

Article

# Image-Based Somatotype as a Biometric Trait for Non-Collaborative Person Recognition at a Distance and On-The-Move

Antonios Danelakis \*  and Theoharis Theoharis

Department of Computer Science, Norwegian University of Science and Technology, 7030 Trondheim, Norway; theotheo@ntnu.no

\* Correspondence: antonios.danelakis@ntnu.no

Received: 13 May 2020; Accepted: 13 June 2020; Published: 17 June 2020



**Abstract:** It has recently been shown in Re-Identification (Re-ID) work that full-body images of people reveal their somatotype, even after change in apparel. A significant advantage of this biometric trait is that it can easily be captured, even at a distance, as a full-body image of a person, taken by a standard 2D camera. In this work, full-body image-based somatotype is investigated as a novel soft biometric feature for person recognition at a distance and on-the-move. The two common scenarios of (i) identification and (ii) verification are both studied and evaluated. To this end, two different deep networks have been recruited, one for the identification and one for the verification scenario. Experiments have been conducted on popular, publicly available datasets and the results indicate that somatotype can indeed be a valuable biometric trait for identity recognition at a distance and on-the-move (and hence also suitable for non-collaborative individuals) due to the ease of obtaining the required images. This soft biometric trait can be especially useful under a wider biometric fusion scheme.

**Keywords:** somatotype trait; somatotype as biometric; biometrics; full body image; deep learning; inception V3 network; siamese network; fusion scheme

## 1. Introduction

Biometric recognition at a distance under high security requirements (e.g., border control posts) remains a challenging task and becomes even more so nowadays that “on the move” scenarios are becoming a trend (<https://www.schengenvisainfo.com/news/frontex-testing-biometrics-on-the-move-border-check-technology-at-lisbon-airport/>). Biometric recognition accuracy largely depends on a set of (complementary with respect to the failure cases) biometric traits under a fusion scheme [1]. It is therefore crucial to expand the set of available biometric traits, especially with traits that can be easily captured by standard sensors, at a distance and on-the-move. One such trait is the somatotype [2]. The somatotype focuses on the measurement of structural aspects of the human body and recent Re-Identification (Re-ID) works show that it can be captured from still full-body images [3].

This paper proposes the use of the human somatotype as a new soft biometric trait [4] based on full-body still images. Current works of the state-of-the-art exploit the somatotype feature for resolving the Re-ID problem which is different to the biometric approach; the Re-ID procedure uses a gallery which is not predefined, meaning that no enrollment has taken place. In this sense, the Re-ID cannot be considered as a biometric problem [5]. Nonetheless, it is possible to compare against Re-ID methods by assuming a fixed gallery and using common biometric evaluation measures; this is the approach taken here and we show that the proposed methods outperform the state-of-the-art in terms of verification, while remaining quite comparable in the identification case.

The somatotype is studied under two common biometric recognition scenarios: (i) identification, and (ii) verification. For the identification case, we implement an appropriately modified variation of the Inception V3 network, while for the verification case, a Siamese network variation is proposed. The resultant deep networks have been trained and assessed on four popular publicly available benchmark datasets: (i) *CUHK03* [6], (ii) *RAiD* [7], (iii) *Market-1501* [8] and (iv) *RGBD-ID* [9,10]. In addition, a much larger dataset has been created by unifying the aforementioned ones and has been used for testing purposes. Finally, an additional synthetic dataset (*SOMASet* [3]) which is appropriate for testing under different camera angles and cloth variations, is recruited.

### Contributions

The main contributions of the present work are:

1. The proposal of the full-body image-based somatotype as a biometric trait suitable for non-collaborative person recognition at a distance and on-the-move, as well as suitable experiments to verify this.
2. The implementation of two fine-tuned pre-trained deep learning networks (Inception V3 network and Siamese network) for person recognition based on somatotype, trained on publicly available benchmark somatotype datasets.
3. The provision of a somatotype recognition benchmark for this and similar techniques. This benchmark includes a unified somatotype dataset, and the definition of the validation and training sets.
4. Detailed evaluation of the proposed schemes, even for challenging scenarios including camera angle and cloth variations. These evaluations show that the proposed biometric recognition methods based on the somatotype generally outperform that state-of-the-art (which consists of Re-ID methods).

The rest of the paper is organized as follows: After reviewing relevant previous work (Section 2), we describe the implemented methods (Section 3). In Section 4, the experimentation process is presented and corresponding results are illustrated. Section 5 records important observations extracted from the experimental procedure. Section 6 highlights interesting future challenges while Section 7 concludes the paper.

## 2. Related Work

Re-identification work (Re-ID), based on full-body still images, implicitly uses the somatotype feature, especially when training is involved. On the other hand, to the extend of our knowledge, somatotype has been hardly used as a biometric trait for person recognition. Thus, the related work presented in this section, is focused on full-body image-based Re-ID techniques, the closest relative and it shall be used to be compared against our work. Most Re-ID techniques are based on deep learning with Convolutional Neural Networks (*CNNs*), using the full body images as input and the recognised identity as output, being the most popular [11–13]. A number of more complex *CNN* variations, especially focused on the Re-ID problem, were also introduced; In [14], a duplicate *CaffeNet* is used. The work in [15], implements a more complex convolutional scheme *OSnet*, in terms of number of convolutions and kernel sizes, than typical *CNNs*. Finally, in [16], an extra layer (*Eigenlayer*) is added to the typical *CNN* architecture for improving the feature representation. In the aforementioned works, there is usually a pre-processing step which resizes the initial images to the appropriate size expected by the corresponding network. There are also some cases, where data augmentation takes place in order to increase the training samples.

The main drawback with *CNNs* is that, although they can be well trained and can learn global features, they may lose important local features. Combining learning from global and local features is critical in order to overcome issues related to pose changes, occlusions and non-rigid deformations of the human body. To this end, several models have been proposed such as the Batch DropBlock

Network (*BDB*) [17], the part-based gradation regularization model (*PGR*) [18], the Multi-Scale ContextAware Network (*MSCAN*) [19], the Multi Task Deep learning network (*MTDnet*) [20], the Attribute-Aware Attention Model (*A<sup>3</sup>M*) [21], the *AlignedReID* method [22], the pyramidal framework [23], the Integration CNN (*ICNN*) [24] and the joint attention person Re-ID (*JA-ReID*) architecture [25].

Furthermore, there are some more methodologies which follow different directions towards solving the Re-ID problem. In [26], the so-called Part-based Convolutional Baseline (*PCB*) scheme for learning part-level features of the images, instead of global ones, is presented. In [27] a re-ranking process is implemented on the initially retrieved images for leveraging the final result. The methodology in [28] proposes a scheme which resolves the so-called Small Sample Size problem (*SSS*) [29] (i.e., the training data sample size is much smaller than the feature sample size) which is a common issue that typical distance metric learning techniques have to address. [30] proposes a spatio-temporal person Re-ID (*st-ReID*) framework that takes into account not only the visual semantic information but the spatio-temporal information as well using consecutive images. A new feature (Local Maximal Occurrence) and a metric learning method (Cross-view Quadratic Discriminant Analysis) is proposed in [31]. In [3], a synthetic dataset (*SOMAs*) has been used to train an Inception V3-based network (*SOMAnet*) in order to model additional discriminative information than just the outfit. [32] studies small-sized randomly initialized models. [33] proposes the adaptation of a limited supervised multi-camera Re-ID setup, thus improving the hardware of the entire process. The method of [34] proposes the *Auto-ReID* framework which enables the automated finding of an efficient and effective CNN architecture for Re-ID. [35] implements a model which involves a generative module for encoding a person's appearance and structure based on full-body images, and a discriminative module that shares the appearance encoder with the generative module. Finally, the work of [36] collects and evaluates training lessons from the literature to improve performance.

Finally, it should be highlighted that there are very interesting state-of-the-art works which either perform Re-ID based on different modalities than full-body still images derived by other sensors than conventional cameras [37–39], or combine the somatotype trait with other biometric modalities (i.e., face) in order to perform improved biometric recognition [40]. Nonetheless, the scope of this paper is focusing on full-body still images derived by standard 2D cameras.

### 3. Methods

After discussing the somatotype as a biometric trait, we present its use in both identification and verification scenarios along with proposed deep learning based schemes.

