE 3 RE-ID MODELS ACCORDING TO THEIR DESCRIPTORS

Model	**Dataset "own"**	
	Accuracy	***Training time***
Binary mask (Silhouette)	38.03%	1min 2s
Color image without background (RGB)	77.81%	1min 56s
Silhouette + RGB	80.21%	2min 9s

Fig. 7 shows the learning curves for the three models obtained with the in-house data set. The blue colored curve shows the result after training the model and the orange-colored curve represents the validation results. Considering the silhouette, a) represents the results of the trained model. The validation converges from 0.2 to 0.3 range. For the RGB color, b) shows the results of the trained model, the validation converges between 0.80 and 0.85. While the multi-input c) neural network training generates an accuracy validation that converges between 0.70 and 0.80.

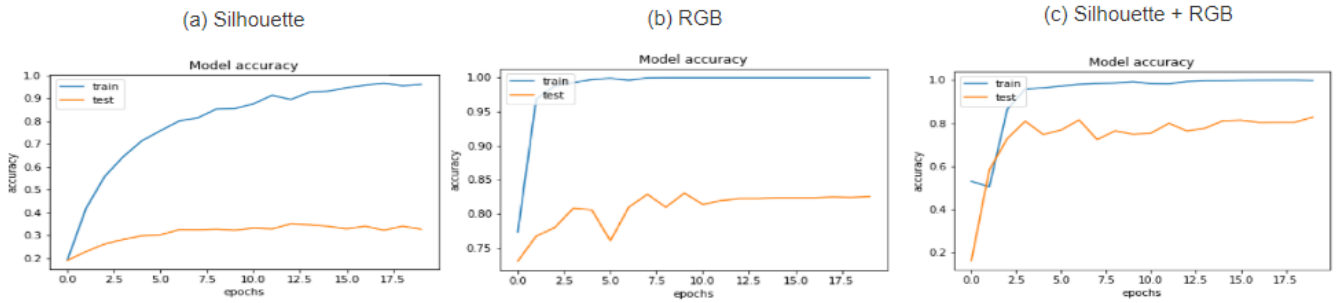

Fig. 7. Learning curves of 3 Re-ID models: (a) model trained with silhouette, (b) model trained with color and (c) multi-input neural network

B. *Performance evaluation*

The proposed multi-input neural network is evaluated using two test video sequences taken in a controlled environment of a corridor with natural lighting. The use of two cameras, opposed to each other, was adopted to acquire the test-evaluation videos. The field of view of the video is 185 degrees, the resolution is 3840 × 2160, the frame rate is 25 frames per second with a time duration of 21 seconds. The recall (2) obtained in it was the highest of the three, with 46.1596%, a value that demonstrates the model's ability to find correctly predicted true positive cases. The same is true for its F1 (3) score, a metric that provides a harmonic average of accuracy and completeness, summarizing these metrics in a single value. Therefore, it is shown that the fusion of the color and silhouette descriptors, increases the ability to recognize the person. Table 2 shows the comparison of the recognition completeness of our proposed network and the other CNN models of person recognition, focusing on only one descriptor color and silhouette correspondingly.

$$Recall = tp / (tp + fn) \quad (2)$$

$$F1 = 2 * (precision * recall) / (precision + recall) \quad (3)$$

TABLE II. EVALUATION VALUES OF THE 3 MODELS OF RE-ID

Model	Recall	F1
Binary mask (Silhouette)	30.72%	0.4699
Color image without background (RGB)	35.77%	0.5262
Silhouette + RGB	46.16%	0.6268

C. *Execution time*

Table 3 shows the detection, identification, processing and preprocessing time for each frame after running the models used for Re-ID. In this case study it can be noted that the combined model for the silhouette and the RGB color image with background subtraction takes a time for identification of 0.0110, a higher value than the other two models. This is due to the computational cost, i.e. the pre-processing of two descriptors before returning a final recognition result. In contrast, the model with the silhouette descriptor shows a time consumption for identification of 0.0092, a lower value than the other two models, this is because this descriptor is simpler and the number of steps for pre-processing is much lower.

