Doctoral thesis

Doctoral theses at NTNU, 2020:262

Igor Barros Barbosa

Unconventional biometrics

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical
Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Igor Barros Barbosa

Unconventional biometrics

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2020

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-4872-6 (printed ver.)
ISBN 978-82-326-4873-3 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:262

Printed by NTNU Grafisk senter

SUPERVISORS

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Professor Theoharis Theoharis, PhD
Department of Computer and Information Science

Professor Agnar Aamodt, PhD
Department of Computer and Information Science

SINTEF

Christian Schellewald, PhD
Department of Seafood Technology

Dedicated to my family.

ABSTRACT

This thesis is a paper collection that focuses on unconventional methods of biometric recognition. Four new approaches are presented and discussed. The first two introduce and explore the concepts behind transient biometrics. Transient biometrics relaxes the hard permanence requirement that is common to biometric identifiers, creating a biometric signature with expiration date which increases acceptability. The third approach investigates a novel method for extracting a capable biometric identifier using Electroencephalography (EEG) and a visual stimulus. The final approach studies the use of synthetic biometric data for training a machine learning approach in the recognition of non-collaborative subjects under the context or person re-identification. Four new datasets have been created for the purposes of this thesis and have been made publicly available. Contributions are on the interface between computer vision, biometrics and machine learning. Ethical implications of this work are discussed, concluding that it is preferable to perform such work in the public domain.

ACKNOWLEDGMENTS

The fulfilment of this thesis involves a large group of people. I want to start by thanking Professor Theoharis Theoharis for the opportunity to work for a doctorate. The chance to come to Norway and do research at NTNU has changed my life in more ways than I could ever imagine. My sincere thanks to my co-supervisors, Professor Agnar Aamodt and Dr Christian Schellewald, for their availability and guidance whenever necessary. Sincere thanks to the coordinators, secretaries and other employees who are behind NTNU's computer science department.

This thesis represents the end of a personal journey of many years. During those years, I was fortunate to meet Amanda, which changed my perception of the world; and that I now have the pleasure of calling wife. Thank you for continually believing in me and for being by my side. I couldn't have gone through this journey without you.

However, this journey was not without sacrifices. The sacrifices behind this journey began two generations ago; where my grandparents decided that the best investment they could make for the family was to invest everything in education. Work and sacrifices are showing results. Two generations ago, we had a family that didn't have a high school diploma. In the current family generation, we have the first member of the family working for a doctorate. I acknowledge all the effort and work, for which I am forever grateful.

None of this would have been possible without my sister Bárbara . Thank you for always taking care of me and for being an unmatched support system. I hope you are proud of me as I am proud of you.

Most importantly, I would like to express my most sincere gratitude to my parents, Juarez and Corália Barbosa: their efforts and sacrifices are unmatched. Few people know that my sister and lived our youth alone and a country apart; everything so that we could have access to better education. Thank you from the bottom of my heart for all the support and love. You both have my sincere gratitude for getting involved and endeavouring with me on this journey as I pursue my dreams.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ANN	Artificial neural network
ASM	Active Shape Model
BRISK	Binary Robust Invariant Scalable Keypoints
CMC	Cumulative Match Curve
CNN	Convolutional neural network
EEG	Electro-Encephalo-Gram
FN	False Negative
FP	False Positive
fNIRS	Functional near-infrared spectroscopy
FPR	False Positive Rate
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GFTT	Good Features To Track
GPU	Graphics processing unit
LBP	Local Binary Patterns

MLP	Multilayer perceptron
PCA	Principal Component Analysis
RANSAC	Random sample consensus
RE-ID	Re-identification
ROC	Receiver operating characteristic
ROI	Region of Interest
SIFT	Scale-Invariant Feature Transform
TBND	Transient Biometrics Nails Dataset
TN	True Negative
TP	True Positive
TPR	True Positive Rate
VEP	Visual Evoked Potentials

Part I

RESEARCH OVERVIEW

1

INTRODUCTION

Biometrics is a science that seeks to understand, explore, and learn how to use the physical, chemical, and behavioral characteristics of a given subject as a reliable proxy for the subject's identity [24]. Modern biometric authentication methods use the aforementioned characteristics, also known as a biometric trait, to produce biometric signatures. A biometric signature is a reliable and robust numeric representation for the biometric trait that is represented by a point in a hyperspace. The comparison of a representative hyperspace observation (a biometric signature) with other previous observations serves as a proxy for identity association. Therefore, one can use a distance computation between biometric signatures to perform biometric recognition. In a few solutions, a binary classifier can replace the distance computation function. A biometric system or solution refers to a complete pipeline, consisting of acquiring and processing the data of a biometric trait and using the generated numeric representation towards identity association. Finally, it is normal to use the term biometrics as an analog for the task of biometric recognition (more details in Section 1.2.1).

An ideal biometric trait conforms to seven established principles [24]:

UNIVERSALITY.

The trait needs to be available on all users of the biometric system. A biometric characteristic that can only be measured for a small fraction of the subjects will have a reduced power of discrimination, making it a less appealing biometric trait.

UNIQUENESS.

The uniqueness of the biometric traits is directly related to the potential use of such characteristics to discriminates between individuals. Hence, to be useful for a biometric solution, a biometric trait should be unique for every person, making it possible to distinguish between identities. A biometric feature that is identical between all subjects has little to no application on a biometric solution.

PERMANENCE.

As per [24], "the biometric trait of an individual should be sufficiently invariant over a period of time with respect to the matching algorithm. A trait that changes significantly over time is not a useful biometric". With the advent of soft biometrics and transient biometrics, this principle may need to be revised. Some applications do not require long-term biometric recognition and can be based on soft/transient biometric traits. Permanence thus needs to be a function of the expected lifespan of a biometric signature. For example, a biometric signature based on DNA will have a more robust/longer-lived permanence than a biometric signature based on facial traits. A biometric signature with short relative permanence (e. g. finger-nails) can still have a lifespan that is worthwhile for short-term authentication tasks (e. g. hotel room authentication) or be extra-useful in scenarios with recurrent interactions.

MEASURABILITY.

The ability to acquire and measure the data of a given biometric trait is fundamental towards the creation of the biometric signature. A biometric characteristic that is hard to digitize has fewer chances of being used in modern biometric solutions. Moreover, the collected data also requires to be fit for the creation of a biometric signature. Two instances of a biometric trait with poor measurability could be a biometric trait that can not be digitized, and a biometric characteristic that, once quantized, requires petabytes of data for the representation of a single observation. Therefore, measurability embodies two distinct points: the ability to acquire a digital representation of the biometric trait and a data representation that can be concisely recorded numerically (the biometric signature).

ACCEPTABILITY.

The users of biometric solutions need to be comfortable in providing the systems with the biometric data. Three factors can influence acceptability: resistance for a specific biometric characteristic to be measured (e. g. biometrics based on genital images are likely to have low acceptability); the user's trust on the computer system performing biometric solutions (e. g. subject might be reluctant to give their fingerprint to hotels since they do not trust the protocols of security, data handling, and encryption); finally, a subject might consider specific biometric traits to be too private for the intended use: e. g. one might expect never to share their DNA records for a biometric authentication system while having no issue in sharing their fingerprints.

CIRCUMVENTION.

Biometric recognition is usually in place to control and confirm that only selected subjects have access to a critical space or object or registry. Biometric recognition solutions inherit from their inception some security implications. A biometric system needs to be resilient to adversarial attacks and spoofing. Moreover, the biometric trait and system need to take into consideration the ability of an ill-intending actor to forge such a trait (e. g. fake fingers) or even to adversarially exploit the system (e. g. using high-resolution photos and videos to overcome face recognition systems). An ideal, and probably utopic, biometric trait should be impossible to circumvent.

PERFORMANCE.

Last but not least, there is the performance of the biometric trait. The trait should have sufficient recognition accuracy for the task as at hand. Performance can also relate to the computational cost of executing the biometric solution using the chosen biometric trait; it needs to be aware of system resource limitations.

Note that *universality* and *uniqueness* are actually requirements for a biometric system, while *permanence*, *acceptability*, *measurability*, *circumvention* and *performance* are not binary requirements but can be satisfied to varying degrees. The seven principles behind biometric traits illustrate how involved it is to associate a subject's identity with the physical or behavioral attributes of the said subject. Nevertheless, biometrics presents an attractive proposition to identity recognition at scale. Biometrics allows agents to perform identity association at a large scale without having to rely on surrogate identifications methods, which would include what a subject knows e. g. username and passwords or what a subject carries e. g. access card or passport)[25]. The disadvantage of surrogate identification methods is that the identification is not being made on a trait of the subject itself, therefore such methods are more vulnerable to theft, copying, and even forging.

While biometric recognition is obviously advantageous with respect to surrogate identification methods, the uptake of biometric recognition methods has been slow and fragmented. One can argue that this is due to the degree to which established biometric traits satisfy *permanence*, *acceptability*, *measurability*, *circumvention* and *performance*. Specifically, one needs more control of the degree of *permanence*, as less permanence can actually increase *acceptability* while greater permanence can be exploited in some critical applications (e. g. medical records). Major concerns with current biometric traits center at *circumvention* and *performance* both of which can be improved by the use of **multimodal biometrics**[†] . One is thus motivated to investigate novel biometric traits which have the potential to improve the aforementioned variables. In addition, such novel traits can also potentially improve *measurability* depending on their acquisition method. For example, the somatotype biometric trait presented in this thesis (publication D [3]) has high measurability as it is based on a simple image of a (collaborative or not) subject; the Electro-Encephalo-Gram (EEG) biometric trait presented in this thesis (publications C [6]) should be extremely hard to circumvent; the fingernail biometric trait presented in this thesis (publications A and B [5, 7]) trades permanence for increased acceptability. Finally, all these novel biometric traits combined in a multimodal system, have the potential to boost performance.

[†] The future of biometric recognition at large scales will depend on multimodal biometrics.

Thus, the main motivation behind this Ph.D. thesis was to attempt to answer the question: Are there novel and/or unconventional biometric traits that have been underexplored by the research community? If so, what are the strengths and weaknesses of these traits, and do they enable new, potentially transformative, biometric solutions? This thesis presents the work done to address these overarching questions. It considers the pipeline from the stage of selection of new biometric traits, the acquisition of the corresponding biometric data, the extraction of a biometric signature, the design of a distance function/algorithm and, of course, performance assessment. Three novel/unconventional biometric traits were explored (Electro-Encephalo-Gram (EEG), fingernails and somatotype) corresponding to four different biometric solutions (two for fingernails).

1.1 RESEARCH GOALS

The overarching theme of this thesis was the study and development of unconventional biometrics that could potentially bring to the research field of biometrics new biometric solutions based on underexplored or novel biometrics traits as well as the investigation of their weakness and strengths. Four research goals were set.

1.1.1 *Novel Biometric Identifiers*

A *biometric identifier* is a set of data derived from biological traits that can be employed to generate capable biometric signatures. Different from a biometric trait, a biometric identifier may derive from combining pieces of information from multiple sets of traits (e. g. somatotype trait is explored in Publication D [3]). A new biometric identifier can most likely be the outcome of one of two methods. The first is to explore one or more well established biometric traits (e. g. fingerprints) and find a better biometric signature, with respect to the state of the art; such a signature would typically contain more distintcive information, increasing its *uniqueness*. The second, more restricted, method involves the derivation of a biometric signature from novel biometric traits, that were not commonly explored by the biometrics community. The idea is that the study of under-explored biometric traits is more likely to bring new data to the field and can potentially establish a biometric signature that is better in one or more of the dimensions dictated by the 7 principles of a biometric trait described above. A novel biometric identifier can also be beneficial in multimodal biometrics. In the sequel we refer to this research goal as RG1.

1.1.2 *Unconventional methods of biometric recognition*

Often a new, unconventional, approach in biometrics involves an entirely different view, making biometric recognition more feasible. A typical example is transient biometrics (publications A and B [5, 7]) where instead of striving for the most permanent biometric traits, one relaxes this requirement, selecting more transient traits, in order to increase acceptability. We refer to applying unconventional methodologies in the field of biometrics as RG2.

1.1.3 *Application of machine learning methods in the context of biometrics*

The third goal of this thesis was to explore if machine learning methods can make a significant difference in biometric tasks, either directly or indirectly. By indirectly we mean for example, using machine learning methods to help segment images of datasets in a task that would otherwise be too laborious (see for example fingernail segmentation in Publication B [7]). Directly means the application of machine learning methods for biometric signature extraction or comparison. It must be pointed out that the use of machine learning was not contrived; the intention was never to force a machine learning solution just for the sake of it. Instead, it was applied when there were no good algorithmic solutions while at the same time, labelled data was available or could be constructed. For example the constraints of the RE-ID problem presented in [3] motivated the use of machine learning, as it was possible to create good labelled synthetic data and then use fine tuning methods for transferring that knowledge to real data. In the sequel we refer to this research goal as RG3.

1.1.4 *Creation of novel datasets*

Much of this thesis explores new or under-explored biometric traits and the required physical data acquisition was laborious and slow. To reduce such strain for future researchers and offer the community the means to better reproduce and improve upon the research presented, a goal was to publish newly created datasets. This research goal is referred to as RG4.

1.2 BIOMETRIC CONCEPTS

This section introduces some classic concepts on Biometrics as well as established performance measurement techniques.