#### 3.1. Somatotype as a Biometric Trait

Somatotype focuses on the measurement of structural aspects of the human body, and includes three main somatotype categories [41]: (i) Ectomorph: long and lean body figures, with little body fat and muscle; (ii) Mesomorph: athletic, solid and strong body figures; (iii) Endomorph: body figures with lots of body fat. Although the human somatotype has (implicitly) achieved remarkable results in the Re-ID field (by using full-body images of humans), as presented in Section 2, it has never before been used in recognition scenarios, to the best of our knowledge. The Re-ID procedure uses a not pre-defined gallery (no enrollment takes place), thus it cannot be considered as a biometric problem [5]. It was not until the work of Barbosa [3] that somatotype was explicitly recognized as a useful trait for the Re-ID problem. In [3], synthetic data with varying somatotypes were used, to show that this trait is capable of re-identifying human beings even after a change in apparel, based on learning somatotypes. It thus makes sense, as a continuation of the aforementioned work, to explore the full-body image-based somatotype trait in non-collaborative person recognition at a distance and on-the-move, potentially opening up a new biometric field.

Since the somatotype is based on a simple whole body image, it can be captured using off-the-shelf cameras, from a distance and on the move. Such cameras are already in place in areas where security

is controlled (e.g., airports, ports, malls). Somatotype is also available in databases of law-enforcement agencies which typically include a full-body image of subjects.

### 3.2. Identification Scenario

This scenario studies the usability of the somatotype trait for person identification. A gallery dataset of identities, containing one or more somatotype instances for each identity is assumed. This gallery dataset would normally be the result of an enrollment process and is used for training. Unknown query somatotype instances are then presented and the trained system is expected to map these queries onto identities of the gallery.

#### 3.2.1. Inception V3 Network Architecture for Identification

The Inception V3 network [42], which has proven extremely successful in classification problems (See *ReCeption* method in classification part of Task 2a of ILSVRC 2015 rankings: <http://image-net.org/challenges/LSVRC/2015/results>), has been implemented for the identification scenario. This is an improved version of a series of deep networks starting with Inception V1 network, widely known as GoogleNet [43]. GoogleNet recruits convolution filters with multiple sizes which operate on the same level. This is done in order to learn salient parts of the input that can be of various sizes. Inception V2 [42] factorizes a  $5 \times 5$  convolution into two  $3 \times 3$  convolution operations to improve computational speed and reduce the representational bottleneck.

Finally, Inception V3 has adopted some upgrades regarding the convolution filter size, the auxiliary classifiers and the prevention of over-fitting. Furthermore, this version of Inception is capable of encoding more attributes related to the identity of a person (i.e., height, obesity and gender) in addition to outfit and body physique [3]. This is the main reason that the aforementioned deep network was chosen. Figure 1 illustrates the Inception V3 network architecture as implemented in this work. There are three different types of modules in the Inception V3 network: (i) Inception *A*, (ii) Inception *B* and (iii) Inception *C* module. All modules are appropriately designed for generating discriminatory features and for reducing the number of trainable parameters. Each module is composed of several convolutional, batch normalization, *ReLU* and pooling layers. Modules *A* and *C* use small convolutional layers which are responsible for the reduction of the trainable parameters of the network. On the other hand, module *B* uses mostly large convolutional layers for feature extraction. As illustrated in Figure 1 three Inception *A* modules, four Inception *B* modules and 2 Inception *C* modules are stacked sequentially (indicated by notation “X3”, “X4” and “X2” in the figure respectively).

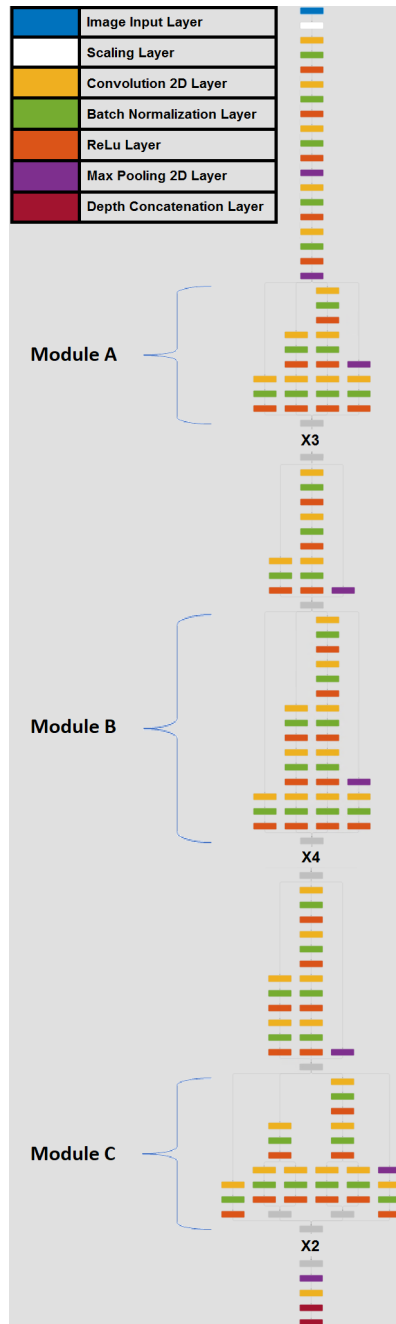


Figure 1. Inception V3 network architecture as implemented in the current work.

### 3.2.2. Identification Procedure

For each training sample  $x$ , the Inception V3 network computes the probability of it belonging to each class  $k \in 1 \dots K$  as indicated in Equation (1). Cross entropy is defined as the loss function of the network (see Equation (2)).

$$p(k|x) = \frac{\exp(z_k)}{\sum_{i=1}^K \exp(z_i)} \tag{1}$$

$$l = - \sum_{k=1}^K \log(p(k)) \cdot q(k) \tag{2}$$

where  $z_i$ ,  $p(k)$  and  $q(k)$  correspond to the output of the network, the predicted class distribution and the ground-truth class distribution respectively.

Transfer learning has been implemented on the pre-trained Inception V3 network. To this end, the final two layers of the pre-trained network have been adapted to each dataset used in the experimentation process. The training set of each dataset is considered as the identities gallery database. The number of outputs (i.e., classes) of the final network equals the number of different identities in the gallery. Each query results in a confidence score in every output of the network and the query is mapped to the class with the highest confidence score. Should the highest confidence score does not exceed a specified threshold  $\epsilon$ , then the identity of the query is considered to be unknown (i.e., no such identity in the gallery).

### 3.2.3. Verification through Identification

The aforementioned scheme can be adopted for verification. In this case, a post-processing operation is implemented. The identity of an individual is verified if the output of the system for the class of the claimed identity has a confidence score higher than a predefined threshold  $\epsilon$ .

### 3.3. Verification Scenario

This scenario studies the usability of the somatotype trait for identity verification. A gallery biometric instance (representing the claimed identity) and a query instance (representing the claimant) are compared by a pre-trained system, producing a similarity score. A confidence threshold  $\epsilon$  on this score decides whether the claimed identity is verified.

#### 3.3.1. Siamese Network Architecture for Verification

For the verification scenario a Siamese network [44] is adopted. Due to the nature of these networks, they are ideal for verification tasks of reduced dimensionality [45]. Siamese networks have mainly been used for face verification, with DeepFace [46] and FaceNet [47] being the most popular implementations. Within the context of the current work, their implementation is extended to somatotype-based verification.

Siamese networks are fed with two somatotype biometric instances and produce a similarity score as output in the interval  $[0, 1]$ . A similarity score close to 0 indicates that the identity of the two inputs is different, while close to 1 that it is the same. The response time is extremely fast. As mentioned already, in the case of identity verification, the first biometric instance will be taken from a gallery dataset which is constructed during an enrollment process. The second biometric instance will typically be captured by a biometric sensor (2D camera) on-the-fly while an individual walks through a security area, having declared a claimed identity. A one-to-one comparison takes place; a similarity score over a predefined confidence threshold  $\epsilon$  confirms the person's identity.

