Person ReID through handcrafted features analysis

Andrea Sanchietti, Francesco Palandra, Riccardo Tittarelli

December 2022

1 Abstract

Re-identification (Re-ID) is the process of identifying people from different angles or cameras. In this report we propose a method for solving the Re-ID problem using classification algorithms to build separate models for three modalities: skeleton, clothing, and face data. We used an one-versus-all setting, the method was tested using the BIWI RGBD-ID dataset and various metrics including precision, recall, f1 score, accuracy, ROC, AUC, and DET for classification models, and average re-ranking accuracy for regression models, showing that they both are effective for Re-ID.

2 Introduction

Re-identification (Re-ID) is the process of identifying people from several angles or cameras. It is a difficult topic in computer vision and has many real-world applications, like crowd analysis and spying.

A Re-ID system's objective is to identify a person's key characteristics that can be utilized to set them apart from other people. This can be accomplished by using a machine learning model that has been trained on a dataset of individual images or video frames to detect people in fresh photo or video frames.

3 Related Work

The first thing that we did was to search related works. While looking for features that could help a model to re identify a person, we found four works that helped us:

- [7] *Person re-identification based on deep learning — An overview.* and [6] *Person Re-Identification by Discriminative Selection in Video Ranking* were useful to have an idea of the problem.

- We then found two papers ([5]*Predicting types of clothing using SURF and LDP based on Bag of Features* and [4] *Classification of clothing with weighted SURF and local binary patterns*) that gave us the idea of exploiting clothes information through Local Binary Patterns to recognize people.

- We then looked for face recognition in cctv videos and found these two works: [1] *Face Recognition in CCTV Systems* and [8] *Face Feature Extraction Based on Principle Discriminant Information Analysis.*

- We finally found a work about re identification based on skeleton: [3]*A Self-Supervised Gait Encoding Approach With Locality-Awareness for 3D Skeleton Based Person Re-Identification.* We used the dataset and our idea was to extract measures from the 3D skeleton and use them to re identify a person.

4 Proposed method

The approach taken to solve the Re-ID problem involves using several different classification algorithms to build separate models for three different modalities: skeleton data, clothing data, and face data. The approach is a one-versus-all, where the model tries to learn how to separate the target class from other people.

The main reason for using machine learning algorithms rather than neural networks is that the dataset we are using is small, and we wanted to use simple, fast algorithms that can be trained quickly. Machine learning models are generally faster to train and easier to interpret than neural networks, and they can be effective for many types of classification or regression tasks.

5 Dataset and Benchmark

The dataset used was BIWI RGBD-ID Dataset[2] composed of a RGB-D dataset of people targeted to long-term people re-identification from RGB-D cameras. It contains 50 training and 56 testing sequences of 50 different people. The dataset includes synchronized RGB images, depth images, persons' segmentation maps and skeletal data, in addition to the ground plane coordinates.

We generated 14 samples made of 1550 entries. Each entry consisted in skeleton, clothes and face data for a given frame. The split was 50 frames from the target that we want to re identify and 1500 from random people. Since facial information is not always available, entries with facial landmarks are usually around 400. Figure 1 and 2 show the information extracted by the dataset. For skeleton recognition, we used the 3D skeleton to extract measures of the person. We then used the LBP algorithm on clothes and extracted an histogram. For facial recognition we used distances of landmark features because the HoG were computed too slowly and the LBP were not robust enough.

In order to test our models, we fit different random samples and then get an average of precision, recall, f1 score, accuracy, ROC, AUC and DET for classification models, while we computed the average re-ranking accuracy for regression models.

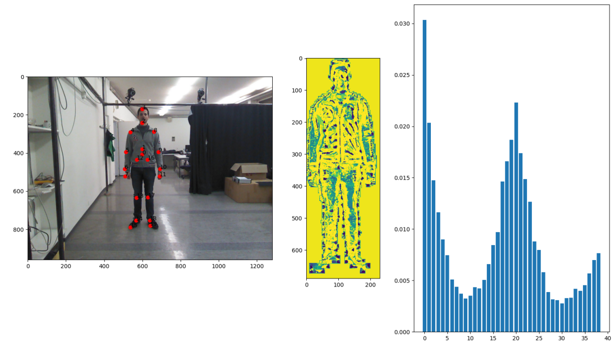

Figure 1: The figure shows an image with the 2D skeleton points, the LBB features and the histogram version of the LBP features with 40 buckets and a radius of 10

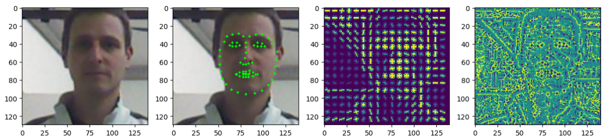

Figure 2: The figure shows a crop of a face, the landmark, the HoG and the LBP features.

