Person Re-identification using soft-biometric features: body silhouette and clothing texture in a multi-camera video surveillance environment

Chango Caisabanda, David Bernardo

Cuyo Chiluisa, Víctor Hugo

Departamento de Ciencias de la Computación

Carrera de Software

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

Dr. Carrillo Medina, José Luis

17 de agosto de 2022

Latacunga

Person Re-identification using soft-biometric features: body silhouette and clothing texture in a multi-camera video surveillance environment

David Chango-Caisabanda
Department of Computer Sciences
Universidad de las Fuerzas Armadas - ESPE
Sangolquí, Ecuador
dbchango@espe.edu.ec

Victor Cuyo-Chiluisa
Department of Computer Sciences
Universidad de las Fuerzas Armadas - ESPE
Sangolquí, Ecuador
vhcuyo@espe.edu.ec

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

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

***Abstract*—People Re-Identification has become a topic of interest due to the increasing use of intelligent video surveillance systems in the security industry. In this paper, we implement a people Re-Id system comprising three important modules: a) a person detection, responsible for detecting people in the image, b) preprocessing module, responsible of extracting the soft-biometric features of the detected persons, and c) an identification module, capable of identifying the detected person. For this purpose, a two branches multi-input and one output network model is built. The first one receives the body silhouette descriptor and the other the clothing texture descriptor. To train this model a dataset of 7 identities was built, with 1862 and 481 images for training and validation respectively, facing problems such as the existing bias in the public datasets. In addition, two videos and one validation image set were used to evaluate the system performance. The results of our proposal are positive, demonstrating that the combination of soft-biometrics features, body silhouette and clothing textures of the person increases the system ability to Re-Identify a person in images an qd videos.**

***Index Terms*—people re-identification, soft-biometric features, video surveillance, deep learning, multi-input model**

I. Introduction

Nowadays, Re-Identification (Re-ID) of people is attracting a lot of attention due to its high demand and the remarkable increase of camera networks installed in public places such as airports, universities campuses and buildings, whose purpose is the security of people [1]. The main task of these systems is the retrieval of instances between cameras that aim to search for people in multiple non-overlapping cameras in a controlled environment and visual ranges [2], [3], so the problem fits in the area of image processing and retrieval. Researches with this type of technologies can provide a large number of useful applications and/or tools for security, specifically in intelligent video surveillance systems that can be key in the fight against crime and terrorism [1].

In general, for Re-ID two types of features can be considered, biometric (hard) and soft-biometric [4]. The first ones are the most common, such as: face, fingerprints, DNA sequence, retinal characteristics, etc. On the other hand, in recent years, another type of characteristics called soft-biometrics, have been defined, which, although they do not have a high discriminatory power as hard biometrics, are of great help in the identification of individuals [5]. Some of these characteristics that are used to identify people are for example: silhouette, height and width, skin color, hair color, gait, distinctive scars, tattoos, etc. In addition, there are complementary characteristics such as the color and/or texture of the people's clothing. According [1], these types of features alone have a low performance, but combined, they are more robust in people Re-ID.

Re-ID is related to multiple scientific fields such as Computer Vision, Pattern Recognition and Machine Learning, which provide different approaches to its implementation [6]. Generally, the implementation of Re-ID systems consists of 3 main stages: 1) person detection, in which the region of interest of the detected person is generated, 2) feature extraction, responsible for extracting relevant data from the image, and 3) person identification, responsible for determining the person identity.

Several studies in the literature address person detection and segmentation using Mask R-CNN [7], [8], which is very popular in this context. Several techniques have been proposed in the literature, such as fully convolutional networks [9] (FCN), which improves semantic segmentation performance of arbitrary sized images through combining semantic information from a deep layer with appearance information from a shallow layer, producing more accurate and detailed segmentation [10], this technique is frequently implemented in CNN models [11].

On the other hand, preprocessing is necessary, because it prevents the model from taking useless data for the system. As far as Computer Vision is concerned, there are several descriptors that focus on texture recognition and this is the case of the local binary pattern (LBP) [12], which describes the orientations of the edges in the images in an acceptable way, but loses the intensity information, however, it is independent of illumination variations. On the other hand, there is the Weber local descriptor (WLD) [13], which is able to preserve intensity information and also detects edges gracefully, but in turn loses their orientations [14]. The histogram of oriented gradients (HOG) [15], extracts information from the distribution of edges in local regions as a means to represent their shape and concentrates on defining an image as a group of local histograms [16]. If we talk about the human silhouette, a novel approach for recognition is presented using Fourier descriptors [17], however, the most commonly used descriptor is the binary mask.

According to the literature, it can be point out that the percentage of accuracy for Re-ID, for example using the clothing texture, the percentage of accuracy is 48.12% [18], using classifiers such as K-means [19] and SVM [20]. With respect to body silhouette an accuracy of 77.40% has been obtained [21], having as descriptors HOG and SVM classifier [22]. Accordingly with these results, it is evident that soft-biometric features, such as body silhouette and clothing texture, are high-performance alternatives for the Re-ID of individuals. However, the lack of public data for Re-ID is a factor that can negatively influence the training of these systems because each of these databases have very low-quality images or few variations of images per identity, therefore, the systems cannot correctly generalize possible data that are not in the database. This factor has encouraged the creation of proprietary databases and/or the use of techniques to augment their data and variations, also known as data augmentation.