It should be highlighted that, re-training of the Siamese network for a new gallery dataset is not required (e.g., when more identities are added to the gallery). This is because the One-Shot-Learning [48] procedure is used and happens for two reasons:

1. Unlike the Inception V3 network, whose output is a  $k$ -dimensional vector (where  $k$  is the number of identities) and this dimension can be modified as the number of identities changes, the Siamese network outputs a scalar similarity score in the interval  $[0, 1]$  for any input pair. This means that the output dimensionality remains constant at all times.
2. The Siamese network is actually trained to learn the similarity function instead of the biometric data itself. This means that once the network is trained on a specific type of data, it knows how to compute the similarity between new biometric instance pairs.

Unlike the typical deep learning networks, Siamese networks are trained to learn how to compare two images rather than the training images themselves. Thus, they only need a small training sample

to be trained upon. In theory, this could even work with a training set containing only one instance per identity.

Each of the two biometric input instances of the Siamese network are fed into two separate but identical CNNs. Then, the output of the CNNs is merged using the  $L_1$  distance (see Equation (3), where  $l$  denotes the output dimension of the CNNs). The estimated  $L_1$  distance value is passed through a *sigmoid* function which produces the similarity score as the network's output. Each CNN begins with three consecutive pairs of convolutional and max pooling layers. The first convolution layer uses 64 filters of dimension  $10 \times 10$ . The second convolution layer uses 128 filters of dimension  $7 \times 7$ . Finally, the third convolution layer uses 128 filters of dimension  $4 \times 4$ . All three layers use *ReLU* as their activation function. The third max pooling layer is succeeded by another convolutional layer using 256 filters of dimension  $4 \times 4$  and *ReLU* activation function. A flattening layer follows before the final dense layer with the *sigmoid* activation function. Figure 2 illustrates the Siamese network architecture as implemented in this work. Figure 3 illustrates the architecture of each CNN utilized within the Siamese network.

$$L_1 = - \sum_{i=1}^l |CNN(out)_1^i - CNN(out)_2^i| \quad (3)$$

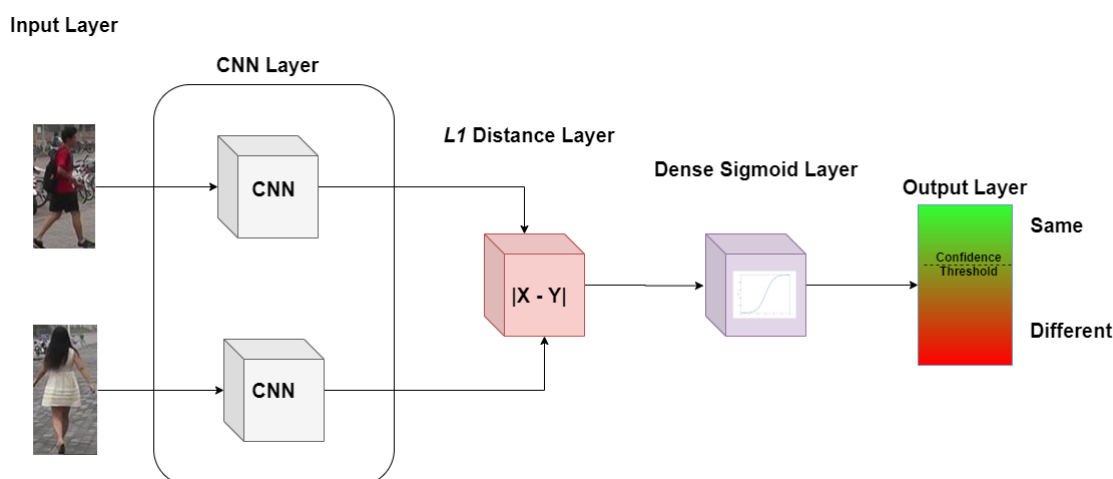


Figure 2. Siamese network architecture as implemented in the current work.

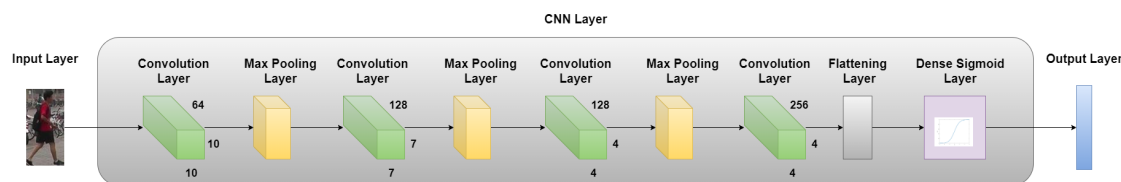


Figure 3. Convolutional Neural Network (CNN) architecture as utilised within the Siamese network of the current work.

### 3.3.2. Identification through Verification

The aforementioned scheme can be also utilised for identification purposes. In this case, the query will be compared against every single instance of the gallery set. To this end, each biometric pair that is formed by the query and each instance of the gallery set will be consumed by the Siamese network as input. If the highest similarity score achieved exceeds a specified confidence threshold  $\epsilon$  and the gallery part of the input pair that achieves the higher score is of the same identity as the query part, then the identity of the query is verified by the system.



## 4. Experiments

Experimentation was conducted on four popular, publicly available datasets, one additional unified dataset and one dataset appropriate for testing variations in camera angle and apparel. The current section illustrates the datasets and the corresponding performance results and gives a comparison of the proposed framework against the state-of-the-art methodologies for each scenario. The methodologies that we compared against came from the Re-ID field. Despite the fact that Re-ID and recognition are not the same tasks [49], we chose to compare against such techniques since, to the extent of our knowledge, there were no previous image-based somatotype recognition techniques to compare against and Re-ID was the closest relative. This is also indicated in [50]. Furthermore, the same evaluation metrics could be recruited.

### 4.1. Datasets

A brief description of the datasets that were used for experimentation follows. *CUHK03*: *CUHK03* is the first relevant dataset large enough for deep learning [6]. It contained 13,164 images of varying dimensions and consists of 1467 identities. The images were taken from five cameras with different acquisition settings. The dataset version where pedestrians in the raw images were manually cropped was selected in order to keep the process simple. Testing (100 identities) and training (1367 identities) partitions were also provided and used for assessment.

*RAiD*: *RAiD* [7] is another related dataset where a limited number of identities (41) was captured and contained 6920 images in total. The dimension of each image was  $128 \times 64$ . The images were collected using a four camera setting (two indoors and two outdoors). The subjects were randomly divided into two sets, training (21 identities) and testing (20 identities).

*Market-1501*: *Market-1501* is the largest relevant real-image dataset [8]. It contained 1501 identities and over 32,668 images. The resolution of the images was  $128 \times 64$ . Five high-resolution and one low-resolution camera were recruited for acquisition. The training set consisted of 750 identities while the testing set consisted of 751 identities.

*RGBD-ID*: *RGBD-ID* [9,10] dataset was captured using *RGB-D* cameras. Although it only included 11 identities, *RGBD-ID* contains many instances for each identity as well as images of the same identities wearing different clothes, which made this dataset very challenging. The total number of images was 22,845 and the image dimension was  $640 \times 480$ . Testing and training images were also provided.

Figures 4–7 illustrate some example images from the four datasets.

*Unified dataset*: We created the *Unified* dataset that combined all the aforementioned publicly available datasets with the aim of maximizing the number of identities, which had a great influence on training and testing, as well as diversifying across different acquisition devices and protocols. In addition, all unique characteristics of each dataset were kept and combined, thus, creating a more challenging dataset. This dataset contained six randomly selected image instances for each identity of the initial datasets. This number was selected two reasons: (i) on the one hand, it excluded just a small number of available identities (identities with less than six recorded instances are skipped but there are only just a few such identities), and (ii) it was a realistic number of images that could be recorded during an enrollment procedure (enrollment in real-world conditions should not be a time-consuming process for the enrolled individual [51]). In total, there were 2939 identities and 17,634 images. The dimension of the images was resized to  $299 \times 299$  which was the standard dimension of the Inception V3 network input. This was done by applying bilinear interpolation (i.e., the output pixel value was a weighted average of pixels in the nearest  $2 \times 2$  neighborhood). Out of the six instances per identity, five were randomly selected as the training set. The remaining instance belonged to the testing set. Furthermore, there was the case where an identity was considered to be unknown. In that case, all six instances of the identity belonged to the testing set. Due to the terms of use of the four initial datasets, the *Unified* dataset could not be directly re-distributed. Nonetheless, we provided the filenames or indices of the selected images, so that one could easily reproduce the dataset.