Given these splits, we decided to decompose and plot the data (we used PCA) so that we could see how it distributes in space. It is clearly visible in figure 3 that samples from the target person tents to cluster or distribute on a line for both skeleton and face. Clothes information are instead all compressed on a line. This behaviour is probably given by the distribution that LBP features from clothes tents to assume if transformed to histograms (see fig. 1)

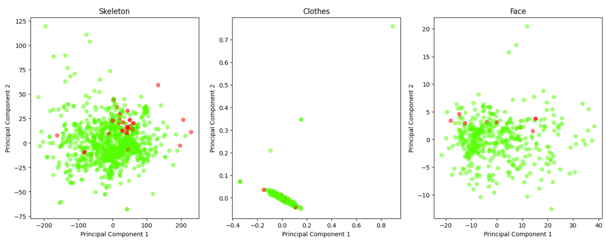

Figure 3: Projection of a sample on two dimensions for skeleton, clothes and face.

6 Experimental results

Firstly, we tested different models for classification. The tested models are: K-Nearest Neighbors, Decision Trees, Random Forests, Logistic Regression, Gaussian Naive Bayes and Support Vector Machines. We used a grid search for each one of these models in order to find the best parameters. Then we evaluated each model on a test set as large as the training set. Figure 4 shows the ROC curve. We can notice that both K-NN and RF performed very well for skeleton and clothes while SVC and Logistic Regression learned well face features. However, we are also interested in the error rate of our models because want to lower the rate of False Positive for our application in an hypothetical real context.

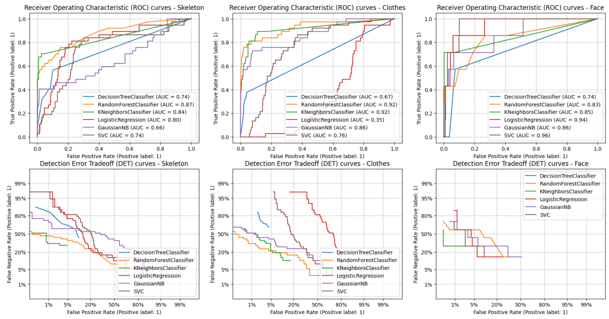

Figure 4: Projection of a sample on two dimensions for skeleton, clothes and face.

We also plotted the confusion matrix for each model in order to evaluate performances. Following (fig 5) the matrix of the models that we choose for the final application.

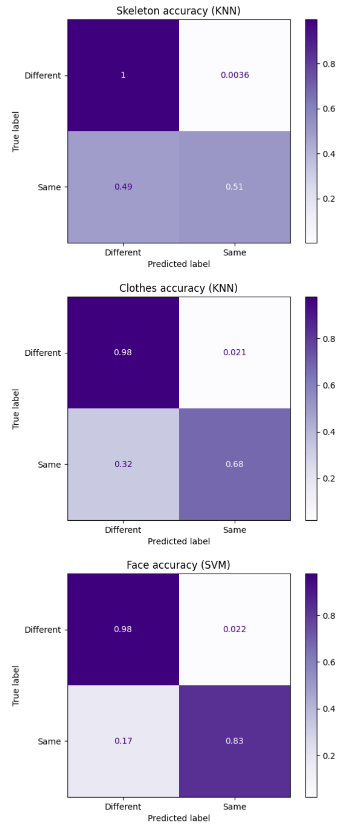

Figure 5: Confusion matrices. Each matrix's name indicates the features, the model used and the parameter optimized.

Given these information, we selected those models with a low false positive rate and experimented two different policies to merge them. The one that gave the best precision was the "argmax" policy, that consists in evaluating probabilities of each model, summing them and taking the argmax. The second policy is the "or" and consists in taking the or of the predictions: if one of the three models thinks that the input is the same person, then we output 1, otherwise we output 0. This second version is less conservative and tends to lower the precision and raise the recall. This kind of behaviour will let more "true" people in, but will also give a slightly higher false acceptance rate.

In table 1 a comparison between the two proposed policy is shown.

Figure 6 shows the roc and det curves of the model.

	False Accept	True Accept
OR	3.3/1500	30/50
ARGMAX	0/1500	22/50

Table 1: Comparison between the two policies for the final model.

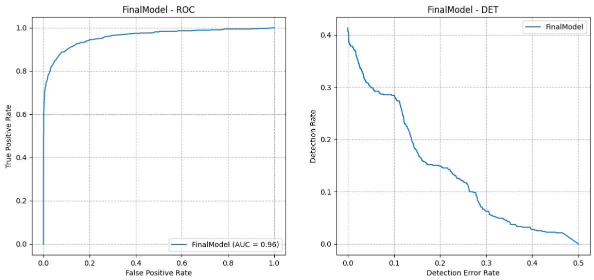

Figure 6: ROC and DET of the final model. It uses probabilities so it does not vary with the model's policy.

We also thought that it was possible to learn a "latent space" that ensembles that person's features. The idea is that a regressor can learn how to optimize the output of a linear combination so that features from the person that we want to re-identify will output a low value (0) while the other people will output an higher value (1).