This paper proposes the development of a Re-ID system for people that are positioned in a closed circuit of non-overlapping cameras in a controlled environment. The system has the particularity of identifying a person at a given point and then seeks to identify him or her along its path by other cameras (as long as he/she does not change clothes) without the need to distinguish his face. To identify them, soft-biometric characteristics, such as body silhouette and texture of the person's clothing are used. A database of finite identities is created, with different variations in order to build an informative pattern that allows the identification of different people in different places. For this purpose, Computer Vision techniques and Machine Learning models are used to increase the accuracy of the Re-ID.

II. Methods and materials

Nowadays, Re-ID systems have become an area of high interest for the Computer Vision (CV) community [20] due to their applications in the security field and the increasing demand by companies and public and/or private institutions in general [1]. Therefore, a non-intrusive person re-identification system was proposed, which allows the automatic analysis of the scene for the detection and identification of people through different cameras.

A. System construction

The proposed system, which is the basis for tracking people, was built using the Python language due to its simplicity, consistency, flexibility and platform independence. The XP (Extreme Programming) software methodology [23] was applied for its development. The functional scheme of the system is shown in Fig. 1, in which three main modules are identified: a) Detection and segmentation of people, which allows determining the regions of interest (of people) in an image or a sequence of images, b) Preprocessing, responsible for removing the noise (background) of the images and the extraction of the soft-biometric characteristics of a person, and **c)** Identification, which allows determining the identity of the detected person.

B. People detection and segmentation

The first step for the Re-ID of people is to detect the regions of interest over an image sequence [24]. Therefore, a module capable of performing this task automatically was developed using as a basis the pre-trained model of the Mask R-CNN object detector [7], which was imported from the Pixellib library [25] to simplify the segmentation of objects in images and videos [26]. The Mask R-CNN model consists of four phases (see Fig 2), a) the backbone network, based on Faster

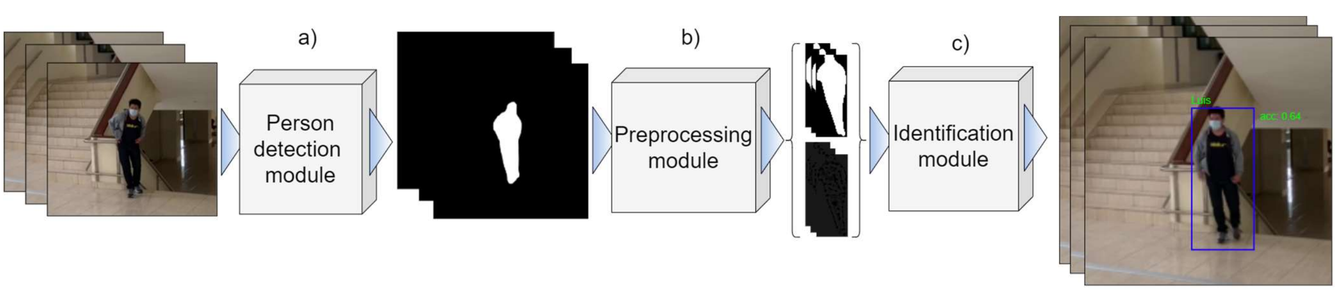

Fig. 1: The functional scheme of the re-identification system consisting of three modules: a) person detection and segmentation module, b) pre-processing module, and c) identification module.

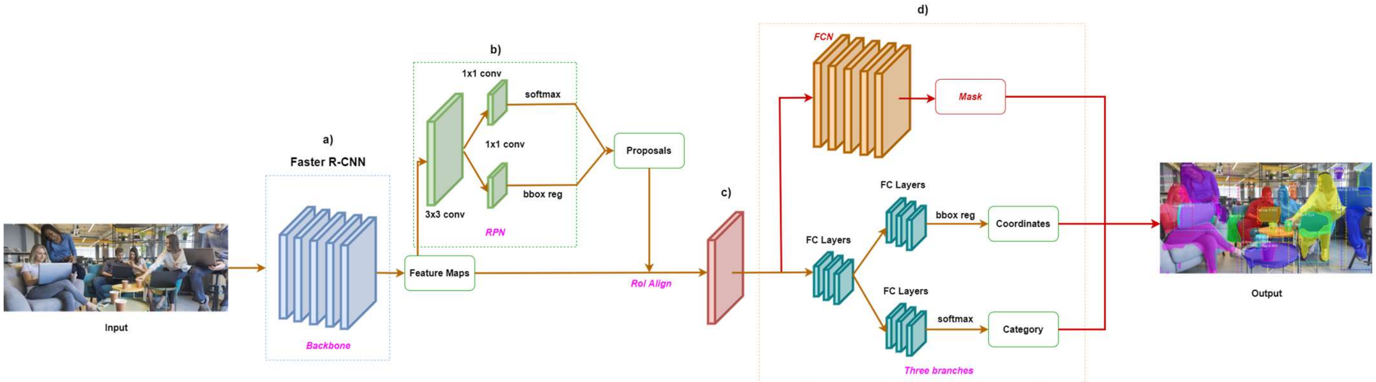

Fig. 2: Mask R-CNN framework for people detection and segmentation. It consists of: a) backbone network based on Faster R-CNN, b) Region Proposal Network, c) mapping of generated regions of interest, and d) binary mask generation, bounding boxes and classification accuracy.