**Figure 4.** *CHUK03* dataset exemplary samples.



**Figure 5.** *RAiD* dataset exemplary samples.



**Figure 6.** *Market-1501* dataset exemplary samples.



**Figure 7.** *RGBD-ID* dataset exemplary samples.

*SOMAs*et dataset: Finally, the *SOMAs*et dataset [3] (<https://www.kaggle.com/vicolab/somaset>) was used in order for the proposed scheme to be tested specifically under different clothing and pose variation scenarios. *SOMAs*et is a synthetic somatotype dataset which included 50 identities (25 males and 25 females). It also accounted for different ethnicities. Each identity wore eleven sets of clothes and assumed 250 different poses, over an outdoor background scene, with varying illumination conditions. This made *SOMAs*et ideal for testing the clothing and pose variation effect on the proposed deep learning scheme. Each image in the dataset had dimensions of  $200 \times 400$  pixels. Figure 8 illustrates example images of *SOMAs*et. Table 1 sums up all the recruited datasets along with their main characteristics. It is important to highlight that, although all datasets contained multiple poses for each subject included, only *SOMAs*et provided a standard pattern to facilitate the elaboration, using programming procedures, of both the pose and clothing variation in the defined experimentation protocol. That is why the aforementioned dataset was chosen for the corresponding experiments.



Figure 8. SOMAset dataset example images.

Table 1. Sum up of the datasets used and their main characteristics.

Dataset	Number of Identities	Number of Images	Image Dimensions	Training/Testing Partition Available	Pose Facilitation	Cloth Variation
CUHK03 [6]	1467	13,164	Varying	Yes	No	No
RAiD [7]	41	6920	128 × 64	Yes	No	No
Market-1501 [8]	1501	32,668	128 × 64	Yes	No	No
RGBD-ID [9,10]	11	22,845	640 × 480	Yes	No	Yes
Unified	2939	17,634	299 × 299	Yes	No	Yes
SOMAset [3]	50	100,000	200 × 400	No	Yes	Yes

## 4.2. Results

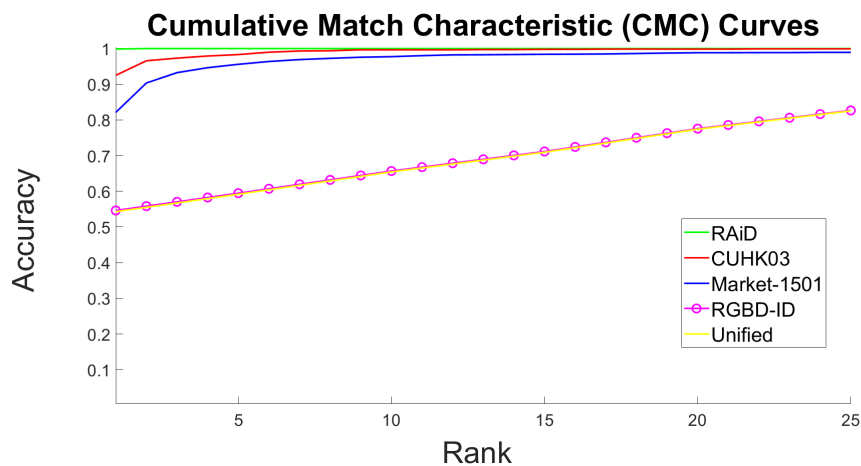
### 4.2.1. Identification Scenario

The identification module of the Inception V3 network, as described in Section 3, is strictly used experimenting on the identification scenario. The identification experiments were conducted using MATLAB toolboxes on a Windows 10 PC with an Intel i9 CPU at 2.30 GHz, 32.00 GB RAM and a NVIDIA Quadro T2000 GPU of 4 GB memory.

Table 2 illustrates the accuracy at rank-1, 5 and 10 of the proposed scheme. Fortunately, it is possible to compare the proposed identification methodology against state-of-art Re-ID methods since rank- $n$  rate is a common way of evaluating Re-ID methods; they are essentially considered as retrieval problems. The Cumulative Match Characteristic (CMC) curves of the proposed somatotype identification system for all datasets are illustrated in Figure 9 and confirm the results presented in Table 2. Note that, the best curve was achieved for the RAiD dataset (green color). The curve which corresponds to the RGBD-ID dataset (magenta color) almost overlaps with the one of the Unified dataset (yellow color). Nonetheless, the CMC curve of the Unified dataset was a bit lower which made it the most challenging case to handle. Table 3 compares the accuracy (at most popular ranks of the state-of-the-art, i.e., rank-1, 5 and 10) of the proposed scheme against recent top-ranked state-of-the-art methods for each dataset. 'N/A' stands for not available value. Some methodologies achieved better results on one dataset (i.e., Market-1501) but worse on another (i.e., CUHK03) compared to the proposed method. It is worth highlighting that all our results were achieved in a single query setting without using any re-ranking algorithms.

**Table 2.** Accuracy at different ranks, per dataset, as achieved by the Inception V3 network.

Dataset	Rank-1	Rank-5	Rank-10
CUHK03	92.50%	98.31%	99.65%
RAiD	99.86%	100.0%	100.0%
Market-1501	82.11%	97.73%	99.58%
RGBD-ID	54.62%	59.46%	65.68%
Unified	54.28%	59.16%	65.42%

**Figure 9.** Cumulative Match Characteristic (CMC) curves per dataset as achieved by the Inception V3 network.

#### Clothing and Pose Variation Effects

Variations in attire and pose are probably some of the most challenging in recognizing somatotype. To this end, experiments have been conducted on the *SOMASet* dataset, which includes such variations, in order to see how the proposed scheme deals with very challenging data.

**Effect of Clothing Variation** All 250 poses for each identity wearing a specific outfit were chosen as the training set. As testing set, a subset (10%) of the poses for each identity wearing an outfit different to the one corresponding to the same identity in the training dataset, were chosen. Thus, all queries are identities with different outfit to the one used in the training of the system on the same identity. This was significantly harder than the scenario employed in the other state-of-the-art methods using the *RGBD-ID* dataset where only a subset of the query identities wore different clothes to the ones in training set. The Rank-1 score dived to 15.70%, highlighting the fact that distinguishing identity based on somatotype under different clothes is a major challenge. This is the main reason for the performance decrease of the proposed scheme in the *RGBD-ID* and the *Unified* datasets. Table 4 illustrates the rank scores of this experiment.

**Effect of Pose Variation** In the pose variation case, the experiments were split into three sub-cases: (i) frontal pose, (ii) back pose and (iii) side pose. For each sub-case, all images of a specific pose (frontal, back or side) for each identity wearing a specific outfit were chosen as training set. As testing set, a subset (10%) of the remaining poses for each identity wearing the same outfit as in the training dataset, were chosen. Thus, all queries are identities with the same outfit but in a different camera angle that the one used for the training of the system on the respective identity. This was significantly harder than the scenarios of mixed poses employed in the other state-of-the-art methods. Table 5 illustrates the accuracy at different ranks. It appeared that the system was better trained with the side pose images. This is in agreement with the results illustrated in [52]. Thus, pose variations can be addressed especially when appropriate training samples are used.

**Table 3.** Accuracy of the Inception V3 network at different ranks (Ranks-1, 5, 10), against state-of-the-art Re-Identification (Re-ID) methods, per dataset, for the identification scenario.