1.2.1 *Biometric recognition*

Biometric recognition aims to associate a subject's identity with the physical or behavioral characteristics of the afore-mentioned subject [24]. These characteristics involve measurements of biological data and their derivatives, hence the term *biometrics*. It is also common to use biometrics as a synonym for the task of biometric recognition.

The task of biometric recognition usually takes two forms: *verification* and *identification* which are respectively presented in Section 1.2.1.3 and Section 1.2.1.2.

1.2.1.1 *Gallery, Probe and out-of-samples (data)sets*

It is typical for the biometric literature [18, 24] to use the terms of *gallery* and *probe* set when discussing biometric protocols. A gallery set is created to enroll biometric data of subjects who consent to be recognized; the gallery represents a reference for recognition tasks. The gallery is usually a database where data from collaborative individuals are enrolled. The probe set typically refers to new samples of biometric data; this is data acquired during recognition (post enrollment), which needs to be matched to the gallery set for the purpose of recognition.

Re-identification often also employs the terms gallery and probe; however its gallery set is normally acquired without collaboration from subjects. This is discussed in more detail in Section 1.2.2

It is also common for current biometrics benchmarks to have multiple disjoint versions of the gallery and probe sets. This enables data-driven methods (e. g. machine learning, pattern recognition methods) that require disjoint training and testing data partitions. Complex protocols are often included in benchmarking datasets [12, 31, 62].

Figure 1.1 shows the data-flow of enrollment, identification, verification and re-identification which are further in the respective sections 1.2.1.2, 1.2.1.3 and 1.2.2.

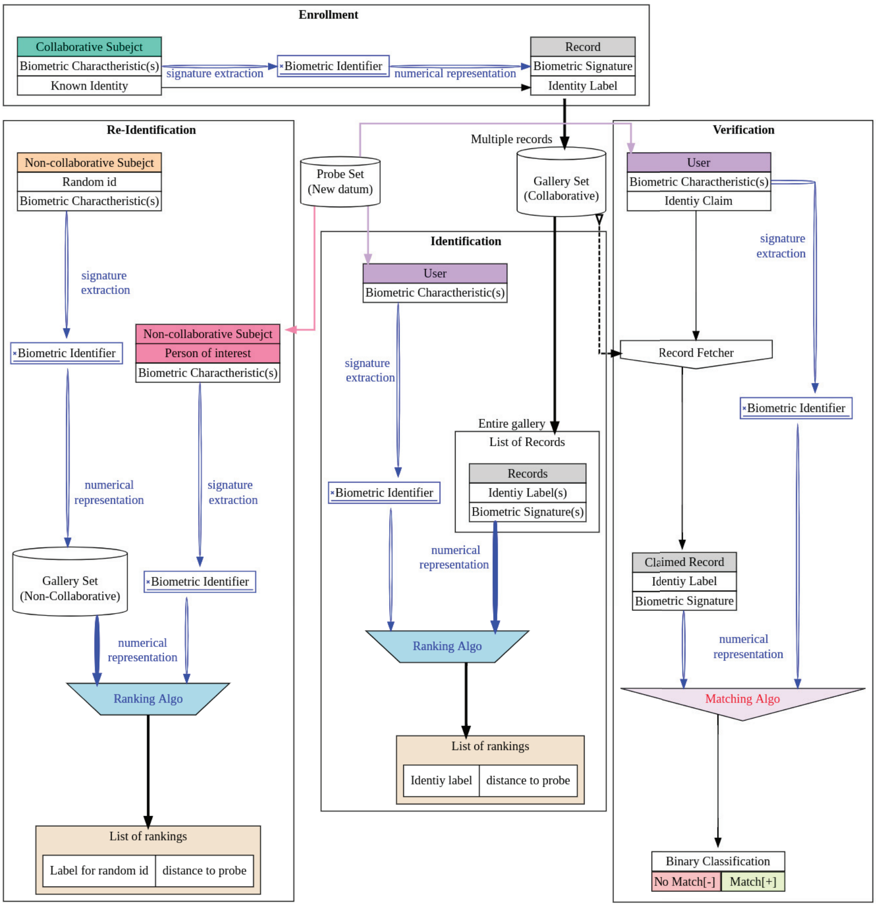


Figure 1.1: Data-flow of biometrics solutions. The figure shows the operational principles behind Enrollment, Verification, Identification and Re-identification. Biometric solutions normally start with enrollment, which is a controlled data collection procedure. Enrollment assesses the quality of the data and assures only good samples are enrolled in the gallery set. After the enrollment process, one can decide to perform either identification or verification. Identification is the process of identify association with no identity claim, whereas in verification there is an identity claim. There is no enrollment stage preceding re-identification; in principle, RE-ID is a self-contained modality, where one records data from non-collaborative individuals to establish a non-quality assured gallery set. Newly observed data is referred to as data from a probe set. In the case of verification and identification, the probe data comes from subjects that want to use the systems (collaborative). In the case of RE-ID newly observed datum (probe data) is from non-collaborative subjects.

1.2.1.2 *Identification*

In identification, a new biometric datum is created by a biometric sensor and presented as a query; in a benchmark the set of these queries forms the probe set. The biometric data of all enrolled subjects (gallery set) are compared against the query and often ranked in order of similarity based on a distance function. Identification can be considered as a one-to-many distance evaluation.

Thus the performance of identification is often measured using a curve known as CMC (see Section 1.2.3.2). The identification task can be used as a replacement for biometric verification, where a given subject does not need to claim a specific identity.

1.2.1.3 *Verification*

In verification, a new biometric datum is created by a biometric sensor and presented as an identity claim. The task verifies if the new datum matches the biometric data of the claimed identity in the gallery set. This is often implemented by a threshold decision on the output of a distance function. We thus have a one-to-one distance evaluation. Verification can be considered as a binary classification problem where the goal is to confirm or deny that new biometric datum matched that of the claimed identity. As a classification task, it is normally assessed using the ROC curve (see Section 1.2.3.1).

1.2.2 *Re-identification*

An essential task of the modern video-surveillance system is the association of observed subjects across multiple views. This is the motivation behind the problem of person RE-ID. **Person re-identification**† consists of recognizing instances of an individual in different locations over a network of cameras with overlapping and non-overlapping views and with *potentially* ample spatio-temporal intervals between observations [18].

†*Definition of the re-identification task*

As RE-ID is a surveillance provoked task, it must deal with crowds in what is usually an uncontrolled environment. Therefore, different from more conventional biometric tasks, RE-ID is required to work with non-collaborative subjects, where there is no cooperation for enrolment. Thus, proven robust biometrics such as fingerprints, iris, or even facial recognition are not ordinarily a viable option. The use of more conventional and proven methods is also limited by the small number of pixels on which a subject is represented on surveillance images. Even with the use of high-resolution cameras, many surveillance videos are set to monitor extensive areas and rely on the use of a large field of view. Consequently, a small subset of the available pixels might be the norm for the RE-ID task.

Different from typical biometrics solutions, the goal of re-identification is not the verification or identification of a subject's identity but to associate where a given subject of interest has been previously observed. In other words, the identity of subjects is not known in RE-ID; there is no enrolment procedure and therefore no collaboration from the subjects is required. Since there is no ground truth for identity, RE-ID algorithms are often evaluated as a ranking task using the CMC (see Section 1.2.3.2).

RE-ID algorithms typically create a short-lived visual-based signature for each observed subject. These signatures need to be robust to variations in pose, view-angle, partial occlusions, lighting conditions, etc. In our work [3] we extended the potential lifespan of RE-ID signatures by making them also robust to attire changes.

1.2.3 *Performance assessment*

In this section, we will go over some conventional methods for measuring the performance of biometrics tasks. These tasks often fall under the general problems of classification and ranking. First, we will present the case of binary classification.

When dealing with multi-class classification problems, it is common to use the same performance metrics as for binary classification but to perform a pairwise comparison of classes. Pairwise comparison means that for each class, one does a binary classification of one class versus all other classes.

Coming back to the simpler binary classifier; It is common practice to denote the two plausible classes as the positive and negative classes respectively denoted here as $+$ and $-$. After a classifier is trained, one can compare its predictions with the real (ground truth) labels. Such comparison generates four cardinalities: the numbers of true positives, true negatives, false positives and false negatives. True Positive (TP) is the count of predictions that were correctly classified as $+$, while True Negative (TN) is the count of predictions that were correctly classified as −. False Positive (FP) is the count of − predictions that were incorrectly classified as + and finally False Negative (FN) is the count of + predictions that were incorrectly classified as −. These cardinalities are shown in Table 1.1.

	Predicted Values	
Ground Truth	Positive prediction (+)	Negative Prediction (−)
Positive Label (+)	True Positive (TP)	False Negative (FN)
Negative Label (−)	False Positive (FP)	True Negative (TN)

Table 1.1: Definition of common scalar metrics for binary classification.

Using these cardinalities it is possible to compute multiple scalar performance measures; some of the more common scalar ones are summarized in Table 1.2, while more complex derived ones are presented in the following sections.

Performance measure	**Formula**	**Also known as**
True Positive Rate (TPR)	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$	Recall
False Positive Rate (FPR)	$\frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}}$	False acceptance rate
Accuracy	$\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$	
Precision	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$	

Table 1.2: Common scalar performance measures for biometrics algorithms.

In the context of biometrics, one can consider the existence of a binary classifier for each every class (identity) in the gallery set (see Section 1.2.1.1). Then, for every query in the probe set, the aforementioned classifiers predict a binary (+ or −) outcome for their class. In the context of biometric classification FPR represents the risk of the system. It is also known as false acceptance rate, and it measures the fraction of imposters accepted by the biometric classifier. The TPR measures the convenience of the system. TPR is also known as the genuine acceptance rate, or recall, and represents the fraction of accepted individuals.

The relationship between the FPR and the TPR is often represented by a ROC, which summarizes the trade-off between convenience and risk of a biometric method, as presented in the following Section.

1.2.3.1 *The receiver operating characteristic curve*

The ROC curve is often used to assess the performance of classification tasks. Since classifications are often based on thresholding a distance (or probability) score, one can use different thresholds and plot the points that describes the relationship between TPR and FPR of a biometric system. By connecting these points one generates the ROC curve. An example of ROC curves is shown in Figure 1.2.

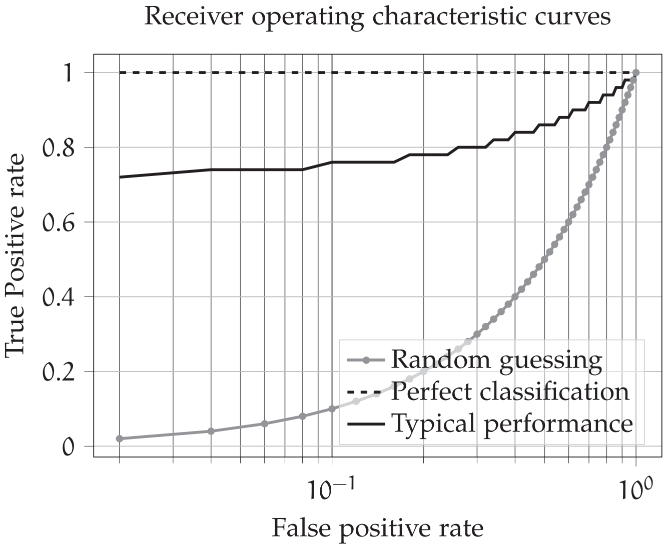


Figure 1.2: Sample ROC curves. This classification problem is emulated from a set of 50 identities. Notice that the FPR is presented in log scale and it is capped at 10^0 which represents 100%. Note that the classification threshold becomes stricter towards the left of the abscissa.

Figure 1.2 presents the ROC curve for three distinct classifiers. A (theoretical) perfect classifier would always predict the correct match, even at zero FPR, thus presented as a horizontal line. The random guessing classifier has equal FPR and TPR. Since the presented graph has the abscissa in log scale, a parabolic curve results. Finally we also present the performance a hyphotetical typical classifier to illustrate a behaviour between the random guessing and perfect classification spectrum.

1.2.3.2 *The cumulative matching curve*

The CMCs is one of the standard methods for evaluating ranking algorithms (such as identification and RE-ID) [8]. Ranking algorithms produce a list of probable matches from a gallery set and ordered by their degree of similarity to a given query input. Therefore, by using a set of query inputs (probe set) and a gallery set it is possible to estimate the probability that the ranking algorithm will retrieve the correct match within the top k most similar matches.

The matching probabilities of a given ranking algorithm are presented by rank. The CMC graphs the probability (ordinate) of having the correct match within the top k ranked items (abscissa). Three CMCs curves are shown in Figure 1.3 illustrating the performance of the theoretically perfect ranking solution, a random guessing ranking solution, and a typical hypothetical solution.