R-CNN that is responsible for extracting feature maps for the generation of regions of interest (ROIs), b) a Region Proposal Network (RPN), the ROIs from c) that are used to extract corresponding features of interest with the shared feature maps using d), a stack of Fully Connected Layers (FC) and a Fully Convolutional Network (FCN) for instance segmentation and object classification, where they are processed to generate the classification accuracies, bounding boxes and binary masks (see Fig. 2).

C. Pre-processing

The preprocessing module is used to remove the background and extract the soft-biometric features from the images obtained by the previous module and to convert them to a system-readable format. The scikit-learn [27] and OpenCV [28] libraries, two of the most widely used libraries in image analysis and processing, were used to develop this module.

1) Background deletion: Detection of people with different cameras may differ due to noise in the environment, affecting image feature extraction [29]. Background removal is widely used in scenarios where foreground objects need to be detected in image sequences [30]. This concept can be applied for background removal to discard irrelevant information [31]. The background subtraction method is used as shown in Fig. 3, where the images to be analyzed by the system in a) these are read through the person detection module and this detects instance binary masks and the coordinates of the Regions of Interest (ROIs) b) are generated, subsequently, these data pass to the background subtraction module where we used the equation (1) (in which M is the binary mask, F us the original image in where the binary mask is applied and R is the resulting image), and finally the images are cropped to extract the region of interest.

With the outputs of the people detection module a), the binary mask and the coordinates of the regions of interest of the detections generated b), equation (1) is applied, where M is the binary mask, F is the original image on which the mask is applied and R is the resulting image with the background removed, and finally c) the image is cropped to extract the region of interest.

$$R = M \cdot F \tag{1}$$

2) Extraction of soft-biometric features: In many occasions it is complicated to identify a person by means of the face in a video surveillance system [1]. In this work we propose to use the person as a single object [32] (holistic representation) and the extraction of two soft-biometric features for its Re-ID, such as: the silhouette of the person and the texture of his clothes.

a) Clothing textures: Different methods have been proposed to describe the textures of a person's clothing in an image, e.g., Fourier spectrum, Gabor, Wavelet and Local Binary Pattern (LBP) filters. In this work, the uniform LBP filter was implemented to generate the texture descriptors because of its computational simplicity, tolerance to illumination changes and its discriminative power. The LBP filter labels each pixel of the image through an analysis of its neighborhood, studies whether the gray level of each pixel exceeds in a certain threshold and encodes this comparison using a binary number, forming a byte of information. To determine whether a pattern is uniform there must be at most two jumps between 0 and 1 or between 1 and 0 as indicated in equation (2). Where P is the number of neighbors to be considered, R is the neighborhood size, and g_c and g_p are the gray values of the central pixel and each the pixels in the neighborhood, respectively, and U(LBP) is the uniform LBP, which is defined in equation (3). By using this variation of LBPU(LBP), the complexity of the descriptor is reduced and invariance to rotations is obtained [33]. To implement this method, the "local_binary_pattern" module of the "scikit-image" library was used, with a radius of 3 and a default number of 24 points. Finally, a filtered image as shown in Fig. 4, is generated.

$$LBP^{riu2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1, & \text{otherwise} \end{cases} \tag{2}$$

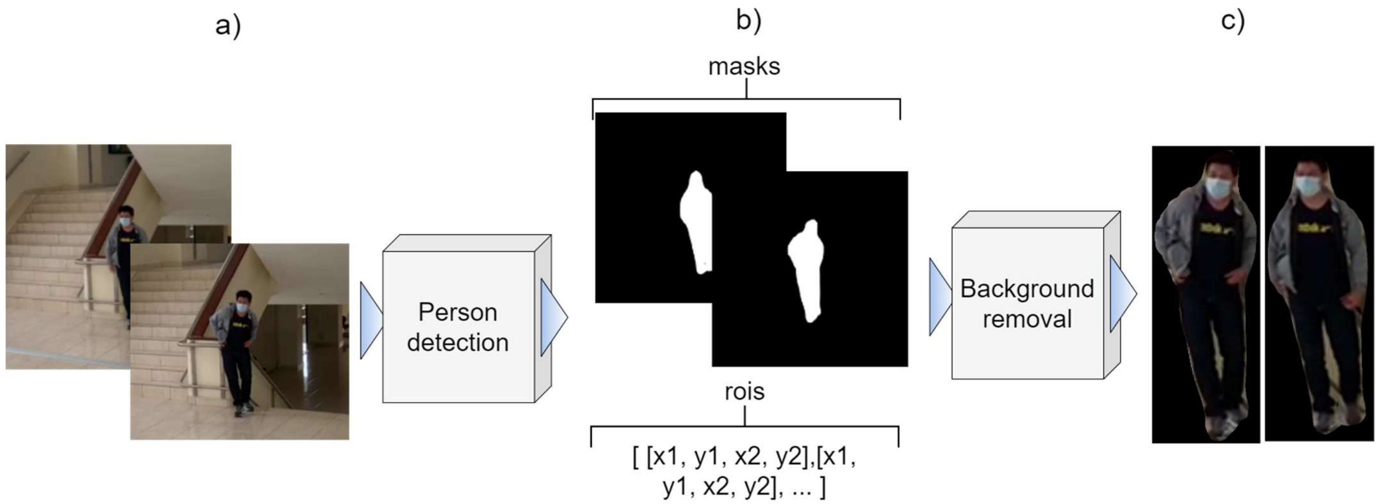

Fig. 3: Background subtraction method, which consists of: a) image reading, b) generation of binary masks and bounding boxes coordinates, and c) image cropping and background removal.