<i>CUHK03 Dataset</i>			
Methodology	Rank-1 Rate	Rank-5 Rate	Rank-10 Rate
[22]	97.80%	99.60%	99.80%
<b>Inception V3</b>	92.50%	98.31%	99.65%
[11]	87.50%	97.70%	98.90%
[3]	83.60%	97.50%	99.20%
[12]	82.10%	96.20%	98.20%
[16]	81.80%	95.20%	97.20%
[17]	79.40%	N/A	N/A
[23]	78.90%	N/A	N/A
[34]	77.90%	N/A	N/A
[18]	74.90%	92.90%	96.70%
[20]	74.68%	95.99%	97.47%
[19]	74.21%	94.33%	97.54%
[15]	72.30%	N/A	N/A
[26]	63.70%	80.60%	86.90%
[53]	62.70%	79.90%	86.20%
[28]	62.55%	90.05%	94.80%
[24]	61.36%	N/A	N/A
[25]	58.00%	N/A	N/A
[31]	52.20%	N/A	N/A
[27]	38.10%	N/A	N/A
[8]	24.30%	N/A	N/A
<i>RAiD Dataset</i>			
Methodology	Rank-1 Rate	Rank-5 Rate	Rank-10 Rate
<b>Inception V3</b>	99.86%	100.0%	100.0%
[3]	95.00%	100.0%	100.0%
[33]	59.84%	N/A	N/A
<i>Market-1501 Dataset</i>			
Methodology	Rank-1 Rate	Rank-5 Rate	Rank-10 Rate
[30]	98.10%	99.30%	99.60%
[32]	96.20%	N/A	N/A
[23]	95.70%	98.40%	99.00%
[36]	95.40%	N/A	N/A
[34]	95.40%	N/A	N/A
[17]	95.30%	N/A	N/A
[35]	94.80%	N/A	N/A
[15]	94.80%	N/A	N/A
[22]	94.40%	N/A	N/A
[26]	93.80%	97.50%	98.50%
[53]	92.80%	97.30%	98.10%
[24]	92.13%	N/A	N/A
[25]	90.40%	N/A	N/A
[11]	88.90%	95.60%	N/A
[13]	87.04%	95.10%	96.42%
[21]	86.54%	95.16%	97.03%
[18]	83.31%	N/A	N/A
[16]	82.30%	92.30%	95.20%
<b>Inception V3</b>	82.11%	97.73%	99.60%
[19]	80.31%	N/A	N/A
[14]	79.51%	N/A	N/A
[3]	77.49%	91.81%	94.69%
[27]	77.11%	N/A	N/A
[49]	72.54%	N/A	N/A
[28]	61.02%	N/A	N/A
[8]	47.25%	N/A	N/A
<i>RGBD-ID Dataset</i>			
Methodology	Rank-1 Rate	Rank-5 Rate	Rank-10 Rate
[10]	98.40%	N/A	N/A
[38]	91.60%	N/A	N/A
[9]	85.20%	N/A	N/A
<b>Inception V3</b>	54.62%	59.46%	65.68%
[54]	30.10%	N/A	N/A
<i>Unified DATASET</i>			
Methodology	Rank-1 Rate	Rank-5 Rate	Rank-10 Rate
<b>Inception V3</b>	54.28%	59.16%	65.42%

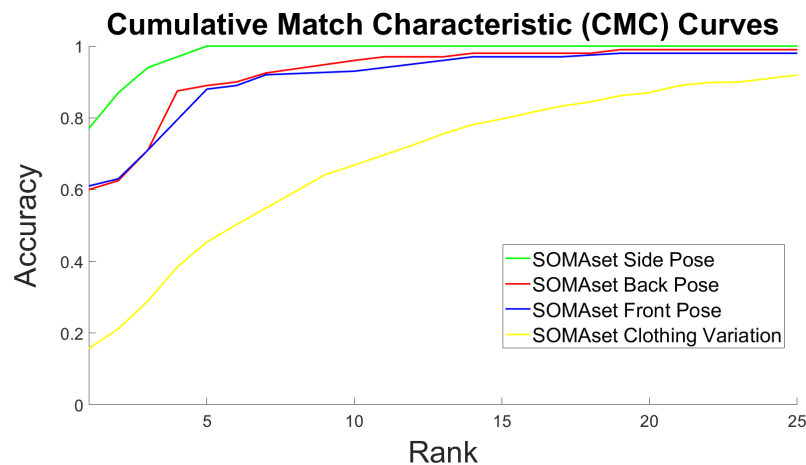
**Table 4.** SOMAset accuracy at different ranks under cloth variation as achieved by the Inception V3 network.

Rank-1	Rank-5	Rank-10
15.70%	45.43%	66.84%

**Table 5.** SOMAset accuracy at different ranks under pose variation as achieved by the Inception V3 network.

Training Pose	Rank-1	Rank-5	Rank-10
Frontal	61.00%	88.15%	93.28%
Back	60.00%	89.33%	96.02%
Side	77.00%	100.0%	100.0%

The corresponding CMC curves of the aforementioned clothing and pose variation effect experimentation are illustrated in Figure 10. The curves visually confirm the results of Tables 4 and 5. The best curve was achieved when training the network with the side pose (green color) images. The curves corresponding to training with the back (red color) and front (blue color) images followed, while the most difficult effect to be handled was the clothing variation with the corresponding (yellow color) curve being at the bottom.

**Figure 10.** SOMAset dataset CMC curves for different poses and clothing variation as achieved by the Inception V3 network.

#### 4.2.2. Verification Scenario

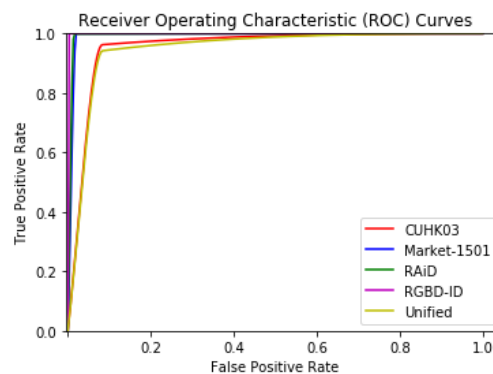
The verification module of the Siamese network, as described in Section 3, was strictly used experimenting on the identity verification scenario. The verification experiments were conducted using Python on a Windows 10 PC with an Intel i9 CPU at 2.30 GHz, 32.00 GB RAM and a NVIDIA Quadro T2000 GPU of 4 GB memory.

Table 6 illustrates the verification accuracy achieved on the evaluation datasets. Note, that this was not the same as identification rank- $n$  scores. Figure 11 illustrates the Receiver Operating Characteristic (ROC) curves for the proposed somatotype verification system for all datasets. The ROC curve is commonly used for evaluating verification experiments. The curves confirmed the results of Table 6. The best curve was achieved for the *RGBD-ID* dataset (magenta color) with the curves of *RAiD* (green color), *Market-1501* (blue color), *CUHK03* (red color) and *Unified* (yellow color) dataset following. The curves corresponding to *RAiD* and *Market-1501* datasets almost overlapped with the one of the *RAiD* being slightly better. The *Unified* was, as expected, the most challenging dataset. Table 7 presents the mean Average Precision ( $mAP$ ) of the proposed verification method compared to recent

top-ranked state-of-the-art Re-ID methods, for each dataset. The  $mAP$  was used in the evaluation of both identification and verification experiments; a comparison of the two proposed learning methods (Siamese vs. Inception V3) is thus illustrated in Table 8 via their  $mAP$  values per dataset.

**Table 6.** Verification accuracy per dataset as achieved by the Siamese network.

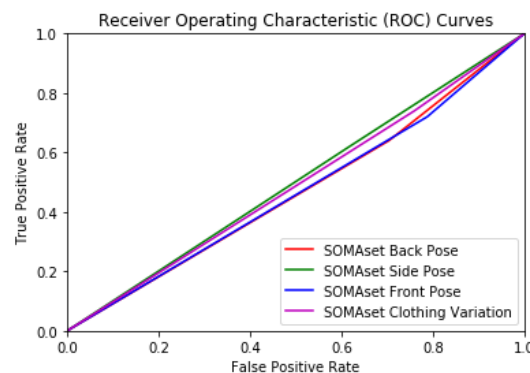
Dataset	Verification Accuracy
CUHK03	93.50%
RAiD	99.32%
Market-1501	99.10%
RGBD-ID	99.95%
Unified	92.27%



**Figure 11.** Receiver Operating Characteristic (ROC) curves per dataset as achieved by the Siamese network. Note that the green and blue color lines almost overlap.

#### Clothing and Pose Variation Effects

The verification scenario methodology was also tested under clothing and pose variations. Experiments were conducted on the *SOMAsSet* dataset. The experimental protocol was the same as in the identification scenario. In contrast to the identification methodology (Inception V3 network), it seemed that the Siamese network was less sensitive to cloth variations but slightly more sensitive to camera pose variations. The verification accuracy is presented in Table 9. The ROC curves are illustrated in Figure 12. Once again, side pose (green color) achieved better results than when front (blue color) and back poses (red color) were used in training. The clothing variation case (magenta color) seemed to be more stable in the Siamese rather than in the Inception V3 network.