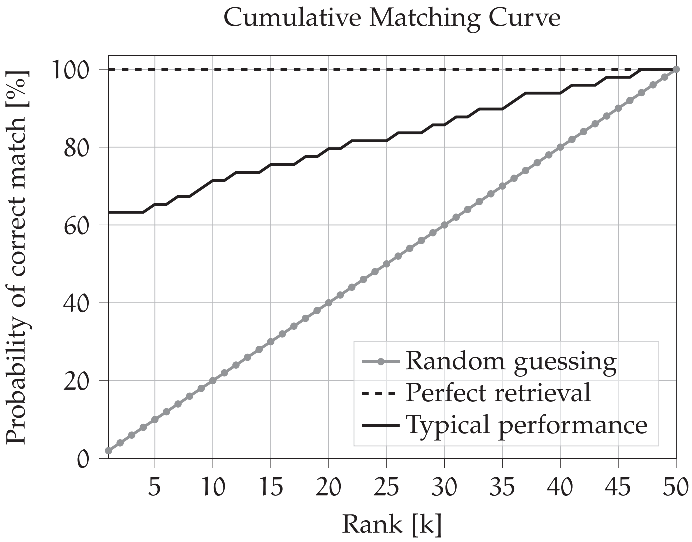


Figure 1.3: Sample plots for CMCs. For this sample ranking a problem the maximum rank (e. g. Number of subjects, number of retrieval classes) is set to 50. Therefore random guessing starts with a 2% probability of match. The typical performance serves to illustrate a hypothetical solution.

A perfect ranking system would always retrieve the correct match at rank 1. The random guess has a $\frac{k}{N}$ chance of retrieving the correct match for a given rank k and number of retrieval classes (e. g. subjetcs) N in the gallery. The CMC is a monotonically increasing function since the ranking system can not diminish accrued number matches as the rank increases.

For highly accurate ranking algorithms it is common to represent a sample of the lower rank values in a table [31, 62] along with a scalar metric, the normalized area under the CMC. The Table 1.3 gives an example.

METHOD	RANK 1	RANK 5	RANK 10
Random guessing	2%	10%	20%
Perfect guessing	100%	100%	100%
Typical performance	63.27%	65.31%	79.59%

Table 1.3: Sample of a performance summaries for CMCs presented in Figure 1.3. For this sample ranking problem the maximum rank (e. g. Number of subjects, number of retrieval classes) is set to 50. Therefore random guessing starts with a 2% probability of a match.

The normalize area under the curve ($nAUC$) is a common scalar metric used to compare CMC curves; it integrates the are under the curve of a given ranking algorithm and divides by the theoretical maximum given by the perfect ranking algorithm. This gives a simple scalar that can compare ranking algorithms with a different number of retrieval classes (e. g. subjects) N. However, the shape of the CMC curve provides critical insight into the workings of an algorithm. This shape information is lost when reducing a CMC to a scalar $nAUC$ since curves with different shapes can have the same $nAUC$.

1.2.3.3 *The precision-recall curve*

The precision-recall curve, similar to the ROC curve, is also commonly employed in the assessment of binary classification tasks, such as biometric verification.

A specific threshold value on the a distance (or probability) score of a biometric binary classifier defines specific scalar precision and recall values, see Table 1.2. The precision-recall curve is formed by defining different thresholds. It serves to illustrate the trade-off that a binary classifier makes. Precision tells what fraction of the *predicted + values* are correct while recall measures what fraction of the *ground truth + labels* are correctly classified. Therefore, a system with high recall and low precision would return most of the *ground truth + labels* data points, but the proportion of data points that is correctly classified as + is small.

A perfect theoretical classifier has zero FP. Thus the precision simplifies to $\frac{\text{TP}}{\text{TP}+0}$, and is always one. In this hypothetical classifier, no matter the threshold, the fraction of *ground truth + labels* is one. Therefore, a perfect classifier is portrayed as a parallel line to the abscissa where precision equals one.

Suppose that a random classifier has equal chances to classify a data point as + or −. This classifier should, in theory, have constant precision, and would be represented by a horizontal line. In this case, precision will be defined by the proportion of *ground truth + labels* to *ground truth − labels*. This is because the random classifier is essentially an unbiased sampler. So TP should be half the number of *ground truth + labels* , and FP should be half of the *ground truth − labels*. For example, if there is one *ground truth + label* for every nine *ground truth − labels*, the precision should be $\frac{0.5}{0.5+4.5} = 0.1$. Figure 1.4 shows examples of precision-recall curves for a theoretically perfect classifier, a rendom classifier and a simulated typical classifier response.

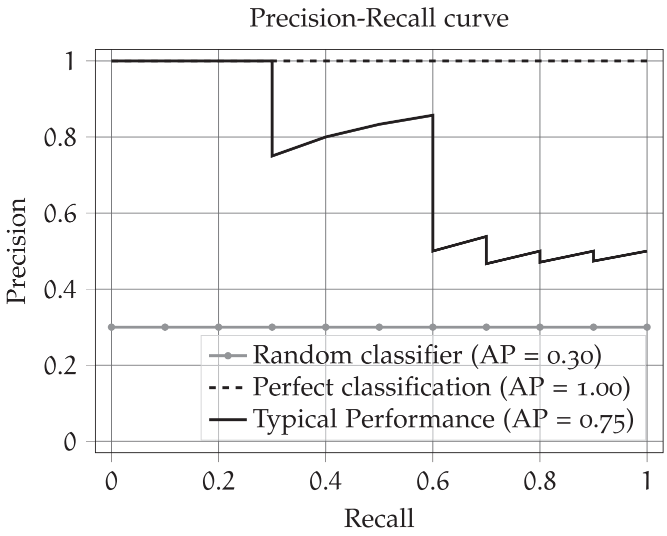


Figure 1.4: Sample plots for precision-recall curves. The illustration of three binary classification tasks are shown. The random guessing classifier has average precision of 0.30 because there are 3 *ground truth + labels* for every seven *ground truth − labels* on this hypothetical classification task. It is important to notice that the behaviour of a typical classifier is not necessarily smooth or linear.

A typical classifier that has good performance should have high precision and high recall. This is represented by the point on the top right corner of the precision recall diagram. Figure 1.4 presents the precision-recall curve for thee distinct classifiers.

The area under a precision-recall curve represents a performance scalar known as average precision (AP).The mean value of different average precision scalars (for different classifiers) is a base metric in recent RE-ID tasks. In this case, multiple observations (e. g. images) of a given class (e. g. subject identity) need to be available. The mean average precision mAP has been proposed as an alternative performance measure for re-identification benchmarks [62].

2

CONTRIBUTIONS AND PUBLICATIONS SUMMARIES

This chapter lists the thesis elected publications and highlights its research contributions. The publication listing follows the chronological order, explaining the evolution of this research on unconventional biometrics. The publication listing covers key contributions furthermore the reasoning and decisions that drove its publications/research.

2.1 PUBLICATION A

Igor Barros Barbosa, Theoharis Theoharis, Christian Schellewald, and Cham Athwal. "Transient biometrics using finger nails." In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. Sept. 2013, pp. 1–6. DOI: 10.1109/BTAS.2013.6712730.

2.1.1 *Motivation*

Our initial goal was to find better alternatives to the prevalent biometric traits. A major concern at the time was that usual biometric traits were not compatible with the right to be forgotten [43] and newer privacy concerns that today have manifested themselves in regulations like General Data Protection Regulation (GDPR) [41]. The intended solution was to study *transient biometrics*. Instead of recording time-invariant data, transient biometrics uses ephemeral data, i.e., data that does change over time and are thus canceled by nature.

Transient biometrics also makes it easier to comply with data life-span requirements, since old data are useless by definition.

Transient biometrics has the added benefit of acceptability. A subject is likely to be more willing to offer their biometric data if they know that such data has a natural expiration.

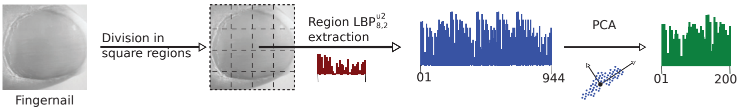

Fingernail
Division in square regions
Region $LBP^{u2}_{8,2}$ extraction
01
944
PCA
01
200

Figure 2.1: First proposal for a transient biometric identifier, presented in [5]. The biometric identifier is derived from Local Binary Patterns (LBP) which are represented by the 200 main components (computed using Principal Component Analysis (PCA)).

2.1.2 *Research contributions*

This publication addressed all four Ph.D. research goals.

2.1.2.1 *RG1: Novel Biometric Identifier*

The results of this publication (as well as of [7]) indicate that fingernail images are indeed a valid transient biometric identifier, since their useful lifespan is up to six months [59].

This biometric identifier utilizes texture features, extracted from fingernail images using Local Binary Patterns (LBP)[38]. The process starts by registering fingernail images, and then uniform LBP histograms are used to extract a 954 dimension vector. Finally, Principal Component Analysis (PCA) is used to reduce the vector to 200 dimensions; thus arriving at a numerical representation of the transient biometric identifier. This process is presented in the signature extraction pipeline shown in Figure 2.1.

2.1.2.2 *RG2: Unconventional methods of biometric recognition*

The introduction of the idea of transient biometrics in addition to the novel biometric trait of the fingernail image constitutes a novel method of biometric recognition, which is very unconventional due to its transient nature. The proposed approach and the collected dataset influenced several subsequent publications by the research community, such as [7, 14, 29].

2.1.2.3 *RG3: Application of machine learning methods in the context of biometrics*

A naive Bayes classifier was used to establish a baseline metric on the identification task. The combined usage of a baseline classifier and dataset created at different points in time allowed us to asses the transitive nature of our approach. The results

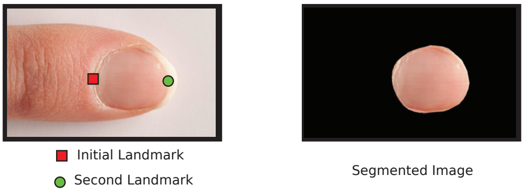


Figure 2.2: Active shape model input and segmented output.

allowed us to the argue that the proposed methodology indeed extracts a transient biometric signature from fingernail images.

2.1.2.4 *RG4: Creation of novel datasets*

A new dataset called Transient Biometrics Nails Dataset (TBND) containing fingernails images was made publicly available. The motivation was to provide a platform for verifying our results in transient biometrics and to give a boost to further research on the topic. This public dataset can be found at `https://www.kaggle.com/vicolab/tbnd-v1` . As of 03.12.2019 one hundred and thirty-three downloads have been made by third parties and multiple publications have referenced our paper.

2.1.3 *Technical contributions*

An Active Shape Model (ASM)[53] was used to segment the nail bed from a fingernail image. The segmented region defines a Region of Interest (ROI) that can be registered in order to facilitate the computation of a distance measure between such objects. (Registration is the task of transforming an input image to a universal and reproducible coordinate system.)

ASM segmentation had one major issue: it needs manual input to specify two landmarks, as shown in Figure 2.2. It was initially believed that segmentation needed to be very accurate by removing every pixel from the image that was not from the nail bed. This hard requirement proved not to be needed in the development of this publication due to the selected descriptor. Thus, a more robust (and coarse) segmentation algorithm that does not require manual input was seen as a requirement for future work.

All fingernails beds have a unique pattern, like a fingerprint, which influences the texture detected on the fingernail plate [27]. Combined with the day-to-day environment interactions that the fingernail sustains, gives a rich but temporal quality to the detected texture. The objective of the first publications was to explore the temporal traits of these textures in order to create a transient biometric descriptor.

LBP was used as the texture descriptor because of its previous success in describing texture in general [37, 38] as well as facial biometric texture in particular [2].

Temporal performance of the fingernail texture as a biometric descriptor was assessed based on the identification task using a dataset collected at different points in time. This dataset was made publicly available to aid reproducibility and stimulate further research.

2.2 PUBLICATION B

Igor Barros Barbosa, Theoharis Theoharis, and Ali E. Abdallah. "On the use of fingernail images as transient biometric identifiers." In: *Machine Vision and Applications* 27.1 (Jan. 2016), pp. 65–76. ISSN: 1432-1769. DOI: 10.1007/s00138-015-0721-y. URL: https://doi.org/10.1007/s00138-015-0721-y

2.2.1 *Motivation*

The motivation behind this work is to mature methods behind transient biometrics using fingernails and to address the shortcomings of *Publication A*.

The first concern was scalability of transient biometrics beyond a handful of participants.

The second concern was to reduce the possibility of injected bias in the evaluation of the transient biometrics features. Therefore for this publication, we opted out of the use of machine learning methods which can inject bias from the training dataset and the machine learning approach used. A direct approach was opted for.

The third concern was to mature the computational methods used to extract and match the transient biometric signatures. The extended number of participants and the direct approach meant that better supporting tools and matching algorithms become a requirement.

2.2.2 *Research contributions*

This publication addressed three of the Ph.D. research goals. Since it constitutes a continuation of the work described in *Publication A* [5], some overlap of the research goals exists.

2.2.2.1 *RG1: Novel Biometric Identifier*

In the work described in *Publication A*, a small dataset of 17 subjects was used to evaluate the temporal identification rate of fingernails. LBP was used as the feature extractor.

The present work exploits both texture features and descriptor-based information extracted from discriminant fingernail keypoints. Texture information is extracted using LBP on the three color channels of the image. Binary Robust Invariant Scalable Keypoints (BRISK)[30] and Scale-Invariant Feature Transform (SIFT)[33] are used to compute features on keypoints determined by Good Features To Track (GFTT) [46].

This work focused on the fusion of matching scores instead of fusion of features to avoid the curse of dimensionality. Texture-based descriptors are matched using the average of cosine similarities while keypoint descriptors are matched using Fast Approximate Nearest Neighbor Search[36] where inliers are filtered using the Random sample consensus (RANSAC)[15] algorithm.