TABLE III. TABLE TYPE STYLES

Model	Time (s)			
	Detection	*Identification*	*Processing*	*Preprocessing*
Binary mask (Silhouette)	1.8736	0.0092	1.8855	0.00150
Color image without background (RGB)	1.9237	0.0093	1.9359	0.00157
Silhouette + RGB	1.8701	0.0110	1.8839	0.00152

D. *Validation of the proposed model*

To validate the proposed model, a validation dataset was performed covering 57 images from each of the 7 classes. With the average precision (AP) as a metric, the 3 models described above were evaluated. Table 4 shows the performance of each model. The multi-input neural network showed the best percentage of accuracy with 51.64%, significantly surpassing the models that analyze only one descriptor (silhouette or color), with 31.32% more than the model that handles the color descriptor, while the model with the silhouette descriptor generates a lower level of accuracy, with 2.39% less than the model of the color descriptor.

TABLE IV. AVERAGE PRECISION OF THE RE-ID MODELS

Model	Average precision (AP)
Binary mask (Silhouette)	17.93%
Color image without background (RGB)	20.32%
Silhouette + RGB	51.64%

E. *Discussion*

The results obtained in the performance evaluation presented in Table 2 demonstrate the superiority of the proposed multi-input neural network as it outperforms the comparison models in the Re-ID accuracy percentage, showing the effectiveness in applying Computer Vision and Deep Learning techniques. The combination of soft-biometric features considerably benefits the performance and accuracy for Re-ID recognition. It was also evidenced that clothing color in RGB space alone is better than silhouette. However, the fact revealed in this work implies that the combination of two soft-biometric features, in this case color plus silhouette,

showed slightly better and reliable performance in accuracy for Re-ID.

There are two motivations for the development of this type of systems making use of descriptors such as color and silhouette with Computer Vision and Deep Learning techniques. Firstly, the use of these types of features, as well as the fusion of them, is more adequate for identifying a person in scenarios where the face cannot be observed with sufficient resolution for recognition. Secondly, considering that CNN architectures are mainstream in the field of Computer Vision it is more feasible to form a robust architecture superior to the traditional fully connected neural network architecture in terms of flexibility and adaptability [40]. It should be emphasized that the features used in this study do not constitute to a conclusive list, expanding it to more than 2 soft-biometric features or combining new features with those already proposed, could introduce other criteria and to influence the performance and accuracy of Re-ID.

IV. Conclusions

In this work a Re-ID system is proposed based on the analysis of information by combining soft-biometric features such as color in a RGB space and silhouette. Main tasks that must be performed in a system to achieve a correct identification of people such as class identification, segmentation, feature extraction and classification were addressed. In fact, a Re-ID model formed by a multilayer perceptron type network with double input and one output is shown, as well as its comparison with 2 other models that process only a soft biometric feature color or silhouette respectively. The combined method obtained the best results in all experiments, using a proprietary database captured in controlled environments. An accuracy of over 50% is recorded with an identification percentage of 51.64% in the average accuracy metric, which corroborates the accuracy of these two features indicated in [25] [26]. Therefore, it can be concluded that soft biometric feature fusion demonstrates its great potential. In the future, this work could be a starting point for other studies related to soft-biometric feature combination for Re-ID considering other types of color spaces such as HSV or HSL, also updating the proposed architecture for a real-time performance evaluation and validation.

Acknowledgments

This study is part of the research project called "Development of video surveillance models and/or algorithms for re-Identification of people based on Soft-Biometrics Features in a camera closed-circuit using computer vision and machine learning techniques". We would like to thank to all the research assistants of the project, also to the Universidad de las Fuerzas Armadas ESPE for supporting the development of this work.

References

[1] L. Nanni, M. Munaro, S. Ghidoni, E. Menegatti, and S. Brahnam, "Ensemble of different approaches for a reliable person re-identification system," *Applied Computing and Informatics*, vol. 12, no. 2, pp. 142–153, Jul. 2016, doi: 10.1016/J.ACI.2015.02.002.

[2] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 152–159, Sep. 2014, doi: 10.1109/CVPR.2014.27.

[3] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8692 LNCS, no. PART 4, pp. 688–703, 2014, doi: 10.1007/978-3-319-10593-2_45.