**Figure 12.** *SOMAsSet* dataset ROC curves for different pose and clothing variations as achieved by the Siamese network.



**Table 7.** *mAP* results of the Siamese network compared to state-of-the-art, per dataset, for the verification scenario.

<i>CUHK03 Dataset</i>	
Methodology	<i>mAP</i>
<b>Siamese</b>	<b>89.72%</b>
[3]	86.79%
[16]	84.80%
[18]	78.80%
[23]	76.90%
[17]	76.70%
[34]	73.00%
[15]	67.80%
[53]	57.60%
[26]	57.50%
[25]	56.50%
[24]	55.78%
[27]	40.30%
[8]	22.70%
<i>RAiD Dataset</i>	
Methodology	<i>mAP</i>
<b>Siamese</b>	<b>99.18%</b>
[3]	95.00%
<i>Market-1501 Dataset</i>	
Methodology	<i>mAP</i>
<b>Siamese</b>	<b>97.87%</b>
[36]	94.20%
[34]	94.20%
[22]	90.70%
[32]	89.70%
[23]	88.20%
[30]	87.60%
[17]	86.70%
[35]	86.00%
[15]	84.90%
[26]	81.60%
[24]	79.01%
[53]	78.70%
[11]	76.70%
[25]	76.01%
[21]	68.97%
[13]	66.89%
[27]	63.63%
[16]	62.10%
[14]	59.87%
[18]	59.69%
[19]	57.53%
[3]	53.50%
[49]	40.54%
[28]	35.68%
[8]	19.47%
<i>RGBD-ID Dataset</i>	
Methodology	<i>mAP</i>
<b>Siamese</b>	<b>99.95%</b>
[10]	99.50%
[9]	98.23%
[54]	88.70%
<i>Unified Dataset</i>	
Methodology	<i>mAP</i>
<b>Siamese</b>	<b>86.30%</b>

**Table 8.** Proposed Siamese versus Inception V3 network results per somatotype dataset.

<b>CUHK03 Dataset</b>	
Methodology	<i>mAP</i>
Siamese Network	89.72%
<b>Inception V3 Network</b>	<b>90.00%</b>
<b>RAiD Dataset</b>	
Methodology	<i>mAP</i>
<b>Siamese Network</b>	<b>99.18%</b>
Inception V3 Network	96.53%
<b>Market-1501 Dataset</b>	
Methodology	<i>mAP</i>
<b>Siamese Network</b>	<b>97.87%</b>
Inception V3 Network	81.39%
<b>RGBD-ID Dataset</b>	
Methodology	<i>mAP</i>
<b>Siamese Network</b>	<b>99.95%</b>
Inception V3 Network	53.37%
<b>Unified Dataset</b>	
Methodology	<i>mAP</i>
<b>Siamese Network</b>	<b>86.30%</b>
Inception V3 Network	50.94%

**Table 9.** SOMAset verification accuracy under clothing and pose variations as achieved by the Siamese network.

Experiment	Verification Accuracy
Clothing variation	49.02%
Back pose	46.62%
Front Pose	46.80%
Side Pose	50.12%

#### 4.3. Implementation Details

The only undertaken pre-processing step for both identification and verification scenario is that the initial images were resized in order to feed the input of the corresponding network. For the identification experiments, the learning rate of the training process was set to 0.0003, the size of the mini batch to 10 and the training epochs to 12 for the four publicly available datasets and to 30 for the unified dataset. For the verification scenario, the learning rate of the training process was set to 0.00006, the size of the mini batch to 32 and the training epochs to 30 for the unified somatotype dataset. The values of the aforementioned hyper-parameters were chosen experimentally. After training, the testing process took place without an intervening validation phase.

As the confidence threshold  $\epsilon$  for both scenarios, the value 0.7 was chosen. It should be mentioned here that, the threshold was parameterised in the proposed scheme. Thus, any desired threshold could be chosen. During this research, many different thresholds were tested and the 0.7 threshold value was experimentally proved to achieve the best results.

#### 4.4. Timings

The time needed on a per-query basis for both scenarios was very low. More precisely, for identifying a query, the time required by the Inception V3 network was approximately 0.05 s, while for verifying an identity the required time by the Siamese network was approximately 0.03 s.

If the Inception V3 network was implemented for verification, the per-query time was not much different to that for the identification case, as the same process took place followed by an identity confirmation post-processing which was of  $O(1)$  time complexity. On the contrary, if the Siamese network was implemented for identification, the per-query time drastically increased. That is because it should be multiplied by the number of instances of each identity in the training set of the corresponding dataset. If we consider that there are  $m$  instances for each identity and  $n$  identities in total, then the per-query time required will be multiplied by a factor of  $O(n \times m)$ . Thus the bigger the dataset the more the time needed for the system to respond. In our case, this time varies from approximately 3 min for the *RAiD* dataset to approximately 10 min for the *Market – 1501* dataset.

### 5. Discussion

Identification and verification are biometric recognition problems with different evaluation metrics. In the case of identification, *CMC* curves and Rank- $n$  accuracy ( $n = 1, 5, 10$ ) are often used. Respectively, in the case of verification, *ROC* curves, *mAP* and *VerificationRate* are commonly used. The nearest comparable state-of-the-art methods that have implicitly or explicitly (only [3]) used the somatotype feature have focused on Re-ID, which is an identification problem. Thus, although the proposed identification method could be directly compared against these state-of-the-art Re-ID methods, this was not the case for the proposed verification method. To overcome this problem, *mAP* values were used, as *mAP* is common to both problems. The only drawback is that *mAP* is not made available as often as other measures in related publications.

The experimental results for the identification scenario were very promising for the majority of the cases. However, in more challenging datasets which include a change of attire (*RGBD-ID* and *Unified*), identification accuracy drops significantly. Thus, while this method may be suitable for constrained cases where individuals are not expected to change attire (e.g., intra-day), it does not appear suitable for identification across longer time periods, where a change of attire is expected. Another problem is that it necessitates time-consuming retraining for every addition to the gallery dataset, making it unsuitable for dynamic galleries of identities. Still, given the ease of capturing the somatotype (still images) it can be an extremely useful and cheap component under a biometric fusion scheme. At the recognition phase, no collaboration is required, since somatotype is based on simple still images of a person's body; the person may or may not have enrolled.

The results of the verification scenario indicate that the proposed method achieves very high verification rates for all datasets. Furthermore, the scheme surpasses state-of-the-art performance, including the Inception V3 network approach, for every single dataset used for experimentation. The increased verification rates in both *RGBD-ID* and *Unified* datasets indicate that, unlike the Inception V3 network, this scheme is much less sensitive to cloth variations. On the other hand, this scheme is slightly more sensitive to pose variations. That can be explained by the fact that the Siamese networks are trained on a comparison metric rather than the images themselves. Thus, if it is trained only in comparing over a specific pose, then comparing over a different pose will be an unknown for the system. An important advantage of the Siamese network is that it needs no re-training when utilised with different datasets as it follows the One-Shot-Learning approach. One can thus assume a dynamic gallery dataset without the time cost of re-training. Due to the nature of the Siamese networks, even the initial training set doesn't have to be as extensive as for typical deep networks. However, when aimed at the identification scenario, the Siamese network is more time-consuming than the Inception V3 network, as its performance is linear in the size of the gallery rather than constant.

The effect of clothing is a major issue that has been highlighted in this paper. Ideally, a person recognition system should be invariant to clothing, as clothing is not a characteristic of a person's

identity. However, since the input modality is images, such a recognition system cannot be completely invariant to clothing or for that reason other image artifacts. Just as a facial recognition system cannot be completely invariant to occlusions. However, here comes the advantage of being able to train on a relatively large dataset with exact ground truth; the machine learning system learns the underlying somatotype because it is presented with the same somatotype wearing different clothes as well as different somatotypes wearing the same clothes. Still, there is the question of what the network actually learned. It was thus deemed useful to investigate whether the network, after training, is better at distinguishing somatotypes or clothes. Note that the latter would seem like an easier task, as it is based on the outer appearance that is obvious in full body images. The mean value and standard deviation of the similarity score for the same subjects wearing different clothes as well as different subjects wearing the same clothes are shown in Table 10 for a balanced set of five different cloth variations and five different somatotype variations. The number of total cloth variations that are common for males and females is five, and that is why this number is chosen. Of course, this number is smaller than the number of somatotype variations in the original dataset. This led to the use of 6250 somatotype images, which is less than the 10% of the full dataset. The mean values corresponding to subjects of the same somatotype are slightly higher than the ones corresponding to subjects wearing the same clothes, while the standard deviations corresponding to subjects of the same somatotype are negligible, in contrast to the ones corresponding to subjects wearing same clothes.