Figure 2.3 illustrates the signature extraction process. These innovations in both the matching procedure and feature representation produced a new transient biometric identifier which constitutes a contribution to research goal 1.

2.2.2.2 *RG2: Unconventional methods of biometric recognition*

This publication extends the research on unconventional methods of biometric recognition beyond *Publication A*. Assessment of unconventional biometrics is extended to a group of 93 subjects. The use of fingernails as a source of information for biometric trait still remains unusual, with a handful of papers, such as [5, 26, 28, 52], exploring it.

2.2.2.3 *RG4: Creation of novel datasets*

A new version of TBND containing fingernail images was made publicly available. This time pre-segmented fingernail images are provided instead of full finger pictures, making the new

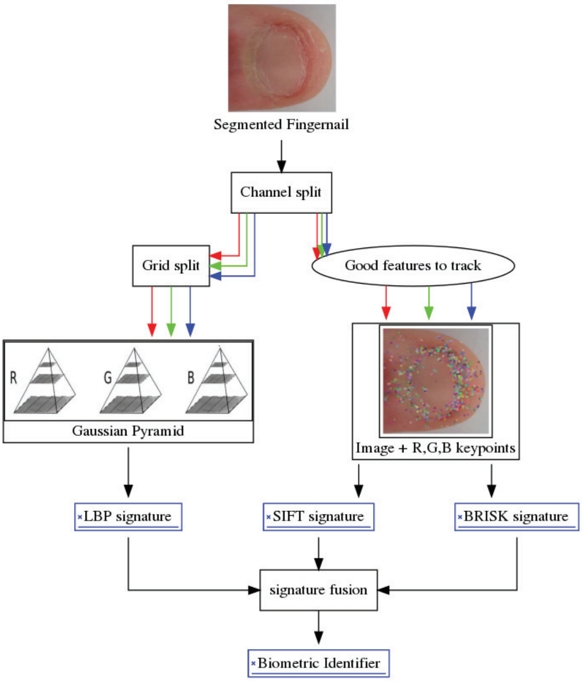


Figure 2.3: Pipeline for the transient biometric identifier, presented in [7]. A segmented fingernail image is split by color channel to allow extraction of signatures per channel. A Local Binary Patterns (LBP) signature is derived using a Gaussian pyramid and a 4×4 grid. A Binary Robust Invariant Scalable Keypoints (BRISK) and a Scale-Invariant Feature Transform (SIFT) based signature are computed using Good Features To Track (GFTT) keypoints. Finally a new algorithm for signature fusion merges the three signatures into the biometric identifier.

dataset easier to use in classification tasks. The dataset is a available at https://www.kaggle.com/vicolab/tbnd-v2. As of 20.8.2019 one hundred forty-two downloads have been made by different research groups and multiple publications have referenced our paper.

2.2.3 *Technical contributions*

The dataset extension to ninety-three subject made the manual selection of keypoints for ASM segmentation infeasible. As it turned out that a pixel-accurate segmentation was not needed, the fingernail was detected and segmented using the ROI defined by the Haar feature-based object detector proposed in [32, 55].

The idea behind this work was to use well-established features descriptors combined with a pre-determined set of feature fusion techniques to achieve a direct and unbiased feasibility evaluation. The direct approach would avoid dataset selection bias, and remove any influence of specific machine learning methods.

Feature description begins by expanding the LBP feature extraction presented in publication A. The goal was a more robust solution that could operate in RGB space. The process starts by dividing the segmented fingernail images into a four by four grid. Each one of the sixteen sub-images was submitted to three stages of Gaussian smoothing, creating a total of 48 sub-images. The final LBP features are extracted by color channel, giving us a total of 144 LBP signatures per fingernail image. Each LBP signature is a 59 bin histogram. A final texture matching score between two fingernail images is computed by averaging the cosine similarity between pairs of such histogram sets.

Further technical contributions arose from the fusion of texture features (LBP) and descriptor-based features extracted at discriminant fingernail keypoints. The typical approach might have been to create a combined feature set representation of the different features. We opted to match each feature individually and have a final matching metric based on the combination of matching scores. A small contribution involves the discussion of how scores based on RANSAC inliers are biased towards a low value and how cosine similarity score from LBP features have the opposite behavior. Consequently, the geometric mean is a fitting solution to compute a final matching score; the geometric mean does not weigh the LBP and descriptor-based features differently.

2.3 PUBLICATION C

Igor Barros Barbosa, Kenneth Vilhelmsen, Audrey van der Meer, Ruud van der Weel, and Theoharis Theoharis. "EEG Biometrics: On the Use of Occipital Cortex Based Features from Visual Evoked Potentials." In: *28th Norsk Informatikkonferanse, NIK 2015, Høgskolen i Ålesund*. Bibsys Open Journal Systems, Norway, Nov. 2015. URL: `http://ojs.bibsys.no/index.php/NIK/article/view/243`.

2.3.1 *Motivation*

Our original intention was to assess if the new generation of cheap **non-invasive Brain-Computer Interfaces**[†] could be useful in a biometric verification system and thus explore if 'brainwaves' can be used as a biometric trait. The concept was to ask different subjects to select a pass-phrase of their choosing. The selected pass-phrase was then silently repeated by the subject with closed eyes. Such a process is known as the recitation task. The rationale was that the recitation task would cause patterns on Electro-Encephalo-Gram (EEG) recordings that would be useful for biometric recognition. After a few trials and assessment of data, it was evident that data acquisition was an obstacle. The results lead us to discuss with a research group with domain-knowledge on EEG acquisition, specifically the NTNU Developmental Neuroscience Laboratory. Given their experience and available data, the project goal pivoted to the re-use of already collected EEG on infants. The work then concentrated on VEP. VEPs are electrical potentials, recorded by EEG, reflecting activity excited by visual stimuli. The objective was set to assess the possibility of extracting biometrically capable features from the recorded EEG.

[†] *The selected brain-computer-interface was the Emotiv Epoc.*

2.3.2 *Research contributions*

This publication addressed four Ph.D. research goals.

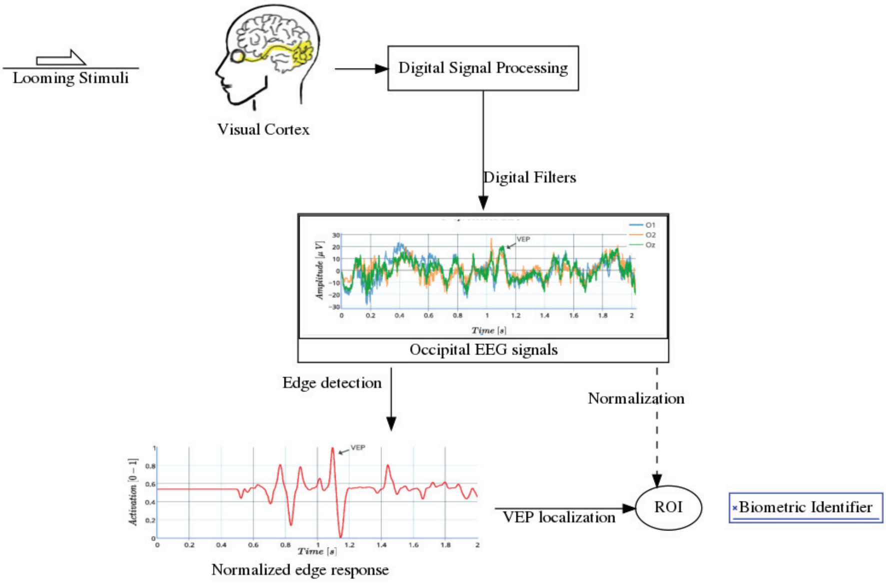


Figure 2.4: First proposal for a looming Visual Evoked Potentials (VEP) biometric identifier, presented in [6]. A looming stimulus (simulation of objects moving in the direction of the observer) is presented to a subject. Such a stimulus results in Visual Evoked Potentials (VEP) which are recorded as Electro-Encephalo-Gram (EEG) signals. Signal processing techniques are employed to segment the VEP response and define a Region of Interest (ROI). The biometric identifier is a numerical representation of the ROI using normalization and edge detection on the EEG signal.

2.3.2.1 *RG1: Novel Biometric Identifier*

The work introduces a methodology for computing a 200-dimensional feature vector that can be used to represent biometric features from brain activity recorded as EEG signals. The experimental results indicate that the extracted features are beneficial for biometric recognition of a small set of subjects.Figure 2.4 outlines the process behind the creation of the biometric identifier.

Biometric recognition from VEP is a niche field within biometrics. This work proposed feature extraction methods from EEG signals as well as initial experimental evidence that these features are useful for biometric recognition.

2.3.2.2 *RG2: Unconventional methods of biometric recognition*

This publication was the first study to assesses looming stimuli for the creation of biometrically useful VEP, i.e. EEG responses due to visual stimuli. Although the results from our initial acquisitions on recitation task were disappointing, the processing of the data from the Developmental Neuroscience Laboratory showed that indeed VEP can be used as an unconventional method of biometric recognition with limited success. In any case, the key to biometric recognition accuracy is the combination of multiple biometrics and, as such, it can be considered. It has the distinctive advantage of being very hard to spoof.

2.3.2.3 *RG3: Application of machine learning methods in the context of biometrics*

This publication makes use of a Multilayer perceptron (MLP) on a two-fold cross-validation task, since the EEG is a multidimensional signal which has been well handled by such classifiers in the past [40].

2.3.2.4 *RG4: Novel datasets for Unconventional Biometrics*

A new dataset of EEG responses to looming stimuli was made available at `https://www.kaggle.com/vicolab/eeg-looming`. This is to the best of our knowledge one of the largest public datasets on Brain activity that concentrates on VEP from healthy individuals. The introduction of this dataset is important since the largest previously published dataset [61] can be considered biased as the data was acquired from alcoholic individuals [13].

2.4 PUBLICATION D

Igor Barros Barbosa, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen, and Theoharis Theoharis. "Looking beyond appearances: Synthetic training data for deep CNNs in re-identification." In: *Computer Vision and Image Understanding* 167 (2018), pp. 50 –62. ISSN: 1077-3142. DOI: `https://doi.org/10.1016/j.cviu.2017.12.002`. URL: `http://www.sciencedirect.com/science/article/pii/S1077314217302254`

2.4.1 *Motivation*

Publications A,B, and C [5–7] focused on brand new or very niche biometrics, thus limiting benchmarking possibilities and potential collaboration with the biometrics research community.

We thus decided to address the RE-ID problem which, being more established, has large benchmarking possibilities and an established research community.

Still, the RE-ID task is unconventional. First, it does not focus on the common biometric problem of identification/verification of personal identity. Instead, it deals with the task of recognizing if an individual has been previously observed.

Second, RE-ID deals with a number of challenges uncommon to more established biometrics (such as 2D facial and fingerprint biometrics): these include non-collaborative subjects, reduced camera resolution, varying illumination, backgrounds and camera position. We decided to also include varying subject apparel in our problem statement in order to address a RE-ID scenario across multiple days. This adds a significant new challenge to the RE-ID problem, making short-lived visual-based signatures based on apparel obsolete.

The above challenges would be difficult to address with a direct approach, so we decided that a machine learning method, such as ANN would be more appropriate. At the time the training of ANNs required copious amount of labelled data and we wanted to investigate whether real training data could be replaced (or aided) by synthetic data. Synthetic labelled data has the advantages of being cheaper and faster to acquire, having no subject privacy issues, be well controlled in terms of the parameters involved and having its size limited only by the modeling capabilities. For example, the classes of a synthetic dataset can be perfectly balanced and specific **domain knowledge**† can be inserted.

† *Domain knowledge in the RE-ID scenario can be extended by changing the subjects' apparel*

2.4.2 *Research Contributions*

This publication addressed all four Ph.D. research goals.

2.4.2.1 *RG1: Novel Biometric Identifier*

The majority of re-id approaches focus on modeling the appearance of people in terms of their apparel. At the time, the deep

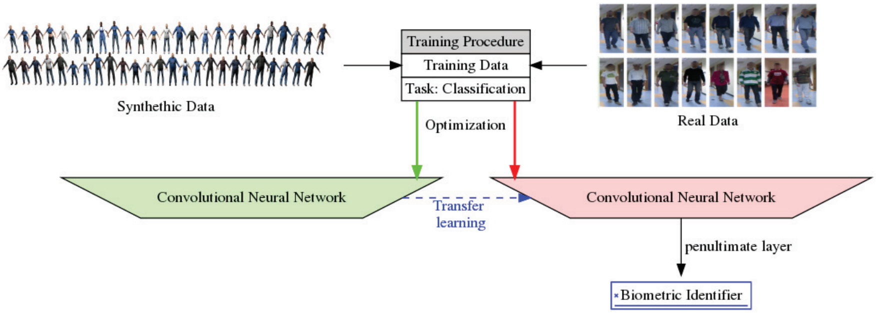


Figure 2.5: Novel Re-identification (RE-ID) based biometric identifier, presented in [3]. A Convolutional neural network (CNN) is first trained on a classification task using synthetic data. The 'knowledge' (weights and new normalized learning rate) from the first CNN is then transferred to another CNN which is subsequently trained on real-data. Finally a biometric identifier is extracted from the numerical output of the penultimate layer (before the classification layer).

learning methods approached the task by designing siamese ANNs which were usually a Convolutional neural network (CNN). Such CNNs receive two input images at once and are trained to return a boolean value indicating if both images are from the same subject. Therefore, siamese networks provided a matching score, but no biometrics signature; potentially limiting the fusion of such biometric systems with other biometric modalities.