$$U(LBP) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} s(g_p - g_c) - s(g_{P-1} - g_c) \quad (3)$$

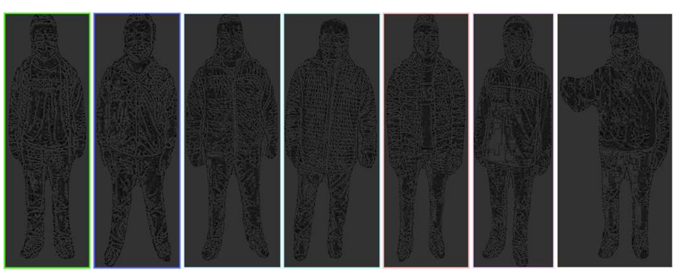

Fig. 4: Filtered images using LBPU.

b) Person silhouette: Person silhouette is another feature that can be used to identify he or her, when it is not possible to see their face, improving the robustness of the system to variations in scale, poses, lighting and clothing changes [34]. We propose the use of a binary mask as a descriptor of the silhouette of a person for its identification. This mask is one of the outputs of the Mask R-CNN model, which is trimmed according to the coordinates of the person's ROI, obtaining binary masks as shown in Fig. 5.

Fig. 5: Binary masks of detected people.

D. Re-identification model

Deep learning models for person re-identification have proven to be more robust than other methods, allowing deep feature extraction [6], so the use of large amounts of information for model training has become an approach in the field of person Re-ID [35], [36]. One of the most widely used approches in deep learning based person Re-ID is the use of classification models, also known as identification models, in which Re-ID is considered as a multi-class classification problem as presented in [36]. Deep learning models with multiple inputs for image classification have enabled the diversification of images, facilitating the use of different data structures, e.g., images, matrices, vectors, scalars, and text, in order to achieve more robust models [36]. This work Re-ID is approached as a classification problem, in which we train the model using seven identities. The proposed Re-ID model in this work is a multi-input classifier, for its design we considered the strengths and patterns in which layers are stacked in neural network architectures as in the literature such as AlexNet and LeNet. To implement this module we use the Keras tool [37], a high-level API written in Python, used for the simple implementation of neural networks. The model consists of two branches, each accepting different types of input. In the first branch we receive an image of the clothing texture of the person of interest filtered with the LBPU algorithm, while the second branch receives the binary mask of the person's silhouette. The designed model has 3,345,543 trainable parameters, its architecture consists of two similar branches with a 40x40 input, a convolution layer with 32 3x3 filters, a MaxPooling layer with a 2x2 filter and a 2x2 stride. These two branches are combined and passed through two fully connected layers of 256 neurons and finally through an output layer of 7 neurons (one for each identity) with a "softmax" activation function, often used for image classification [36]. The outputs of the model correspond to a vector of 7 elements with the probability of correspondence of the inputs with each class, where the index of the highest value corresponds to the inferred class (Fig. 6).

E. Training and validation dataset

Public datasets used in the literature for re-identification of individuals using CNNs are limited due to: small number of images per identity, images captured with few variations, small number of cameras to capture the samples, etc. (data bias), causing overfitting in models trained with this data [38], [39], this problem is common when using deep learning models [40]. For this reason, a proprietary set of images (dataset) was collected in this work for testing and validation of the Re-Id system. This contains 7 different identities. Each one has at least 105 images divided into two groups: training (with 85 images) and test (with about 20 images).

These samples were captured in two scenarios with natural illumination. To solve the problem of low variation of the public datasets, different captures were taken in different perspectives of the participants, as can be seen in Fig. 7. In addition, data augmentation technique was used to improve the robustness and accuracy of the approach. We generated 10 variations of each capture, using the "ImageDataGenerator" module of Tensorflow [41]. The data generator configuration parameters were a rotation range of 10 degrees, a horizontal and vertical displacement range of 0.1, a zoom range of 0.1 and with horizontal rotation. Finally we have a total of 1862 training images and about 481 test images.

III. Experimental results

In this section we report the experiments and the corresponding results of the proposed approach. A proprietary database with 12,295 images of 7 identities was used to evaluate its effectiveness. An MSI GL62M 7RDX computer running on Windows 10, an Nvidia GTX 1050 video card with 4 GB GDDR5, 16 GB of RAM and an Intel core i7-7700HQ processor were used.

A. Training of the re-identification model

For the training of the Re-ID model we used the validation and training data of the data set described in section II. A loss function "Categorical Crossentropy", an optimization function "RMSProp" and a learning degree of 0.001 were used. To avoid overfitting the model, it was decided to train the model using the "k-fold" cross-validation technique, in which the data are partitioned into k equal or similar segments on which training and testing is done in k iterations, so that in each iteration we leave one segment for testing and the rest of k-1 segments for model training [42]

A value of k equal to 3 was taken. Therefore, three models were trained for the Re-Id, from which those with the best learning curve and the best accuracy results during the evaluation were selected. The first one was trained with the descriptor of clothing textures, the second one with the descriptor of people's body silhouette, and the third one with the combination of these two descriptors. The results of the three models are shown in Table 1, where the combination of the LBPU and Binary Mask descriptors (model 3) has better results than the other resulting models (models 1 and 2) during training. In model 3, an accuracy of 71.58% is recorded, having an increase of 2.94% and 28.59% with respect to models 1 and 2, respectively. However, the training time is the highest, by more than one minute. Reflecting an increase of 81 and 89.2 seconds over models 1 and 2, respectively.