**Table 10.** Mean values and standard deviations of the similarity scores, produced by the Inception V3 and the Siamese networks among subjects belonging to the same somatotype and subjects wearing same clothes.

Comparison	Network	Similarity Scores	
		Mean Value	Standard Deviation
Same somatotype/Different Clothes	Inception V3	0.9897912	$3.7876 \times 10^{-6}$
Same somatotype/Different Clothes	Inception V3	0.9087757	$1.0225 \times 10^{-2}$
Same somatotype/Different Clothes	Inception V3	0.9999905	$1.0120 \times 10^{-7}$
Same somatotype/Different Clothes	Inception V3	0.9889345	$1.5363 \times 10^{-3}$

An exhaustive study regarding the influence of the different movements of the structural aspects of the human body on the derived accuracy is not applicable for the time being. Due to the plethora of such movements this would require many resources and would be very time-consuming. Moreover, the very nature of the datasets themselves make the aforementioned study unrealistic. The datasets contain only still images which cannot capture the entire movement spectrum of the structural aspects of the human body. In that sense, the assessment of the proposed deep networks under three different pose variations can be thought of as an assessment under a small subset of different movements of the structural aspects of the human body. As mentioned in Section 6, the recruitment of videos (image sequences) instead of the still images, adding extended temporal information as well, can offer a better basis for such a study, but the problems of resource allocation and time complexity will remain, possibly restricting the study just to a bigger movement subset than the one analysed here.

Summing up the discussion, Table 11 illustrates the advantages and disadvantages of the two proposed schemes.

**Table 11.** Advantages and disadvantages of the two proposed schemes.

Scheme	Advantages	Disadvantages
Inception V3 Network	- High identification ranks for most cases - Fast identification	- Cloth variation sensitive - Re-training required for use with different datasets
Siamese Network	- High verification accuracy - Fast verification - Re-training is not required - Less sensitive to cloth variation - Small training set required	- Slow identification

## 6. Future Challenges

One promising way of extracting more accurate somatotype features and thus potentially improving accuracy when there exist variations in attire, is the use of video sequences instead of still images. Thus, the entire movement spectrum of the structural aspects of the human body could be encoded by the deep networks. Preliminary works in this direction could be used as a basis [55,56]. Moreover, synthetic somatotype datasets (e.g., *SyRI* [57], *SOMAsSet* [3] and *PersonX* [52]) are extremely useful in training, especially for creating controlled subject variations in e.g., attire and pose. In the future we expect such datasets to be of increasing importance in transfer learning scenarios, combined with emerging more sophisticated deep networks such as Inception V4 [58] and ResNet [59].

## 7. Conclusions

The human somatotype, as can be captured from simple full-body images, is an extremely attractive biometric trait as it can be used on-the-move and at a distance, thus being suitable even for non-collaborative subjects. In addition it requires no specialised capture equipment. This paper studied the possibility of using the somatotype as a biometric recognition trait and proposed two respective learning networks for both identification and verification scenarios. Evaluation has been conducted on well-established publicly available datasets for Re-ID, as well as on a new unified dataset that was created by merging the aforementioned ones and contains the largest number of identities.

The proposed Inception V3 network is suitable for non-collaborative and on-the-move identification and/or verification using a pre-constructed, static, gallery dataset of identities. However, application of the scheme on a dataset of identities that can be dynamically modified is problematic, because the Inception V3 network should undergo a time-consuming re-training process for every change in the gallery. Moreover, this scheme is sensitive to cloth variations. On the other hand, the proposed Siamese network scheme is suitable for non-collaborative identity verification on-the-move. It can also be utilised for the identification process at the cost of a linear time penalty. The implementation of One-Shot-Learning ensures that system training is performed only once and offline, making it suitable for a dynamically varying gallery.

Under constrained conditions, where individuals are not expected to change attire across captures and where recognition failures are not catastrophic (e.g., access to non-critical events), the somatotype feature appears extremely attractive due to the ease and low-cost associated with its capture. However, under conditions of large time ranges, where individuals are expected to change attire across captures, the somatotype trait may not be adequate as a stand-alone biometric, particularly if high recognition accuracy is required. Still, under such more general conditions, the somatotype is useful under a biometric fusion scheme, where it can boost a biometric system's performance at a very low cost.

**Author Contributions:** Conceptualization, A.D. and T.T.; methodology, A.D. and T.T.; software, A.D.; validation, A.D. and T.T.; formal analysis, T.T.; investigation, A.D.; resources, A.D.; data curation, A.D.; writing—original draft preparation, A.D.; writing—review and editing, T.T.; visualization, A.D.; supervision, T.T.; project administration, T.T.; funding acquisition, T.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 833704. 

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Theoharis, T.; Passalis, G.; Toderici, G.; Kakadiaris, I.A. Unified 3D Face and Ear Recognition Using Wavelets on Geometry Images. *Pattern Recogn.* **2008**, *41*, 796–804. [[CrossRef](#)]
2. Sheldon, W.H.; Stevens, S.S.; Tucker, W.B. *The Varieties of Human Physique*; Harper: New York, NY, USA, 1940.
3. Barbosa, I.B.; Cristani, M.; Caputo, B.; Rognhaugen, A.; Theoharis, T. Looking beyond appearances: Synthetic training data for deep CNNs in re-identification. *Comput. Vis. Image Underst.* **2018**, *167*, 50–62. [[CrossRef](#)]

4. Castrillón-Santana, M.; Lorenzo-Navarro, J. Chapter 7—Soft Biometric Attributes in the Wild: Case Study on Gender Classification. In *Human Recognition in Unconstrained Environments*; Marsico, M.D., Nappi, M., Proença, H., Eds.; Academic Press: Cambridge, MA, USA, 2017; pp. 145–176. [\[CrossRef\]](#)
5. Jain, A.K.; Flynn, P.; Ross, A.A. *Handbook of Biometrics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2007.
6. Li, W.; Zhao, R.; Xiao, T.; Wang, X. Deep ReID: Deep Filter Pairing Neural Network for Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 152–159.
7. Das, A.; Chakraborty, A.; Roy-Chowdhury, A.K. Consistent Re-identification in a Camera Network. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Volume 8690, pp. 330–345.
8. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
9. Munaro, M.; Ghidoni, S.; Dizmen, D.T.; Menegatti, E. A Feature-Based approach to People Re-Identification Using Skeleton Keypoints. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 5644–5651.
10. Munaro, M.; Basso, A.; Fossati, A.; Van Gool, L.; Menegatti, E. 3D reconstruction of freely moving persons for re-identification with a depth sensor. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 4512–4519.
11. Xiao, Q.; Luo, H.; Zhang, C. Margin sample mining loss: A deep learning based method for person re-identification. *arXiv Preprint* **2017**, arXiv:1710.00478.
12. Jin, H.; Wang, X.; Liao, S.; Li, S.Z. Deep person re-identification with improved embedding and efficient training. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 261–267.
13. Lin, Y.; Zheng, L.; Zheng, Z.; Wu, Y.; Hu, Z.; Yan, C.; Yang, Y. Improving person re-identification by attribute and identity learning. *Pattern Recognit.* **2019**. [\[CrossRef\]](#)
14. Zheng, Z.; Zheng, L.; Yang, Y. A discriminatively learned CNN embedding for person reidentification. *ACM Trans. Multimed. Comput. Commun. Appl.* **2018**, *14*, 13. [\[CrossRef\]](#)
15. Zhou, K.; Yang, Y.; Cavallaro, A.; Xiang, T. Omni-Scale Feature Learning for Person Re-Identification. *arXiv Preprint* **2019**, arXiv:1905.00953.
16. Sun, Y.; Zheng, L.; Deng, W.; Wang, S. SVDNet for Pedestrian Retrieval. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
17. Dai, Z.; Chen, M.; Gu, X.; Zhu, S.; Tan, P. Batch DropBlock Network for Person Re-identification and Beyond. *arXiv Preprint* **2018**, arXiv:1811.07130.
18. Liu, S.; Hao, X.; Zhang, Z. Pedestrian retrieval via part-based gradation regularization in sensor networks. *IEEE Access* **2018**, *6*, 38171–38178. [\[CrossRef\]](#)
19. Li, D.; Chen, X.; Zhang, Z.; Huang, K. Learning deep context-aware features over body and latent parts for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 384–393.
20. Chen, W.; Chen, X.; Zhang, J.; Huang, K. A Multi-Task Deep Network for Person Re-Identification. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–10 February 2017.
21. Han, K.; Guo, J.; Zhang, C.; Zhu, M. Attribute-Aware Attention Model for Fine-Grained Representation Learning. In Proceedings of the 2018 ACM Multimedia Conference on Multimedia Conference, Seoul, Korea, 22–26 October 2018; pp. 2040–2048.
22. Zhang, X.; Luo, H.; Fan, X.; Xiang, W.; Sun, Y.; Xiao, Q.; Jiang, W.; Zhang, C.; Sun, J. Alignedreid: Surpassing human-level performance in person re-identification. *arXiv Preprint* **2017**, arXiv:1711.08184.
23. Zheng, F.; Deng, C.; Sun, X.; Jiang, X.; Guo, X.; Yu, Z.; Huang, F.; Ji, R. Pyramidal Person Re-Identification via Multi-Loss Dynamic Training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 8514–8522.
24. Zhang, Z.; Si, T.; Liu, S. Integration convolutional neural network for person re-identification in camera networks. *IEEE Access* **2018**, *6*, 36887–36896. [\[CrossRef\]](#)