Publication D diverges from these typical solutions by focusing on the creation of a re-usable biometric identifier that is extractable from a single image. The publication indicates that the output of the penultimate layer of a CNN trained to classify subjects could be used as a numerical representation for a biometric signature. Experimental results showed that RE-ID was possible even for subjects that were never seen by the network (during training). The proposed methodology was comparable to human performance on the re-id task for subjects that were observed in different days and with apparel changes. It was thus concluded that, with the proposed approach, we had created a rather robust descriptor for the RE-ID and possibly other biometric tasks. Figure 2.5 shows the steps needed to achieved the proposed biometric identifier.

2.4.2.2 *RG2: Unconventional methods of biometric recognition*

Compared to the leading biometric methods based on the fingerprint, face or iris, RE-ID is a niche area. The debut of a synthetic dataset to aid and inject domain knowledge to re-identification biometrics was a new unconventional method of training a machine learning algorithm for a biometric problem.

When coupled with the departure from the dominant siamese networks, this work significantly contributes to research goal two.

2.4.2.3 *RG3: Application of machine learning methods in the context of biometrics*

The work proposes to use a variation of the inception-V3 network [50] to tackle the RE-ID problem. The results indicate that a non-siamese CNN trained on synthetic data can achieve comparable performance to the siamese networks proposed at the time of publication. The divergence from the siamese structure simplifies the training process. A direct approach, different from siamese networks, does not require pairs of images to be trained. Therefore, the proposed direct method removes the chance of injecting bias through the selection of sample-pairs for the training dataset.

This work also made a contribution to the more general field by demonstrating, back in 2017, how a synthetic dataset could be used to train deep learning methods in visual tasks that can then be domain adapted and applied to real tasks. Such practice has become quite popular since, and has been used in further research in RE-ID [48, 54, 63].

2.4.2.4 *RG4: Novel datasets for Unconventional Biometrics*

A new dataset called *SOMAset* has been introduced by this publication. *SOMAset* is composed of 100 thousand images, which are synthetic renderings of 25 female and 25 male body prototypes. The 50 human prototypes resulted from a mixture of the three main somatotypes: ectomorph, mesomorph, and endomorph. Thus, the rendered subjects can represent mixtures of long and lean bodies, athletics bodies, and obese bodies. The synthetic dataset accounts for ethnicity by rendering subjects with different skin colors. From the 50 body prototypes, 18 have beige skin tones, two model Asian , and the remainder 32 prototypes are equally divided between caucasian and darker skin

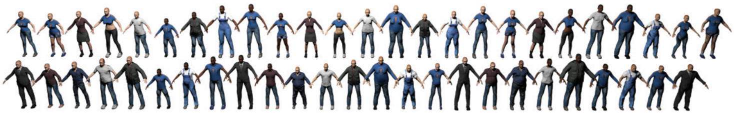


Figure 2.6: *SOMAset* [3] rendering of 50 humans prototypes.

tones. *SOMAset* can be considered as a **re-identification bootstraping dataset**† of 50 body prototypes (identities) that are instantiated by pose and appearance as described below.

The 50 body prototypes, shown in Figure 2.6, are rendered in 250 distinct poses each, using a different viewpoint. The dataset is further enlarged by generalising the re-id problem across appearance, by rendering the prototypes using 8 sets of clothes. 5 unisex sets of clothes are shared across all prototypes. Three sets of clothes are exclusive to female prototypes and Three exclusive to male ones. The enlargements of the initial dataset with different poses and different clothes introduces domain knowledge; the re-identification task can occur across different poses, different subjects that wear the same clothes (uniform simulation) and across different dates where a given subject is more likely to wear different clothes.

SOMAset can be publicly downloaded from `https://www.kaggle.com/vicolab/somaset` . As of 20.8.2019 two hundred and eighty-four downloads have been made by third parties and multiple publications have referenced our paper.

†SOMAset is meant to be a bootstrapping mechanism to be used in conjunction with fine-tuning on the training partition of a re-identification dataset. It was designed to provide a better initialization of the weights and thus achieve better priors.

2.4.3 *Technical contributions*

The proposed methodology showed the potential of our neural network which could respond to non-appearance features of the human silhouette — capturing structural aspects of the human body- which boosted re-identification performance. The following specific technical contributions were made in the application of machine learning on visual data:

A MODIFICATION ON THE INCEPTION NETWORK.
The work presents a reduction of the inception network that generates a 256 latent vector as the penultimate layer. The result is one of the few networks that could be trained from scratch and achieve competitive performance [22].

A TRANSFER LEARNING METHODOLOGY.
The presented methodology allows for a more aggressive

gradient update on the final layers, while the rest of the network has a reduced gradient update rule. The experimental results are evidence that the transfer learning method is successful in transferring knowledge from the synthetic dataset to the real dataset.

A METHOD FOR PROBING SPECIALIZED NEURONS

Given two small subsets of data, where a given property is part of the first subset but missing in the second subset, this work presents a method to find a group of neurons that are discerning towards the property that is unique to the first subset. Thus a new method for finding discerning neurons, in line with the distributed representation theory, is presented. This solution gives new insights on where and what the networking has learned to represent/detect.

3

DISCUSSION AND FUTURE WORK

3.1 DISCUSSION ON TRANSIENT BIOMETRICS

The use of fingernail images for the extraction of a transient biometric signature is a topic of both publication A [5] and publication B [7]. The assessment of possible degradation of the biometric signature, and thus working evidence of its transitive nature, is done using datasets acquired at three different points in time. The first dataset works as a reference point. The second dataset is acquired shortly after the first and shows what performance one can expect when the signature does not have enough time to degrade. The third dataset acquisition happens after a significant amount of time passes; long enough so that the fingernail can accrue significant temporal changes. The third dataset can help quantify the degradation of the biometric signature. Both publications A and B use the identification task to assess performance; in this task, one measures the probability of matching a signature from the reference dataset (first dataset) within the first P-ranked signatures (subjects) from the second or third dataset. Table 3.1 integrates the identification performance from Publication A and Publication B .

The results of publication A [5] show the potential of a fingernail transient biometric signature. Out of the subjects that were present on Day 1, 24 subjects returned on Day 8 and, out of those, 17 subjects returned on day 70. On the 17 subjects, a massive reduction of 58.82% in rank 1 matching probability was observed across Day 1 and Day 70. This substantial performance degradation agrees with the biological literature that indicates the lifespan of fingernails to be up to six months [59]. The reduced set of 17 subjects is sadly not large enough to measure scalability and also limits the potential of assessing different signature extraction techniques. The proposed method-

DATASET	ACQUISITION DAY	RANK 1	RANK 2
17 Subjects [5]	Day 8	100%	100%
	Day 70	41.18%	88.24%
24 Subjects [5]	Day 8	91.66%	100%
93 Subjects [7]	Day 2	86.02%	94.62%
	Day 30	56.98%	69.89%

Table 3.1: Integrated fingernail identification performance from [5, 7]. The reference datasets were acquired on day 1 in each case, while the test datasets were acquired on the future dates indicated by the *Acquisition Day* column.

ology achieved perfect re-identification in the short time frame for 17 subjects and nearly perfect for 24 subjects.

Publication B [7] is based on a significantly larger dataset of 93 subjects, allowing to answer questions of scalability and signature extraction robustness better. The enlarged dataset also carries new data acquisition challenges. It is considerably harder to get a bigger group of subjects to provide data for a prolonged period. To avoid the pitfalls observed on the first dataset acquisition, we had to limit the dataset acquisition period to one month. The initial exploration showed that the LBP signature extraction technique presented in [5] was not robust enough. Publication B thus proposes an extension of the LBP signature that can work in RGB colorspace. Subjects were ranked using the cosine similarity. The classification results give a true-positive-rate of 0.581 for a false-positive-rate of 0.01 for fingernail images across Day 1 and Day 2. That practically means that 58.1% of subjects were correctly matched while one in a hundred imposter attempts were wrongly accepted. However, even better performence was achieved by a combination of LBP, BRISK, and SIFT for signature extraction; this gave us a true-positive-rate of 0.774 for a false-positive-rate of 0.01 across Day 1 and Day 2. The new, more robust, signature extraction technique still observes the temporal degradation of the fingernail biometric signature; a true-positive-rate of 0.247 for a false-positive-rate of 0.01 is achieved when matching fingernail images across Day 1 and Day 30.

Transient biometrics were initially motivated to address privacy concerns in biometric systems. The fingernail has a relatively fast temporal degradation and is thus suitable as a transient biometric trait valid for days. Any transient biometric system needs to be designed with the concept of a biometric signature lifespan; old transient signatures should either be deleted or updated regularly.

The fingernail is a trait the subjects have control over them. The subject can assert control over the fingernail shape through clipping, and determine the fingernail texture through the use of file or nail polish. Therefore, the proposed solutions give any subject the chance of forcing the annulment of their temporal biometrics. Consequentially transient biometrics using fingernails can be canceled in two way. One way is through natural degradation by time, and a second way by giving the subject the freedom to control the validity of their biometric.

3.1.1 *Future Work*

Transient biometrics is a new and open research field, with various possibilities for further exploration and exploitation:

EXPLORATION OF OTHER TRANSIENT BIOMETRIC TRAITS.
It is an open question whether fingernails are the best source of information for a transient biometric. Our work has shown the temporal potential of fingernails, but we hope to see transient biometrics being derived from other biometric traits.

NEW SIGNATURE EXTRACTION TECHNIQUES.
The present work introduced transient biometrics using the fingernail as the transient biometric trait. Therefore only a small set of techniques, from the multitude that are available in the computer vision armoury, were explored for this problem.

FURTHER SCALABILITY STUDIES.
As new methodologies for signature extraction push performance boundaries, new research should also push the limits of scalability. Fingernail biometric signatures, given their new status, have a long way until performance allows to scale to the level seen in other well-established biometrics. To this end, much larger datasets must be acquired.

DATA PRIVACY AND GDPR COMPLIANCE.
Data privacy is a concern that will only increase as our so-

ciety becomes more and more data-centric. We believe that transient biometrics is a step in the right direction: helping in both the temporal validity and acceptability of biometrics. Transient biometrics might be one of the solutions that could help achieve better privacy-acceptable biometrics. It also has the potential to be more easily made GDPR compliant; the acquired data already has a validity lifespan and can be removed as useless from archives after expiration. The removal of expired data should have no impact on the performance of the transient biometric system.

SYNTHETHIC DATA.

Novel transient biometric signature extractors could be trained using synthetic data. Their knowledge can then be transferred onto real data. We envision two main approaches for creating such synthetic fingernail data. The first is to use computer graphics rendering. In this case we would algorithmically generate fingernail images along with micro patterns on those to define the nail textures. The second approach is to use a Generative Adversarial Network (GAN)[19]. See also Section 3.4 for more details on synthetic data.

3.2 DISCUSSION ON EEG

The concept of EEG based biometrics is appealing because it represents a potential solution robust to biometric *circumvention*. One can create fake fingerprints or spoofing attacks for visual-based methods of biometric recognition. However, to spoof an EEG pattern would require more effort from potential fraudsters. However, the main problem with EEG is that measuring a reliable and reproducible signal is arduous. The rationale for publication C [6] was that the new generation of cheap non-invasive Brain-Computer Interfaces could be a reliable hardware solution to the data acquisition problem. However, this did not prove to be the case. The failure to extract a reliable EEG signal is most likely due to compound errors, which were not trivial to diagnose. There could have been methodological problems in our acquisition procedures or lack of domain-knowledge in EEG acquisition and processing.

The above acquisition failures prompted us to use readily available data from the NTNU Developmental Neuroscience Laboratory. The biometric identifier would be defined by the behaviour of the Occipital electrodes, that measure a VEP response. We attempted to automate the process of detecting the VEP response within the EEG signal.

Previous literature points to the fact that VEP occurs in frequencies higher then 1.8Hz [1, 35, 57]. Our work estimated that looming VEP occur in frequencies lower than 20Hz. We thus opted for three infinite impulse response filters to refine the data. The first filter is a high pass filter with a cutoff frequency of 1.6Hz. The second is a low pass filter at 20Hz, and the third is a notch filter to clean the influence of alternating currents at 50Hz. After this filtering, an edge detection algorithm was applied to detect the peak of the VEP response in the EEG signal. Motivated by the sequence of operations in the classic Canny edge detector [10], we applied the following sequence of signal processing techniques: moving average, finite derivation and Gaussian Filtering.

The biometric identifier depends on extracting capable biometric information from the segmented VEP response, done by a feature extraction step. For feature extraction, our work proposes a ROI that is 200ms long, where the VEP occurs at the 80ms mark of that ROI. The biometrics signature is composed of two signals per ROI: the first is the normalised EEG readings and the second consists of the output of the proposed edge detector.

This data is finally fed to a MLP classifier. The results indicate that the proposed feature extraction can produce biometrically capable EEG based features. The performance was, however, quite a bit lower than that achieved by classical biometrics methodologies. Nevertheless, one can argue that it helped to produce biometric signatures based on EEG that requires fewer samples (VEP trials) per subject when compared to competing methodologies.