[4] R. Satta, "Appearance Descriptors for Person Re-identification: a Comprehensive Review," Jul. 2013, doi: 10.48550/arxiv.1307.5748.

[5] Y. Manabe and K. Sugawara, "Soft biometric verification by integrating static and dynamic features based on fuzzy inference," *Proceedings of the 12th IEEE International Conference on Cognitive Informatics and Cognitive Computing, ICCI*CC 2013*, pp. 242–247, 2013, doi: 10.1109/ICCI-CC.2013.6622250.

[6] A. K. Jain, S. C. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?".

[7] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 3, pp. 441–467, Mar. 2016, doi: 10.1109/TIFS.2015.2480381.

[8] S. Zhang, Y. Wang, T. Chai, A. Li, and A. K. Jain, "RealGait: Gait Recognition for Person Re-Identification".

[9] A. Yadav and D. Kumar Vishwakarma, "Person Re-Identification using Deep Learning Networks: A Systematic Review."

[10] G. Zou, G. Fu, X. Peng, Y. Liu, M. Gao, and Z. Liu, "Person re-identification based on metric learning: a survey," *Multimedia Tools and Applications 2021 80:17*, vol. 80, no. 17, pp. 26855–26888, May 2021, doi: 10.1007/S11042-021-10953-6.

[11] S. Salehian, P. Sebastian, and A. B. Sayuti, "Framework for Pedestrian Detection, Tracking and Re-identification in Video Surveillance System," *Proceedings of the 2019 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2019*, pp. 192–197, Sep. 2019, doi: 10.1109/ICSIPA45851.2019.8977728.

[12] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, Mar. 2017, doi: 10.48550/arxiv.1703.06870.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation".

[14] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996, doi: 10.1016/0031-3203(95)00067-4.

[15] T. Weng, Y. Yuan, L. Shen, and Y. Zhao, "Clothing image retrieval using color moment," *Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, ICCSNT 2013*, pp. 1016–1020, Nov. 2014, doi: 10.1109/ICCSNT.2013.6967276.

[16] A. Savakis, R. Sharma, and M. Kumar, "Efficient eye detection using HOG-PCA descriptor," *Imaging and Multimedia Analytics in a Web and Mobile World 2014*, vol. 9027, p. 90270J, Mar. 2014, doi: 10.1117/12.2036824.

[17] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: a review," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 803–855, Aug. 2019, doi: 10.1007/S10462-018-9614-6.

[18] S. Sakib, N. Ahmed, A. J. Kabir, and H. Ahmed, "An Overview of Convolutional Neural Network: Its Architecture and Applications," Feb. 2019, doi: 10.20944/PREPRINTS201811.0546.V4.

[19] Vasant Manohar, Matthew Boonstra, and Valentina Korzhova, "In Conjunction with IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) In Cooperation with IEEE Computer Society Supported by EU PASR ISCAPS Performance Evaluation of Tracking and Surveillance (PETS 2006)," 2006.

[20] J. Ferryman and A. Shahrokni, "PETS2009: Dataset and challenge," *Proceedings of the 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS-Winter 2009*, 2009, doi: 10.1109/PETS-WINTER.2009.5399556.

[21] A. Bialkowski, S. Denman, S. Sridharan, C. Fookes, and P. Lucey, "A database for person re-identification in multi-camera surveillance networks," *2012 International Conference on Digital Image Computing Techniques and Applications, DICTA 2012*, 2012, doi: 10.1109/DICTA.2012.6411689.

[22] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom Pictorial Structures for Re-identification", Accessed: Jul. 25, 2022. [Online]. Available: http://profs.sci.univr.it/~swan

[23] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable Person Re-identification: A Benchmark", Accessed: Jul. 25, 2022. [Online]. Available: http://www.liangzheng.com.cn.

[24] D. Moctezuma, C. Conde, I. M. de Diego, and E. Cabello, "Soft-biometrics evaluation for people re-identification in uncontrolled multi-camera environments," *Eurasip Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–20, Dec. 2015, doi: 10.1186/S13640-015-0078-1.