25. Jiao, S.; Wang, J.; Hu, G.; Pan, Z.; Du, L.; Zhang, J. Joint attention mechanism for person re-identification. *IEEE Access* **2019**, *7*, 90497–90506. [[CrossRef](#)]
26. Sun, Y.; Zheng, L.; Li, Y.; Yang, Y.; Tian, Q.; Wang, S. Learning Part-based Convolutional Features for Person Re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**. [[CrossRef](#)] [[PubMed](#)]
27. Zhong, Z.; Zheng, L.; Cao, D.; Li, S. Re-Ranking Person Re-Identification with k-Reciprocal Encoding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 22–25 July 2017; pp. 1318–1327.
28. Zhang, L.; Xiang, T.; Gong, S. Learning a Discriminative Null Space for Person Re-Identification. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1239–1248.
29. Chen, L.F.; Liao, H.Y.M.; Ko, M.T.; Lin, J.C.; Yu, G.J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognit.* **2000**, *33*, 1713–1726. [[CrossRef](#)]
30. Wang, G.; Lai, J.; Huang, P.; Xie, X. Spatial-Temporal Person Re-Identification. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 5 September 2019; Volume 33, pp. 8933–8940.
31. Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person Re-Identification By Local Maximal Occurrence Representation and Metric Learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7 June 2015; pp. 2197–2206.
32. Lawen, H.; Ben-Cohen, A.; Protter, M.; Friedman, I.; Zelnik-Manor, L. Attention Network Robustification for Person ReID. *arXiv Preprint* **2019**, arXiv:1910.07038.
33. Panda, R.; Bhuiyan, A.; Murino, V.; Roy-Chowdhury, A.K. Adaptation of person re-identification models for on-boarding new camera(s). *Pattern Recognit.* **2019**, *96*, 106991. [[CrossRef](#)]
34. Quan, R.; Dong, X.; Wu, Y.; Zhu, L.; Yang, Y. Auto-ReID: Searching for a Part-aware ConvNet for Person Re-Identification. *arXiv Preprint* **2019**, arXiv:1903.09776.
35. Zheng, Z.; Yang, X.; Yu, Z.; Zheng, L.; Yang, Y.; Kautz, J. Joint Discriminative And Generative Learning For Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 2138–2147.
36. Luo, H.; Gu, Y.; Liao, X.; Lai, S.; Jiang, W. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Trento, Italy, 9–13 September 2019; pp. 1–9.
37. Gohar, I.; Riaz, Q.; Shahzad, M.; Ul Hasnain Hashmi, M.Z.; Tahir, H.; Ehsan Ul Haq, M. Person re-identification using deep modeling of temporally correlated inertial motion patterns. *Sensors* **2020**, *20*, 949. [[CrossRef](#)]
38. Liu, H.; Hu, L.; Ma, L. Online RGB-D person re-identification based on metric model update. *CAAI Trans. Intell. Technol.* **2017**, *2*, 48–55. [[CrossRef](#)]
39. Gonzalez-Sosa, E.; Vera-Rodriguez, R.; Fierrez, J.; Patel, V.M. Exploring body shape from mmw images for person recognition. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2078–2089. [[CrossRef](#)]
40. Gonzalez-Sosa, E.; Vera-Rodriguez, R.; Hernandez-Ortega, J.; Fierrez, J. Person Recognition at a Distance: Improving Face Recognition Through Body Static Information. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 3439–3444.
41. Roedelein, J.E. Sheldon’s Type Theory. In *Dictionary of Theories, Laws, and Concepts in Psychology*; Greenwood: Westport, CT, USA, 1998; pp. 427–428.
42. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
43. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
44. Bromley, J.; Guyon, I.; LeCun, Y.; Säckinger, E.; Shah, R. Signature Verification Using a “Siamese” time delay neural network. In Proceedings of the Advances in Neural Information Processing Systems, Denver, CO, USA, 28 November–1 December 1994; pp. 737–744.
45. Yuan, X.; Liu, Q.; Long, J.; Hu, L.; Wang, Y. Deep Image Similarity Measurement Based on the Improved Triplet Network with Spatial Pyramid Pooling. *Information* **2019**, *10*, 129. [[CrossRef](#)]

46. Taigman, Y.; Yang, M.; Ranzato, M.; Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition, Columbus, OH, USA, 24–27 June 2014; pp. 1701–1708.
47. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7 June 2015; pp. 815–823.
48. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese Neural Networks for One-Shot Image Recognition. In Proceedings of the ICML Deep Learning Workshop, Lille, France, 6–11 July 2015; Volume 2.
49. Zheng, L.; Yang, Y.; Hauptmann, A.G. Person re-identification: Past, present and future. *arXiv Preprint* **2016**, arXiv:1610.02984.
50. Bedagkar-Gala, A.; Shah, S.K. A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **2014**, *32*, 270–286. [[CrossRef](#)]
51. Vielhauer, C.; Dittmann, J.; Drygajlo, A.; Juul, N.C.; Fairhurst, M. *Biometrics and ID Management: COST 2101 European Workshop, BioID 2011, Brandenburg (Havel), March 8–10, 2011, Proceedings*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011; Volume 6583.
52. Sun, X.; Zheng, L. Dissecting Person Re-Identification from the Viewpoint of Viewpoint. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 608–617.
53. Chen, P.; Jia, T.; Wu, P.; Wu, J.; Chen, D. Learning Deep Representations by Mutual Information for Person Re-identification. *arXiv Preprint* **2019**, arXiv:1908.05860.
54. Haque, A.; Alahi, A.; Fei-Fei, L. Recurrent Attention Models for Depth-Based Person Identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016.
55. Wu, L.; Wang, Y.; Gao, J.; Li, X. Where-and-when to look: Deep siamese attention networks for video-based person re-identification. *IEEE Trans. Multimed.* **2018**, *21*, 1412–1424. [[CrossRef](#)]
56. Collins, R.T.; Gross, R.; Shi, J. Silhouette-Based Human Identification from Body Shape and Gait. In Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition, Washington, DC, USA, 21–21 May 2002; pp. 366–371.
57. Bak, S.; Carr, P.; Lalonde, J. Domain Adaptation through Synthesis for Unsupervised Person Re-identification. *arXiv* **2018**, arXiv:1804.10094.
58. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-Resnet and the Impact of Residual Connections on Learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–10 February 2017.
59. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