3.2.1 *Future Work*

Reliable data acquisition seems to be a big issue with EEG based biometrics. A possible line of future work for acquiring VEP responses would be to use Functional near-infrared spectroscopy (fNIRS) signals. fNIRS uses infrared light to measure the change of oxygenation level in hemoglobin in the brain. It might realise an engaging alternative for VEP acquisition.

Without a better and very reliable acquisition process, a personal recommendation would be not to spend time and effort on EEG or VEP based biometrics.

3.3 DISCUSSION ON RE-ID

Publication D [3] focused on a topic of great current interest, which is the re-identification of people using deep ANNs. What makes the paper unique in the re-identification literature is the way that the network is trained from scratch using a synthetically generated dataset. So far, collecting datasets for re-identification has been a costly operation, and our approach solves this issue. Besides, our framework allowed us to break the traditional barrier of re-id, that is, that people are not allowed to change their clothes between camera acquisitions. The work also presented a network, which at the time departed from the commonly used siamese architectures in the re-id task.

SOMAset is the name of the synthetic dataset, which is to the best of our knowledge the first `synthetic dataset` designed to aid in the re-identification task The use of synthethic dataset to aid in the Re-id task is now an approach explored by other solutions [48, 54, 63]. *SOMAset* was created to bootstrap the training of machine learning methods such as ANN. The idea is that domain-knowledge is injected during the synthesis of data. The injected knowledge gives the machine learning methods better prior (represented as network weight parameters in our case) for tackling real problems.

A more in-depth discussion on synthetic data happens at Section 3.4.

The work paired the presentation of *SOMAset* to the proposal of *SOMAnet*. *SOMAnet* is a modification on the Inception v3 [50] network designed to produce a small descriptor based on subjects appearance and somatotypical traits.

At the time of publication, the work showed that using synthetic data could achieve state-of-the-art performance on four public re-identification datasets: CUHK03 [31], Market-1501 [62], RAiD [12] and RGBD-ID [4]. The experiments with CUHK03 and Market-1501 show that the proposed network and synthetic dataset could perform comparably to the state of the art methods from that time; even when large dataset are already available for training. We achieved those results with a simple network and the novel idea of synthetic data while others used complex networks. The results also indicate that the synthetic data can aid established datasets and networks.

The experiments with the RGBD-ID and RAiD datasets showed that synthetic data is vital for training machine learning methods that are constrained by small datasets. The proposed method shows excellent performance on both datasets, even though they pose troublesome challenges. The RAiD dataset presents images of 41 subjects collected by both indoor and outdoor cameras resulting in a minute dataset with challenging illumination properties on which one would not expect a deep ANN to be trainable at that time. The proposed solutions established a new state-of-the-art for all proposed tasks on this dataset.

The RGBD-ID dataset presents 79 subjects that are observed on different days and with different apparel. The dataset presents both subjects with different clothes, and in some cases, subjects are using the same attire (Uniform simulation). Thus, RGBD-ID is characterized for breaking the appearance constraint that is common in the re-identification task. The proposed biometric framework using both SOMAset and SOMAnet achieves excellent performance on the RGBD-ID task. When using SOMAset to bootstrap SOMAnet, the achieved rank-1 re-id performance of the network was 63.29%, while the average human performance is 65%. If no synthetic data is used, the rank 1 performance is 22.78%. The achieved performance results using hand-crafted features set by the previous state of the art [23] was 17.72% at rank 1.

The achieved performance in the RGBD-ID dataset paired with the new method for probing specialized neurons gives substantial evidence that this biometric solution can see beyond appearance. For example, as is shown in Figure 7 of the paper, specialised neurons (see Section 3.3.1) were found which respond for ectomorph subjects. The results indicate that the network is capturing structural aspects of the human body, such as the somatotype. The retrieval results presented in Figure 3.1 serve as an illustration of re-id beyond appearance. Arguably, the presented approach has pushed re-id performance in this area by dissociating structural body cues from clothing cues. Therefore, it is now conceivable that the re-id concept can be included within the realm of person recognition technologies and, we feel, that our paper is a step in this direction.

This work proposed a method to assess the impact of illumination, pose and camera viewpoints on the training of the ANN. Experimental results allow the determination of the influence of these variables when modeling a synthetic dataset for re-id.

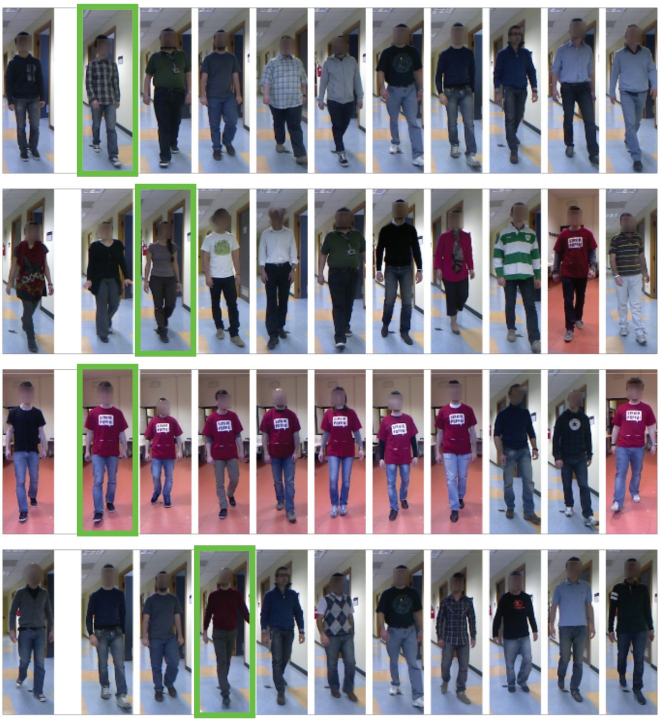


Figure 3.1: Ranking results of RGBD-ID[3]. Probe images are shown in the left column. The top 10 ranked gallery images are shown on the right. The ground-truth match is highlighted with a green frame

The presented solution gives the research community quantifiable clues on the influence of these variables in other datasets. It turns out that pose has the highest impact out of the three variables, making it attractive to include a large number of poses. The impacts of camera viewpoint and illumination are smaller and comparable to each other.

A further assessment was done on the effect of the number of rendered poses vs the number of rendered subjects. The results indicate that the number of subjects is more critical for performance. One can thus conclude that the order of importance of the variables tested is: subject variation, pose variation and finally illumination and viewpoint variation. One should therefore expend more effort in injecting variation by modeling and rendering the variables in that order.

The re-identification task presents a set of motivating challenges to the biometrics community. Non-collaborative subjects, reduced image quality, lack of control during acquisition, camera variation and unconstrained illumination are some of these challenges that are not found in most biometrics systems. The ground truth and labeling of re-identification datasets is also a challenge. The solution for labeling and generating such a dataset is either programmed/posed as in [5] or most likely derived from tracking individuals on video streams. The tracking solutions can generate label data by cross-matching two frames of a tracked subject. As soon as video tracking solutions identifies a subject observed by two cameras, a "ground-truth" label is produced. Nevertheless, most of the systems do not have a solution in place for detecting when the same individual might be re-observed with different apparel. Therefore, for observation over many days where the same subject is likely to return using different clothes (work, university, supermarket, public square); There is a disconnect between ground truth the provided data labels. Nevertheless, RE-ID can also be an attractive task for the biometrics community. Data acquisition for re-identification is cheap; it has almost no impediment by user behaviors; it can be acquired and assessed at a vast scale. As RE-ID methods evolve and break away from the matching of appearance, one can talk about non-collaborative person recognition at a distance, with a strong potential for fusion with other biometrics modalities.

3.3.1 *Understanding a neural network through visualization*

Given the importance of specialised neurons as discussed above, in this section, we would like to discuss unpublished work and our findings that led to the development of our proposed methodology to find specialized neurons as presented in Section 5.2.2 of Publication D [3].

We apply a new visualization technique (see Appendix I) in order to explore different layers of an ANN that we constructed based on Google inception V1 [49]. Specifically, we have chosen to concentrate on a shallow, middle and deep layer, as indicated by the red, blue and green stars respectively in Figure 3.2. We expect that deeper layers will encode features with higher abstraction power.

3.3.1.1 *Visualizing neurons in a shallow layer*

The first probed layer is the max-pooling indicated by the red star in Figure 3.2. This pooling layer follows the first convolution layer. While the convolution layer works as a filter bank, max-pooling will eliminate non-maximum values and add some spatial translation invariance to data from filter bank. Probing this layer shall give an understanding of how the first filter bank is processing images.

To probe this layer using our visualization technique (see Appendix I for more details on the visualization methodology), it was selected a random image from a *pre-soma* dataset. *Pre-soma* is the first versions of our synthetic dataset, and it differs from the published version of *SOMAset*[3] for rendering images with a large field of view. The large field of view is not ideal for re-identifications tasks because most of the information from the pixels are non-descriptive for the subject identity. This finding that is now obvious was made clear by our study covering the visualization of different layers of this network, motivating to change the rendering field of view that yielded *SOMAset*. The image from the *Pre-soma* dataset is shown in Figure 3.3a, while Figure 3.3b uses our visualization technique to probe which pixels of the input image influence activations of the first max pooling layer

The visualizations of Figure 3.3b indicate what has been learned by the first convolution layer. For most of the activation images, the rendered subject has discerning activation values. It is also clear that some of the images have different activations for

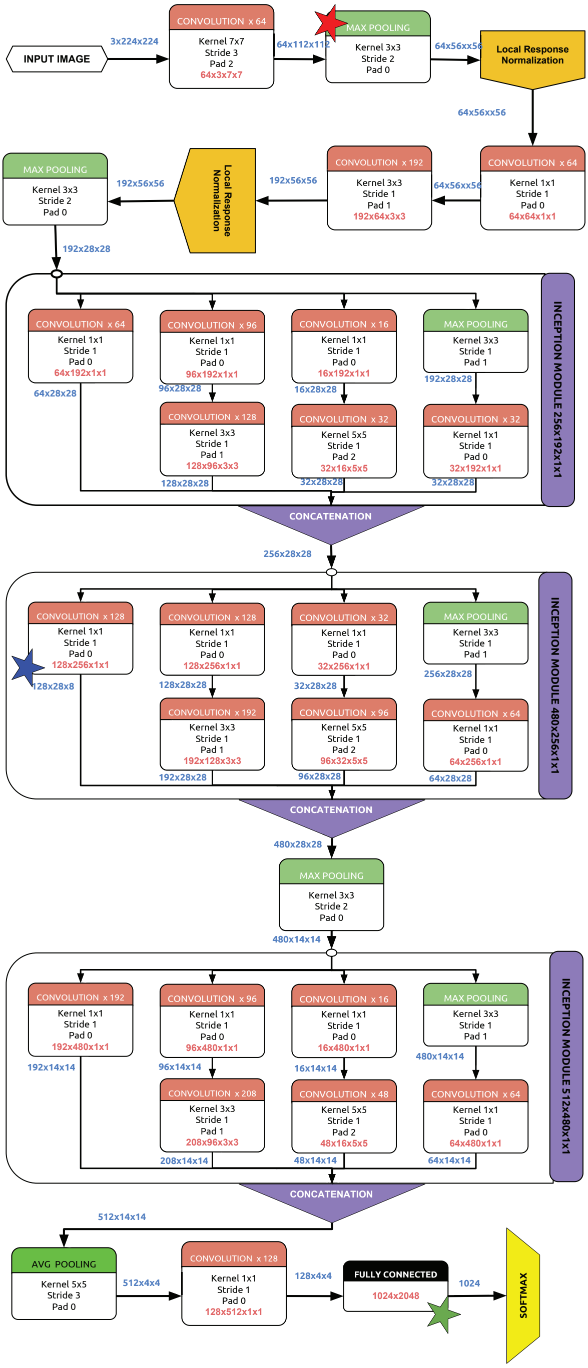
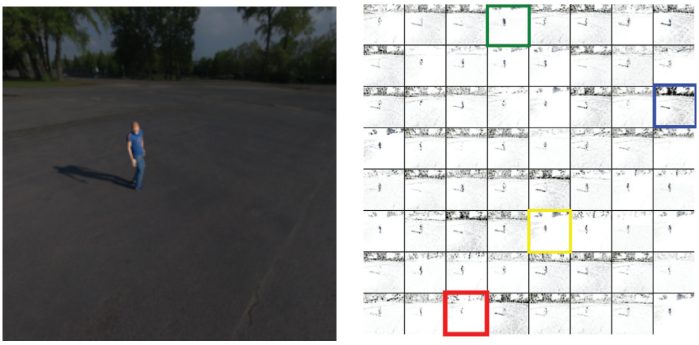


Figure 3.2: Our first trained ANN for RE-ID using syhntehtic data.



(a) Sample Input image.

(b) Probing images - First max pooling.

Figure 3.3: Images generated by probing a shallow layer.

different parts of the subject body. Some activations are high for the torso, some others for the shoulder lines, some for the arms and so on. The surprising effect is that although the same background was rendered for different subjects, the filter banks learned by this network still generate discerning activations for the surroundings. The images highlighted with color contours in Figure 3.3b are visible with more detail in Figure 3.4. The results seem to indicate that the network uses a filter bank specialized in relating the background to the user, presumably to have a better description of the subject.