Given that, we used a polynomial feature with degree 2, a standard scaler and finally an AdaBoostRegressor with SVR as base learner in a pipeline in order to learn the features of a person. To measure if the model learned the target's features, we passed to the model the tests, sorted them based on the output and then plotted (fig. 7) the number of true labels as we increase the threshold of samples:

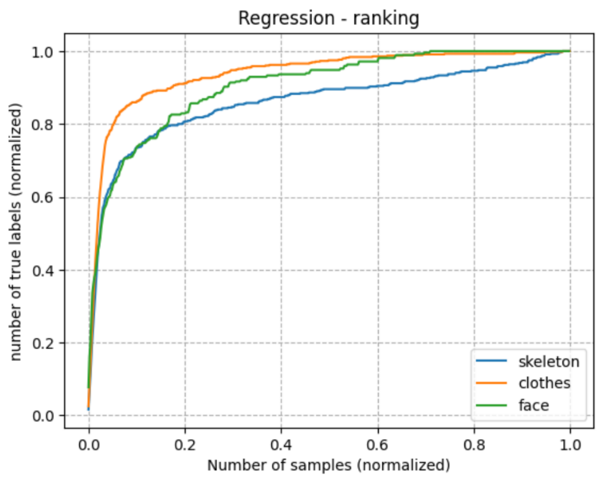

Figure 7: plot of the number of true positive labels (normalized) on the y axes and the number at which we put the threshold (normalized) on the x axis.

these plots confirms that the regressor is able to lower the value of the target even in the test set. By combining these three regressors we can achieve a good result 8:

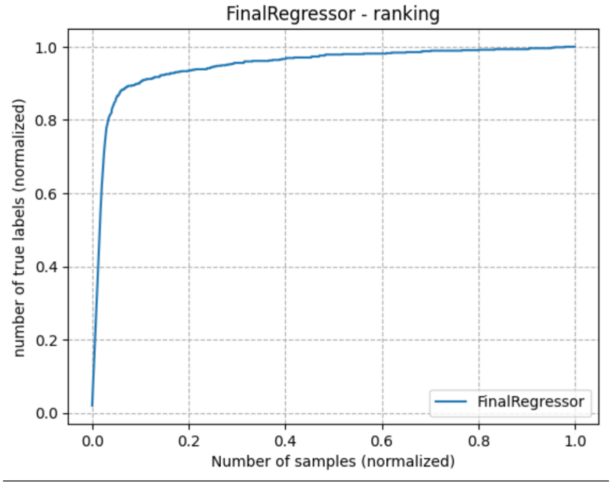

Figure 8: plot of the number of true positive labels (normalized) on the y axes and the number at which we put the threshold (normalized) on the x axis for the FinalRegressor.

This curve (8) is slightly better than the one given by the clothes regressor in figure 7, making this last model a good choice if used together with a threshold (it could be learned or set by the user after the query).

7 Conclusions and Future Work

In the future, the main objective is making the system live so that it can be deployed in the real world. Alternatively, to further increase the performance without resorting to neural networks it may be worth investigating the use of other modalities, such as gait or body shape. Finally, it would be interesting to test the proposed method on a larger and more diverse dataset to further evaluate its effectiveness.

References

[1] Jinu Celine and Sheeja Agustin A. "Face Recognition in CCTV Systems". In: *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. 2019, pp. 111–116. DOI: 10.1109/ICSSIT46314.2019.8987961.

[2] Matteo Munaro. *BIWI RGBD-ID Dataset*. July 2013. URL: http://robotics.dei.unipd.it/reid/index.php/8-dataset/2-overview-biwi.

[3] Haocong Rao et al. "A Self-Supervised Gait Encoding Approach With Locality-Awareness for 3D Skeleton Based Person Re-Identification". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2022), pp. 6649–6666. DOI: 10.1109/TPAMI.2021.3092833.

[4] Wisarut Surakarin and Prabhas Chongstitvatana. "Classification of clothing with weighted SURF and local binary patterns". In: *2015 International Computer Science and Engineering Conference (ICSEC)*. 2015, pp. 1–4. DOI: 10.1109/ICSEC.2015.7401409.

[5] Wisarut Surakarin and Prabhas Chongstitvatana. "Predicting types of clothing using SURF and LDP based on Bag of Features". In: *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 2015, pp. 1–5. DOI: 10.1109/ECTICon.2015.7207101.

[6] Taiqing Wang et al. "Person Re-Identification by Discriminative Selection in Video Ranking". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.12 (2016), pp. 2501–2514. DOI: 10.1109/TPAMI.2016.2522418.

[7] Wenyu Wei et al. "Person re-identification based on deep learning — An overview". In: *Journal of Visual Communication and Image Representation* 82 (2022), p. 103418. ISSN: 1047-3203. DOI: https://doi.org/10.1016/j.jvcir.2021.103418. URL: https://www.sciencedirect.com/science/article/pii/S1047320321002765.

[8] Jianqin Yin, Yuelong Li, and Jinping Li. "Face Feature Extraction Based on Principle Discriminant Information Analysis". In: *2007 IEEE International Conference on Automation and Logistics*. 2007, pp. 1580–1584. DOI: 10.1109/ICAL.2007.4338824.