TABLE I: Model results during training

Model	Descriptors	Accuracy	Training time
1	*LBPU*	*68.54%*	*59sec*
2	*Binary mask*	*42.99%*	*50.8sec*
3	*Binary mask + LBPU*	***71.58%***	***2 min 20sec***

B. System performance

To evaluate the performance of the proposed system, two videos captured with two disjunctives 12 and 13 megapixel cameras placed in a corridor with natural lighting were used, trying to cover the maximum possible field of view (see 8). The two cameras were positioned opposite each other, so that camera 1 points towards the stairs of the corridor (starting point of the trajectory of the filmed persons) and camera 2 points to the opposite side. The time duration of the videos was 21 seconds, read with a latency of 25 frames per second and a resolution of 3840x2160 pixels

System processing times and accuracy were measured to identify the correct classes in the proposed experiments. Accuracy indicates the fraction of correct predictions of the model, this metric can be measured using equation (4), which relates the number of true Positives plus true Negatives to the total number of inferences made. Indicators in Table 2 measure the processing times, the first one measures the time the system takes to process a video, the second one the time in which each frame is preprocessed, and the third one the time needed to identify the detected person. Additionally, the accuracies obtained during this test were compared, the results indicate that: 1) model 3 is the slowest to process a video, with a duration of 166.722 seconds and an increase of 34.356 and 43.368 seconds with respect to models 1 and 2, 2) model 1 is slower during preprocessing, with a time duration of 0.0079 and an increase of 0.0016 and 0. 0021 seconds with respect to models 1 and 2, 3) models 3 and 1 are the slowest during person identification with a duration of 0.0101 and 0.0107 respectively and an increase of about 0.0005 seconds with respect to model 2, and 4) model 3 has the best accuracy

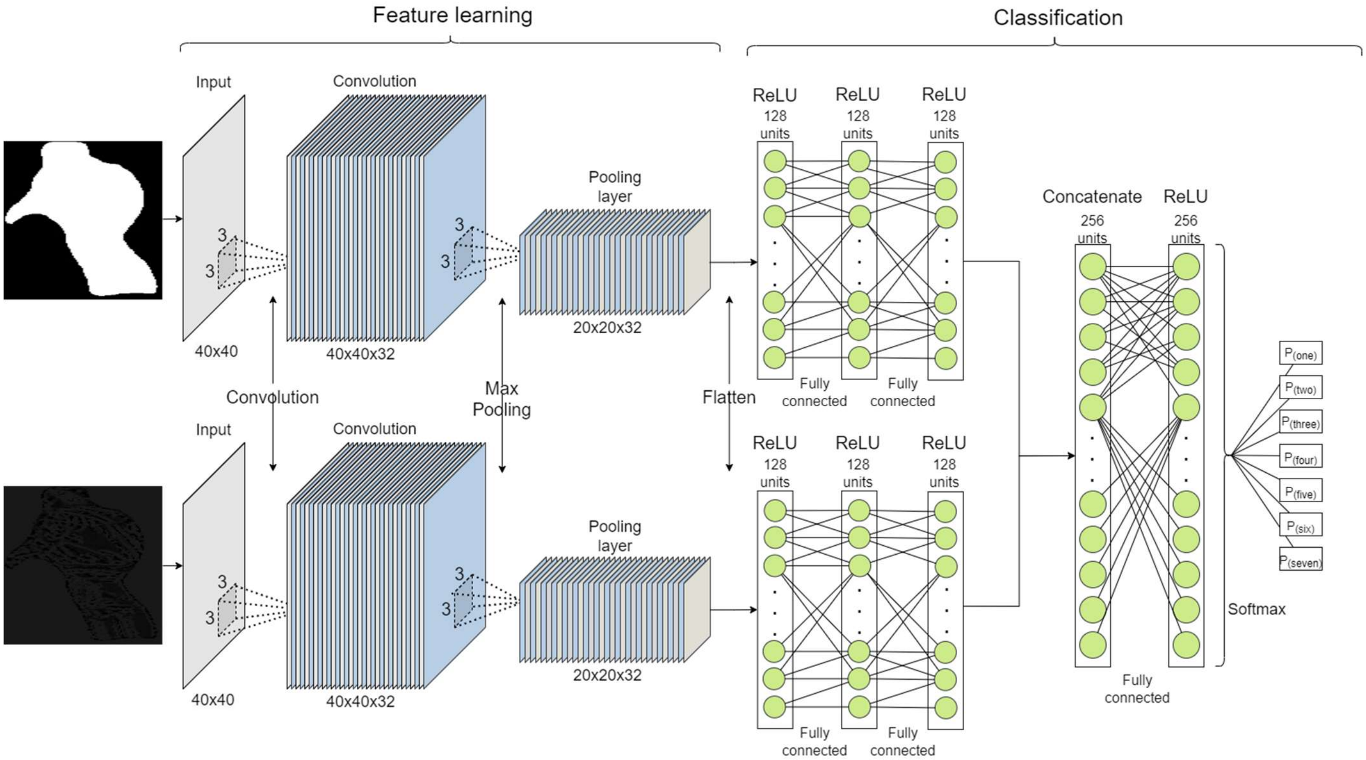

Fig. 6: Design of the people re-identification model architecture.