The selected vizualizations of Figure 3.4 illustrate discerning segmentation behaviour. The white pixels indicate image regions that do not activate the probed neuron. Colored pixels indicate image regions that are important for the neuron activation, the transparency value giving the degree of importance. The first image is segmenting out background and skin (head and arm), the second and third images are from neurons segmenting the blue color. The fourth image shows segmentation of background features (shadows). These segmentations seem like simple filters for edges, brightness and color.

3.3.1.2 *Visualizing neurons in a middle layer*

An analysis of a second and third layer from the ANN should give us an intuition of what are the intermediate and deep features selected for this task. For a middle layer we will analyse activations inside the second inception module, more specifically we will look at the activations of the 1x1 convolution indicated in Figure 3.2 by a blue star. This layer does a sparse

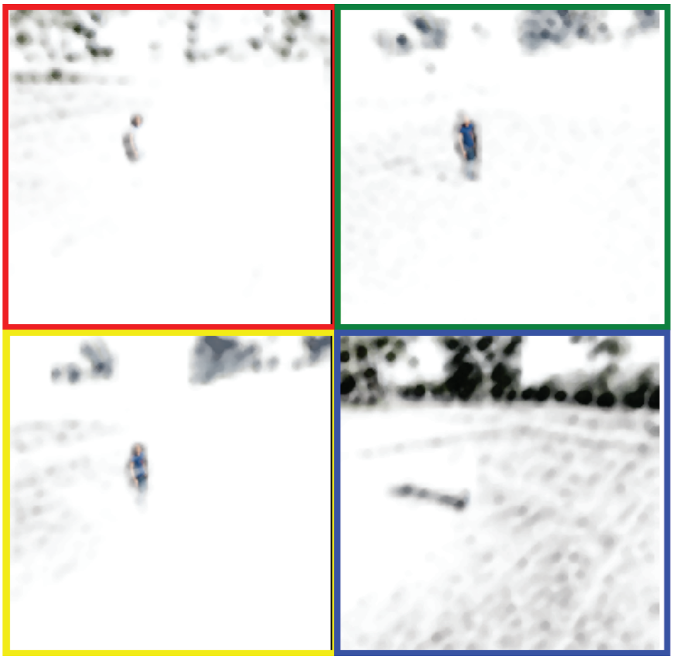


Figure 3.4: Selected visualization from shallow ANN layer.

dimensionality reduction of the previous convolution windows. Therefore the probing of this layer shows the high correlation clusters identified as discriminative. A few images computed by probing this layer are shown in Figure 3.5.

These image from Figure 3.5 show what the network learned to be discriminative clusters of information. The images on the first row are showing the full subject or most of it. The third image shows torso and legs and exemplify a neurons that fires for a subject body's parts. The final image show that the network found the horizon line informative/discriminative. It is also possible to notice that neurons might have higher activations for specific body parts. On the first image is visible that the selected neuron has the biggest activation from the all the four presented neurons. These results give us a few indications on the middle level features. For most of the images with visible activations, the rendered subject has the highest values, indicating that ANN has learned to detect subjects and that they are key to define the class. Again some neurons have activations for the background, if the background did not carry helpful information for defining the subject identify (the class of the image), we would expect to not so many activations images for it. The behaviour to look for hints in the background to iden-

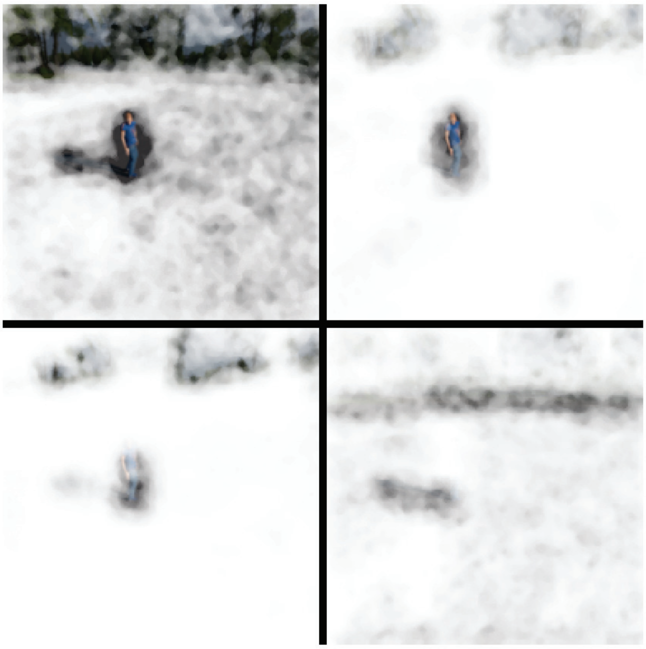


Figure 3.5: Images rendered by probing four random neurons. The selected neuron are from the layer indicated a blue star in Figure 3.2.

tify the subject is probably a reflection of the way the synthethic dataset was rendered.

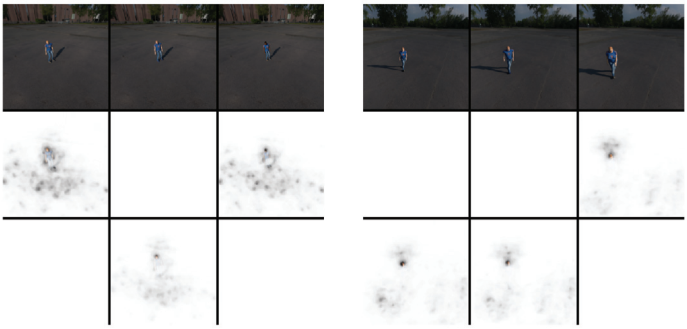
3.3.1.3 *Visualizing neurons in a deep layer*

A striking fact when visualizing this deep layer (see green star in Figure 3.2) is that, in contrast to shallow and middle layers, the images generated (activated pixels) are, at first hand, remarkably similar semantically. Whenever neurons of this layer are probed, they seem to always focus on the same pixel regions. Typical responses are shown in Figs. 3.6a and 3.6b. The responses have activation peaks around the shoulder lines/-head and the floor. The high activation for head region and floor seems to agree with the results of [4] where it was found that head and floor regions were the source of the most descriptive set of features. Although different neurons respond in similar image regions, we believe that they are describing a different semantic property. Hence, probing a deep layer solely by generating neuron activation images may not be sufficient. This motivated us to develop the a method for finding specialized presented in Sec.5.2.2 of [3]. Using such methods allowed us to easily find specialized neurons. For example, Figure 3.6b shows the result of probing neurons that respond to the obese/non-obese subjects. We were also able to find neurons at the deep layer that were responding to gender, as shown in Figure 3.6a

The first row of Figure 3.6b shows three different synthethic input images from non obese female, non obese male and obese female subjects respectively. Columns represent subjects. The second row shows the resulting probe images for a neuron specialized in obese subjects. The third row shows the resulting probe images of a neuron specialized in not-responding to obese subjects.

The first row of Figure 3.6a shows three different SOMAset input images from female, male and female subjects respectively. Columns represent subjects. The second row shows the resulting probe images for a neuron specialized in female subjects. The third row shows the resulting probe images of a neuron specialized in not-responding to female subjects.

The results of specialized neurons for gender and weight are indicative that this layer has neurons which represent a higher level of abstraction. We also found that fine-tunning the trained ANN to a real dataset also elected similar responses for obesity, as show in Figure 3.7



(a) Visualization of two neurons responding to female/male subjects.

(b) Visualization of two neurons responding to obese/non-obese subjects.

Figure 3.6: Visualization of neurons responding to gender and obesity.

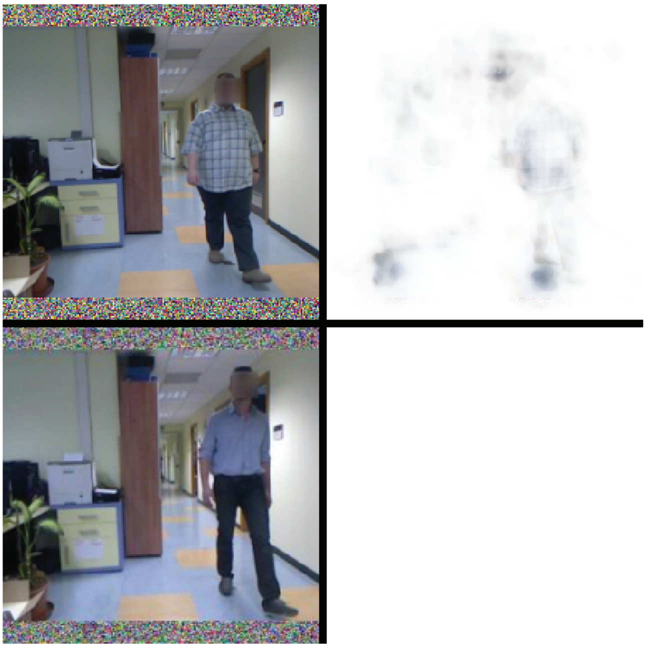


Figure 3.7: Visualization of a neurons responding to obese subject on non-synthethic images.

Figure 3.7 shows different RGBD-ID[4] input images from obese male and non obese male, respectively. The second column shows the resulting probe images for a neuron specialized in obese subjects.

The sparsity of the pixel activation observed in both shallow, middle and deep layers also motivated us to change the rendering pipeline. The change came because a large part of the pixels, showcasing the 3d surrounding, was not electing responses from neurons of a deep layer. Thus, it was decided to have rendering with a narrow lens, where most of the rendered pixels would show the rendered synthetic subject. Figure 3.8 shows the difference between dataset rendering.

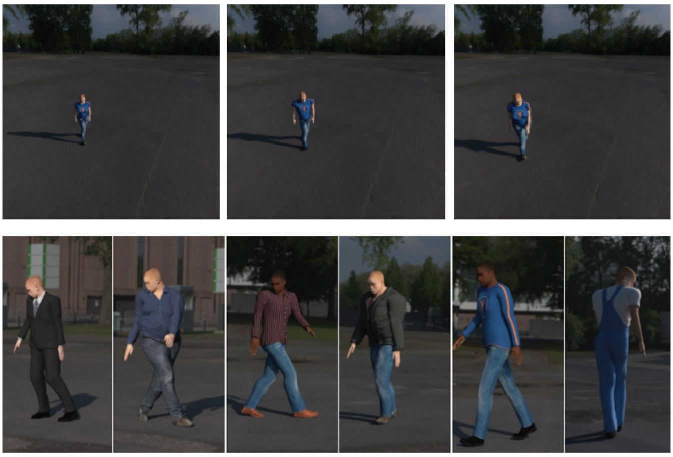


Figure 3.8: Difference from *pre-soma* and *SOMAset* [3]. Top row shows images of different subjects from *pre-soma* dataset while bottom row shows sample images from *SOMAset*.

3.3.2 *Future Work*

MORE ADVANCED NETWORKS AND BECHMARKS.

Publication D [3] used what was at the time a state-of-the-art artificial neural network [50]. However, the development of efficient ANN's is moving quickly, and new topologies and configurations are continuously presented at an increasing rate. A possible lineup of future work is to explore the use of more modern ANN topologies like Resnet [21], U-Net [42] and attention-based [56] to cite a few. One can also assess which topology is better suited to transferring knowledge from synthetic to real data.

RE-ID benchmarks, which are crucial for research, could significantly improve by including computational cost and its growth rate with respect to dataset size. A neural network that performs a one-to-one match, such as a siamese network, will not scale up well with dataset size. The scaling issue arises because every query image needs to be verified/matched against every image in the gallery set. It is not easy to propose a standard way to measure the scalability of proposed approaches. But such a benchmark would create incentives for coming up with ANN topologies that are computationally efficient while maintaining accuracy. One possibility for such a benchmark would be to assess RE-ID using only pre-computed embeddings along with their computational cost.

EXPLORE RE-IDENTIFICATION IN MULTIMODAL SOLUTIONS.

Re-identification is a very interesting biometric approach, as it eliminates the need for subject collaboration while current solutions achieve very respectable ranking performance on very large datasets. Nevertheless, the lack of quality assessment in the dataset maintenance process creates challenges. A very large dataset is not immune to labelling errors and is not easy to maintain. One line of work can be to exploit re-identification techniques in scenarios where subject collaboration is available. This means using re-identification based signatures in multimodal fused biometrics. In that case of course we would be talking about classic recognition tasks instead of RE-ID.

NON-COLLABORATIVE PERSON RECOGNITION AT A DISTANCE.

The indication that RE-ID solutions can associate identity despite changes in apparel can lead to interesting future work in non-collaborative person recognition at a distance. Such solutions would need to perform identity association over extended periods of time and they could borrow a lot from current RE-ID benchmarks and solutions. A difference would be the necessity to provide ground truth on identity across time. Therefore, data acquisition would be more expensive. Based on our experience ([3, 4]), we expect that such solutions could perform close or above humans.

3.4 DISCUSSION ON SYNTHETIC DATA

Publication D [3] presents a synthetic dataset named SOMAset. SOMAset was designed to work as a bootstrapping mechanism for deep learning and ANN that depend on large amounts of labeled training data. Thus, a significant contribution of this work was the demonstration that a synthetic dataset can be used to provide better priors for a Neural Network. The work also shows that domain knowledge can be injected and modeled in the synthetic dataset, helping machine learning algorithms perform better on harder tasks.