[25] C. Wang, Z. Li, and B. Sarpong, "Multimodal adaptive identity-recognition algorithm fused with gait perception," *Big Data Mining and Analytics*, vol. 4, no. 4, pp. 223–232, Dec. 2021, doi: 10.26599/BDMA.2021.9020006.

[26] F. Fang, K. Qian, B. Zhou, and X. Ma, "Real-time RGB-D based people detection and tracking system for mobile robots," *2017 IEEE International Conference on Mechatronics and Automation, ICMA 2017*, pp. 1937–1941, Aug. 2017, doi: 10.1109/ICMA.2017.8016114.

[27] M. Flores-Calero *et al.*, "Ecuadorian traffic sign detection through color information and a convolutional neural network," *2020 IEEE ANDESCON, ANDESCON 2020*, Oct. 2020, doi: 10.1109/ANDESCON50619.2020.9272089.

[28] M. Jiang, Z. Li, and J. Chen, "Person Re-Identification Using Color Features and CNN Features," *2019 IEEE 4th International Conference on Image, Vision and Computing, ICIVC 2019*, pp. 460–462, Jul. 2019, doi: 10.1109/ICIVC47709.2019.8980977.

[29] M. A. Malbog, "MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation," *ICETAS 2019 - 2019 6th IEEE International Conference on Engineering, Technologies and Applied Sciences*, Dec. 2019, doi: 10.1109/ICETAS48360.2019.9117217.

[30] M.-O. D. Alejandra, "Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video-vigilancia," *Ingeniería, Investigación y Tecnología*, vol. 17, no. 2, pp. 257–271, Apr. 2016, doi: 10.1016/J.RIIT.2016.06.010.

[31] Y. Wang, "Human Detection Based on Improved Mask R-CNN," pp. 1–9, 2020, doi: 10.1088/1742-6596/1575/1/012067.

[32] C. Xu *et al.*, "Fast Vehicle and Pedestrian Detection Using Improved Mask R-CNN," *Mathematical Problems in Engineering*, vol. 2020, 2020, doi: 10.1155/2020/5761414.

[33] M. Chen, F. Bai, and Z. Gerile, "Special Object Detection Based on Mask Rcnn," *Proceedings - 2021 17th International Conference on Computational Intelligence and Security, CIS 2021*, pp. 128–132, 2021, doi: 10.1109/CIS54983.2021.00035.

[34] A. D'Angelo and J.-L. Dugelay, "People re-identification in camera networks based on probabilistic color histograms," *Visual Information Processing and Communication II*, vol. 7882, p. 78820K, Jan. 2011, doi: 10.1117/12.876453.

[35] L. Ren, J. Lu, J. Feng, and J. Zhou, "Uniform and variational deep learning for RGB-D object recognition and person re-identification," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 4970–4983, Oct. 2019, doi: 10.1109/TIP.2019.2915655.

[36] K. Zhou, A. Paiement, and M. Mirmehdi, "Detecting humans in RGB-D data with CNNs," *Proceedings of the 15th IAPR International Conference on Machine Vision Applications, MVA 2017*, pp. 306–309, Jul. 2017, doi: 10.23919/MVA.2017.7986862.

[37] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," Mar. 2018, Accessed: Jun. 30, 2022. [Online]. Available: http://arxiv.org/abs/1803.08375

[38] M. M. I. Rahi, F. T. Khan, M. T. Mahtab, A. K. M. Amanat Ullah, M. G. R. Alam, and M. A. Alam, "Detection of Skin Cancer Using Deep Neural Networks," *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2019*, Dec. 2019, doi: 10.1109/CSDE48274.2019.9162400.

[39] D. Anguita, S. Ridella, and F. Rivieccio, "K-fold generalization capability assessment for support vector classifiers," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, pp. 855–858, 2005, doi: 10.1109/IJCNN.2005.1555964.

[40] I. E. Livieris, S. D. Dafnis, G. K. Papadopoulos, and D. P. Kalivas, "A multiple-input neural network model for predicting cotton production quantity: A case study," *Algorithms*, vol. 13, no. 11, pp. 1–14, Nov. 2020, doi: 10.3390/a13110273.