Fig. 7: Dataset samples.

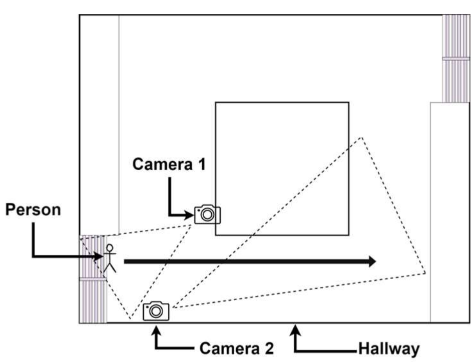

Fig. 8: Location of cameras to capture test videos.

results, yielding a rating of 38.57% outperforming model 1 and 2 with 5.44% and 8.26% respectively. This denotes that the integration of the U(LBP) filter in the system flow, increases the computational cost to process a complete video and to identify a person, this is evidently due to the complexity of this descriptor. Furthermore, it can be determined that the combination of the U(LBP) and binary mask descriptors improves the Re-ID results of this proposal

TABLE II: System performance results

Model	Time			Accuracy
	Total	Pre-processing	Identification	
1	*2.2061min*	*0.0079sec*	*0.0107sec*	*33.13%*
2	*2.0559min*	*0.0063sec*	*0.0095sec*	*30.31%*
3	*2.7787min*	*0.0058sec*	*0.0101sec*	*38.57%*

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

C. System validation

To validate our approach, we prepared a set of 57 images (from all the 7 identities) that were not included during training. We measured: 1) precision (5), 2) recall (6), 3) F1-score (7) and 4) accuracy (4) of the models. These metrics allows us to:

- **Precision:** measure the proportion of positively predicted cases that are correct.
- **Recall:** measures the proportion of true positive cases that have been correctly predicted.
- **F1-score:** is the harmonic mean of precision and completeness, allowing the ratio between these two variables.

Finally, accuracy allows us to measure the percentage of model hits [43]. Table III shows that model 3, in general, has the best prediction results with respect to models 1 and 2, with an accuracy of 45.61% and an F1-score of 0.4555, reflecting an accuracy improvement of 5.26% 14.03%, and an F1-score improvement of 0.0314 and 0.1810 respect to models 1 and 2, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{7}$$

The results obtained in this research work demonstrate that

TABLE III: Measurements for the evaluation of Re-ID models.

Model	Precision	Recall	F1	Accuracy
1	*0.5095*	*0.4035*	*0.4241*	*40.35%*
2	*0.2895*	*0.3158*	*0.2745*	*31.58%*
3	***0.6245***	***0.4561***	***0.4555***	***45.61%***

the use of soft-biometric features positively influences the Re-ID of people in situations where other types of features cannot be extracted, either due to low image quality, illumination changes, occlusions, and others. The use of deep learning techniques and soft-biometric features allowed to improve the performance of the Re-ID system because these models can learn from deep features. By extracting these features, we focus their attention on the important information, and the proposed combination of soft-biometric features (clothing texture and body silhouette) allowed a more robust characterization of a person. The obtained measurements, shows that combination of features, improves the Re-ID system capacity, with an increase of 0.0314 from the use of clothing textures and 0.1810 from the use of silhouette. However, this can be further improved by using more descriptors.

The results obtained from the evaluation of our system are close to obtained in [44], in which they used Color Histograms (CH), Gabor Filter Banks (GBF), Color Structure Descriptors (CSD), and Local Binary Pattern (LBP). CH+GBF, CSD+LBP and CH+GBF+CSD+LBP obtained 0.28, 0.37 and 0.40 precision, respectively. Results are also close to the results obtained in [45], where they use CH and Multi Block Local Binary Pattern (MB-LBP) with a difference of ± 0.03. Accordingly, it can be indicated that our model could perform better than those proposed in [44] and [45].

IV. Conclusions

This paper proposes a deep learning-based system for the Re-ID of people using two descriptors, the first one the texture of their clothing and the second one their body silhouette(soft-biometric features). This system addresses all the tasks necessary to perform Re-ID, which includes person detection, image and video pre-processing, and person identification. From a multi-input classifier, the attributes of a proprietary dataset have been trained. Our proposal can be used in a single and multi-camera video surveillance environment.

The results obtained by the model that combines the proposed soft-biometric characteristics reflect an accuracy in the Re-ID of people of 71.58%, which is much better than the results obtained by the other models. For the validation stage, two types of tests were proposed and used to determine the performance of the system. In the first test, two videos were used, taken with two cameras placed across a corridor, showing a single person. For the second, a set of 57 images were prepared by mixing captures of the 7 validation data identities from our proposed image set.

The results of the first test on the system, shows that the combination of body silhouette and clothing texture descriptors give better results than the individual use of these descriptors. This resulted in an accuracy of 38.57% with respect to the second proposed test where the system achieved an accuracy of 45.61% and an F1 score of 0.4555. These results demonstrate the robustness and efficiency of the system when using the combination of features proposed in the Re-ID of persons.

According to these results, a great potential of using soft-biometric features in combination with deep learning model and/or algorithms, which have gained relevance in this field, is demonstrated. Therefore, this work can be a solid basis for developing robust re-identification systems that can be used by public and private entities to improve security in their facilities.

Acknowledgement

The present work is part of the research project ”Development of video surveillance models and/or algorithms for the re-identification of people based on soft biometric features in a closed circuit camera using computer vision and machine learning techniques”. In addition, we would like to thank the research team of the project, as well as the Universidad de las Fuerzas Armadas ESPE, for the support to achieve the culmination of this work.

References

[1] M.-O. D. Alejandra, “Re-identificación de personas a través de sus características soft-biométricas en un entorno multi-cámara de video-vigilancia,” Ing. Investig. y Tecnol., vol. 17, no. 2, pp. 257–271, April 2016, doi:

[2] X. Wang, M. Liu, X. Hu, D. S. Raychaudhuri, S. Paul, Y. Wang, and A. K. Roy-Chowdhury, “Learning Person Re-Identification Models From Videos With Weak Supervision,” IEEE Trans. Image Process., vol. 30, pp. 3017-3028, 2021. doi: 10.1109/TIP.2021.3056223.

[3] M. Cao, C. Chen, X. Hu, and S. Peng, “From Groups to Co-Traveler Sets: Pair Matching Based Person Re-identification Framework,” in 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 2573-2582. doi: 10.1109/ICCVW.2017.302.

[4] C. Patruno, R. Marani, G. Cicirelli, E. Stella, and T. D’Orazio, “People re-identification using skeleton standard posture and color descriptors from RGB-D data,” Pattern Recognit., vol. 89, pp. 77-90, 2019. doi: https://doi.org/10.1016/j.patcog.2019.01.003.

[5] A. Dantcheva, C. Velardo, A. D’Angelo, and J.-L. Dugelay, “Bag of soft biometrics for person identification,” Multimed. Tools Appl., vol. 51, no. 2, pp. 739-777, 2011. doi: 10.1007/s11042-010-0635-7.

[6] M. Flores-Calero et al, “Ecuadorian traffic sign detection through color information and a convolutional neural network,” in 2020 IEEE ANDESCON, Oct. 2020, pp. 1-6. doi: 10.1109/ANDESCON50619.2020.9272089.
[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 386–397, Mar. 2017. doi: 10.48550/arxiv.1703.06870.
[8] M. A. Malbog, “MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation,” in 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2019, pp. 1–5. doi: 10.1109/ICETAS48360.2019.9117217.
[9] N. L. Tun, A. Gavrilov, and N. M. Tun, “Multi-Classification of Satellite Imagery Using Fully Convolutional Neural Network,” in 2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 2020, pp. 1-5. doi: 10.1109/ICIEAM48468.2020.9111928.
[10] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440. doi: 10.1109/CVPR.2015.7298965.
[11] L. Su, Y. Huang, and Z. Hu, “An empirical study on fully convolutional network and hypercolumn methods for UAV remote sensing imagery classification,” in IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, 2020, pp. 2567-2570. doi: 10.1109/IGARSS39084.2020.9323562.
[12] R. Davarzani and S. Mozaffari, “Robust Image Description with Weighted and Adaptive Local Binary Pattern Features,” in 2014 22nd International Conference on Pattern Recognition, 2014, pp. 1097–1102. doi: 10.1109/ICPR.2014.198.
[13] X.-H. Han, Y.-W. Chen, and G. Xu, “High-Order Statistics of Weber Local Descriptors for Image Representation,” IEEE Trans. Cybern., vol. 45, no. 6, pp. 1180–1193, 2015. doi: 10.1109/TCYB.2014.2346793.
[14] R. Sasidharan and D. Menaka, “Dynamic texture segmentation of video using texture descriptors and optical flow of pixels for automating monitoring in different environments,” in 2013 International Conference on Communication and Signal Processing, 2013, pp. 841–846. doi: 10.1109/iccsp.2013.6577175.
[15] L. Zhang, W. Zhou, J. Li, J. Li, and X. Lou, “Histogram of Oriented Gradients Feature Extraction Without Normalization,” in 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), 2020, pp. 252–255. doi: 10.1109/APCCAS50809.2020.9301715.
[16] S. A. Korkmaz, A. Akçiçek, H. Bínol, and M. F. Korkmaz, “Recognition of the stomach cancer images with probabilistic HOG feature vector histograms by using HOG features,” in 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), 2017, pp. 339–342. doi: 10.1109/SISY.2017.8080578.
[17] R. Diaz De Leon and L. E. Sucar, “Human silhouette recognition with Fourier descriptors,” in Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, 2000, vol. 3, pp. 709–712 vol.3. doi: 10.1109/ICPR.2000.903643.
[18] M. Fayyaz, M. Yasmin, M. Sharif, J. H. Shah, M. Raza, and T. Iqbal, “Person re-identification with features-based clustering and deep features,” Neural Comput. Appl., vol. 32, no. 14, pp. 10519–10540, 2020. doi: 10.1007/s00521-019-04590-2.
[19] H. Cai, J. He, L. Luo, and Y. Wu, “Research on unsupervised people re-identification based on k-means clustering,” in 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT), 2021, pp. 405–408. doi: 10.1109/ISCIPT53667.2021.00087.
[20] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, “Deep Learning for Person Re-identification: A Survey and Outlook,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 6, pp. 2872–2893, Jan. 2020. doi: 10.48550/arxiv.2001.04193.
[21] R. Layne, T. Hospedales, and S. Gong, “Person Re-identification by Attributes,” in Proceedings of the British Machine Vision Conference, 2012, pp. 24.1–24.11. doi: http://dx.doi.org/10.5244/C.26.24.
[22] X. Gu, X. Zou, J. Liu, and L. Zhang, “Person re-identification by using a method combining DPM and SVM,” in 2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2017, pp. 124–127. doi: 10.1109/ICCWAMTIP.2017.8301463.
[23] T. Goto, K. Tsuchida, and T. Nishino, “EPISODE: An Extreme Programming Method for Innovative Software Based on Systems Design,” in 2014 IIAI 3rd International Conference on Advanced Applied Informatics, 2014, pp. 780–784. doi: 10.1109/IIAI-AAI.2014.157.
[24] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian Detection: An Evaluation of the State of the Art,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 4, pp. 743–761, Apr. 2012. doi: 10.1109/TPAMI.2011.155.
[25] A. Olafenwa, “PixelLib,” 2021. https://github.com/ayoolaolafenwa/PixelLib
[26] A. Olafenwa, “Simplifying Object Segmentation with PixelLib Library,” 2021.
[27] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, no. Oct, pp. 2825–2830, 2011.
[28] G. Bradski, “The OpenCV Library,” Dr. Dobb’s J. Softw. Tools, 2000.
[29] L. Qi, J. Huo, L. Wang, Y. Shi, and Y. Gao, “MaskReID: A Mask Based Deep Ranking Neural Network for Person Re-identification,” Apr. 2018, doi: 10.48550/arxiv.1804.03864.
[30] N. A. Mandellos, I. Keramitsoglou, and C. T. Kiranoudis, “A background subtraction algorithm for detecting and tracking vehicles,” Expert Syst. Appl., vol. 38, no. 3, pp. 1619–1631, 2011. doi: https://doi.org/10.1016/j.eswa.2010.07.083.
[31] K. KC, Z. Yin, D. Li, and Z. Wu, “Impacts of Background Removal on Convolutional Neural Networks for Plant Disease Classification In-Situ,” Agriculture, vol. 11, no. 9, 2021. doi: 10.3390/agriculture11090827.
[32] X. Sun, N. Zhang, Q. Chen, Y. Cao, and B. Liu, “People Re-Identification by Multi-Branch CNN with Multi-Scale Features,” Proc. - Int. Conf. Image Process. ICIP, vol. 2019-Septe, pp. 2269-2273, Sep. 2019. doi: 10.1109/ICIP.2019.8803796.
[33] Ó. García-Olalla Olivera and E. Alegre Gutiérrez, “Descripción de textura en imágenes utilizando Local Binary Pattern (LPB),” 2016, Accessed: Jun. 27, 2022. [Online]. Available: http://hdl.handle.net/10612/11063.
[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems, 2012, vol. 25. [Online]. Available: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
[35] M. Jiang, Z. Li, and J. Chen, “Person Re-Identification Using Color Features and CNN Features,” in 2019 IEEE 4th Int. Conf. Image, Vis. Comput. ICIVC 2019, pp. 460–462, Jul. 2019. doi: 10.1109/ICIVC47709.2019.8980977.
[36] A. Yadav and D. K. Vishwakarma, “Person Re-Identification using Deep Learning Networks: A Systematic Review,” Dec. 2020. doi: 10.48550/arxiv.2012.13318.
[37] F. Chollet et al., “Keras,” GitHub, 2015. [Online]. Available: https://github.com/fchollet/keras
[38] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person Re-identification: Past, Present and Future,” Oct. 2016. doi: 10.48550/arxiv.1610.02984.
[39] Z. Jamali, J. Deng, J. Cai, M. U. Aftab, and K. Hussain, “Minimizing vehicle re-identification dataset bias using effective data augmentation method,” Proc. - 15th Int. Conf. Semant. Knowl. Grids Big Data, AI Futur. Interconnect. Environ. SKG 2019, pp. 127–130, Sep. 2019. doi: 10.1109/SKG49510.2019.00030.
[40] S. Madan et al., “When and how CNNs generalize to out- of-distribution category-viewpoint combinations,” arXiv, 2020. doi: 10.48550/ARXIV.2007.08032.
[41] Mart´ın Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” 2015. [Online]. Available: https://www.tensorflow.org/
[42] S. Yadav and S. Shukla, “Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification,” in 2016 IEEE 6th International Conference on Advanced Computing (IACC), Feb. 2016, pp. 78–83. doi: 10.1109/IACC.2016.25.
[43] D. Powers, “Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness and Correlation,” Mach. Learn. Technol., vol. 2, 2008.
[44] A. Schumann and E. Monari, “A soft-biometrics dataset for person tracking and re-identification,” in 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Aug. 2014, pp. 193–198. doi: 10.1109/AVSS.2014.6918667
[45] I. A. Siradjuddin, H. A. Achmani, and A. Muntasa, , “Adaptive Boosting Classifier for Pedestrian Attributes Identification with Color and Texture Features,” in 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM), Nov. 2018, pp. 275–279. doi: 10.1109/CENIM.2018.8710845