The approach used in publication D was computer graphics rendering; for that we need to have a model of the object that we want to generate (in the case of publication D, human objects). However, in many cases, no such computer graphics model exists.

An alternative possibility for synthetic image generation could be based on the use of a GAN[19]. A GAN is usually composed of two ANNs. The first network is a generator that transforms samples of latent variables (a random seed) into samples of a probability distribution that we would like to learn. The second network is a classifier that attempts to discriminate samples from the generator network and real samples from a training dataset. The two ANNs of a GAN are trained in an adversarial setting, competing with each other. Therefore a GAN converges when the discriminator can no longer differentiate between real data and samples generated by the generator. GANs have been successfully used in vision-based classification tasks [17, 45] and can be employed as a solution for synthetic data.

4

CONCLUSIONS

4.1 CONCLUSIONS

4.1.1 *A Race for peanuts?*

The massive proliferation of machine learning methods based on deep learning has created a hype evidenced by the volume of scientific publication activity in this area. This hype has partly proven useful for the research communities in machine learning. New tools driven by the hype are being developed at a fast pace; such tools are now becoming a scientific commodity and have spread to areas outside of deep learning. For example parallelism in scientific computing based on Graphics processing units (GPUs) has become very accessible as a result of the investment of major companies in open source projects like TensorFlow and Pytorch. Such projects have given users access to ultra-fast processing (typically tens of teraflops) for a fraction of the cost of yesteryear. The hype in deep learning methods has also been accompanied by the proliferation of learning materials that constantly brings new users to the exciting field of machine learning, statistics, and pattern recognition.

However, a downside of the hype is the creation of a scientific publication engine that feeds off an avalanche of publications in this field. An engine that does not scale well. For example, in CVPR 2019 (one of the major conferences in the field), according to the official statistics [16], 5160 paper submissions were registered. With an average of 3 reviews per paper, one would expect in excess of 15 thousand reviews. It has proved extremely difficult to find qualified reviewers and of the 2887 reviewers registered, over 700 were students where the minimal qualification was to have authored two papers.

Another downside of the hype and the resulting 'mechanical' reviewing process that has resulted, is that a large portion of publications in machine learning are focusing on beating performance on benchmarks. One can thus observe that a large part of the contributions are ephemeral, where publications demonstrate methods that apply an overfitting solution which achieves an incremental performance gain on a particular benchmark. This craze sends wrong messages to novice reviewers who get biased to expect that a new contribution solely means to outperform previous methods on a benchmark. Is such a race for peanuts worth it?

4.1.2 *The Challenges of Novel Biometrics*

This race for incremental performance gains on benchmarks, which is based on the exploitation of the reviewing system, essentially discourages the exploration of potentially groundbreaking new avenues.

The first research goal (and a major part of the effort) of this Ph.D. was the investigation of new biometric identifiers. The assessment of unconventional biometric identifiers is fundamental for potential breakthroughs in biometric systems' performance and/or cost. First, the concept of transient biometrics was proposed, an innovative, unconventional biometric modality. Second, a specific transient biometric trait was investigated, the fingernail. Fingernails are not a usual source of data for biometrics sytems; thus, a new set of biometric signatures for fingernail images were proposed. Third, the EEG was explored as a potential unconventional biometric identifier, where we demonstrated the potential of looming visual evoked potentials as biometrically discriminant data. Fourth, the somatotype was explored as a potential new biometric identifier, based on the RE-ID task. The proposed solution based on ANNs, achieved state-of-the-art results while generalising beyond **visual appearance cues**[†] .

[†] *Visual characteristics of attire were often the major characteristic exploited in previous RE-ID works.*

Given the novelty of the above works, their publication was challenging as one belongs, by definition, to a small non mainstream community. These works are thus normally reviewed by researchers working in main-stream biometrics. A bias then arises, as such reviewers expect to see performance comparable (at least) to established biometric identifiers. It is commonplace to make unfair comparisons of the performance of an emerging biometric, dictating that such new methods are not suitable replacements for established traits X, Y, or Z. These reviewers need to be reminded that biometrics is not a zero-sum game where one solution needs to be found to rule them all. Rather, novel biometric traits establish potential new sources of data, that can potentially mature in very robust methods later. Such novel biometric traits are also potentially ideal candidates for biometric fusion schemes, if they provide complementarity in failure cases to existing biometric traits. Biometric fusion is considered as a unique option to achieve 100% accuracy [51].

Another challenge of novel biometrics is related to data acquisition. Data for new biometrics is not ordinarily available in the form of established datasets and benchmarks. Thus, data acquisition becomes a big and expensive part of such research. In doing so, the biometrics community is given access to new data that can potentially spark a new field. Since the acquisition cost of a new biometrics dataset is high, it should be expected that such datasets are a fraction of the size of established ones. This challenge motivated us to investigate the use of synthetic datasets for biometric tasks.

The publication of a new dataset accompanied every paper of this Ph.D.

4.1.3 *The philosophical contrast of biometrics*

> *Always the eyes watching you and the voice enveloping you. Asleep or awake, working or eating, indoors or out of doors, in the bath or in bed — no escape.*
>
> *Nothing was your own except the few cubic centimetres inside your skull.*
>
> — 1984 - George Orwell [39]

As humankind developed from a small population of tribes of hunter-gatherers into larger communities of collaborative and interactive societies, identity recognition became a fundamental task. Groups of hunter-gatherers had complex dynamics, and identification of friends or foes was vital for survival; identification was, however, efficiently performed by individuals given that the size of the groups was small. Identification was solely based on a personal ability to recall all known identities from their experience and previous observation of biometric characteristics (like face, voice, body shape) or context (such as location and contacts). This is how humans still perform biometric identity association [25].

As hunter-gatherers evolved into more populous societies with complex interactions between groups, it became impossible to rely on an individual's ability to identify other subjects. Thus, at first, a substitute for the identification of friends/foes relied on a shared belief (e. g. entity/God) or a common exchange currency [20]. Then, as humankind started organizing in more prominent civilizations, the ability to identify a subject and associate it with other social constructs (e. g. Nationality) became fundamental. As the biometric identification ability of humans did not scale and was not transferable, thus evolved identity association via surrogate representations. Surrogate identity representation can be based on what a subject possesses (e. g. ID card or passport) or what they know (e. g. username and password) [25].

Automatic biometric methods scale up the original biometric characteristics used for identity association by hunter-gatherer groups

The advent of modern **automatic biometric methods** , makes it possible to perform identity association based on biometric characteristics for extensive groups. We thus rely less on surrogate identification solutions, which can be lost, forged, copied, shared, stolen, etc.

Robust identity association methods are crucial to modern society as they are fundamental for law-enforcement, international border control, access to secure sites (e. g. Nuclear powerplants, military base) and so on. Automatic biometric methods are a big step towards the goal of social safety.

The economies of scale that accompany the maturity of automated biometric methods have led to a significant reduction in the cost of capable biometric sensors, enabling the possibility of performing biometric authentication on personal devices. Currently, it is common to use fingerprints or even face recognition as a form of access control on mobile phones and computers. The widespread use of biometrics for access control of personal devices appears to be motivated by convenience rather than security concerns. Biometric authentication on these devices commonly replaces token entry but does not add a new layer of security (which would increase access control). In the case of face recognition, the average user probably does not know that information to falsify/deceive biometric authentication is publicly available [11, 58] or that in some cases there is no way to provide updates to prevent spoof attacks [9].

However, this ubiquitous adoption comes with a cultural shift: the acceptance of biometrics is increasing throughout society as a whole, which may lead to a new generation that accepts the indiscriminate use of biometrics and regards it as standard/-common. Acceptance of hard and robust biometrics is growing. For example, every day thousands of people volunteer their fingerprint information to gain access to a Disney amusement park or to purchase restricted sales items such as alcohol or tobacco without human interaction in places like Norway. Questioning of the necessity behind the use of selected biometrics does not seem to grow similarly. Is it important to be given the choice to remain anonymous? Can the loss of privacy as a result of the use of biometrics result in the reduction of an individual's ability to protest and demonstrate? Do modern biometric solutions propagate biases from their training sets (e. g.against specific ethnic groups that were not in the training set)? Do biometrics in the hands of suppressive regimes result in a tool of seggregation and crowd control?

Biometrics and its derived applications are a tool of modern society as any other tool it is unaware of any ethical implications of its use. A simplistic argument goes that the final responsibility rests solely at the hands of the agent employing such a tool. Nevertheless, as researchers, we need to ask ourselves about our responsibility and intent in developing biometric methods. To this end we need to delve into the philosophical contrasts generated by the use of biometrics.

The philosophical contrast of biometric solutions is multifaceted. Biometrics arose out of personal relations, and yet automatic biometric methods currently promote unilateral connections between people and machines. Identity association was in place to ensure individual and collective freedom from danger; automatic biometric methods can be used to isolate and track minorities, potentially placing them in greater danger. Automatic biometric methods evolved to be convenient, scalable and in some cases not even require direct collaboration; now they can also be used by powerful entities that can forcibly extract private biometric information of those who would like to remain anonymous[34].

As for any technology that has both positive and negative implications for society, the question arises whether it is beneficial, or not, to perform public domain research on such a technology. Is it better to not do any research and assume that the current state-of-art of biometric solutions will remain unchanged? I believe this is not a realistic option: biometrics is woven into our societal fabric, and there is no control of research studies driven by valid or spurious motivations. Is it better to perform such research only by private entities that can invest in it, or should it be publicly driven in the hope that public knowledge is the best tool a society can have to establish a fair system? Since there is nothing to stop private entities to develop biometric solutions for spurious motivations, I believe that it is better to have similar research results in the public domain in order to understand their capabilities and pitfalls, thus enabling society to better understand and regulate them.

Part II

APPENDIX

I

NEURAL NETWORK ACTIVATION VISUALIZATION

Loss function optimization can lead to local minima that represent the path of least resistance to the optimization function. Therefore, image based ANNs can learn classification based on a biased representation. One example is a CNN that learns that every image with snow is from the category of skiing (because the dataset only has images of snow when dealing with the class skiing - dataset representation bias). One way to investigate the representation learned by an ANN is to identify which neurons respond to properties of interest and visualize which pixels of the source image are responsible for such response in different layers. For example, is the ANN responding to the parameters that we assume (somatotype, identity) or is it rather responding to other parameters, such as apparel? Is the network responding to variations in gender or obesity? For Publication D [3], the ability to probe the neural network for specialized neurons and to verify expected behavior allowed us to tune the data generation pipeline.

Our focus is neuron activation for a given pixel $\mathcal{X}_i$ of an input image $\mathbf{X}$ where $\{\mathbf{X} \in \mathbb{N}^{w \times h \times 3} \mid 0 \leqslant \mathcal{X}_i \leqslant 255\}$. Thus the established visualization techniques that provide a class model visualization [47], or select which input image would activate a neuron the most, or direct visualization activation [60] are not suitable solutions. The most suitable alternatives were based on image saliency visualization [47].

Saliency visualization is a guided gradient backpropagation visualization technique, which relies on the backpropagation of a neuron activation all the way to the input image. We implemented a variation of this technique that allowed us to distill a better-localized response for our data. Previous visualization techniques had too many high-frequency artifacts or hid important details by not showing the activation on the original colorspace. These issues motivated our novel neuron visualization approach.

Our first synthetic dataset, named *pre-soma* which is unpublished (see more details in Section 3.3.1.3 and in Figure 3.8), had a higher out-of-class similarity when compared to the normally visualized ILSVRC-2013 dataset[44], Figure I.1.

The proposed visualization uses an input image $\mathbf{X}$ and shows by how much, and which pixels, affect any given neuron. We propose modifications to the saliency map visualization of [47] that are less affected by high-frequency artifacts. We present the saliency maps in the original image colorspace. The result is a more straightforward neuron visualization.

Given the input image $\mathbf{X}$ and the activation of a probed neuron $A_{L,N}$, where L denotes the neuron layer and N denotes the neuron index within the layer. One can calculate the derivative of $A_{L,N}$ with respect to the image $\mathbf{X}$ using backpropagation algorithms. The magnitude of the derivative in the input space can be used to indicate which pixels contribute the most for the given activation.

Equation I.1 shows how the Saliency map $\mathcal{S}$ is derived: the gradient of the activation $A_{L,N}$, with respect to the image $\mathbf{X}$, is computed; the gradient is then projected from the three RGB channels to a single channel input space by the function $\phi(\cdot)$, which computes the L2 norm of the three color channels for each pixel. The salience $\mathcal{S}$ is finally computed after $\phi(\cdot)$ is convolved with a Gaussian kernel $\mathcal{N}$ with a square window of size $\mathbf{W}$. This convolution operation works as a further regularization tool [60].

$$\mathcal{S} = \phi\left(\frac{\partial A_{L,N}(\mathbf{X})}{\partial \mathbf{X}}\right) * \mathcal{N}(\mathbf{W}) \tag{I.1}$$

Ideally, one would perform regularization for every derived layer inside the backpropagation chain, but this comes at a very high implementation cost. Instead, [60] proposes to convolve $\phi(\cdot)$ with a Gaussian kernel $\mathcal{N}$, using a *large* size $\mathbf{W}$, directly in the input space. We have taken an approach that tries to give better localization using a small window and less aggressive

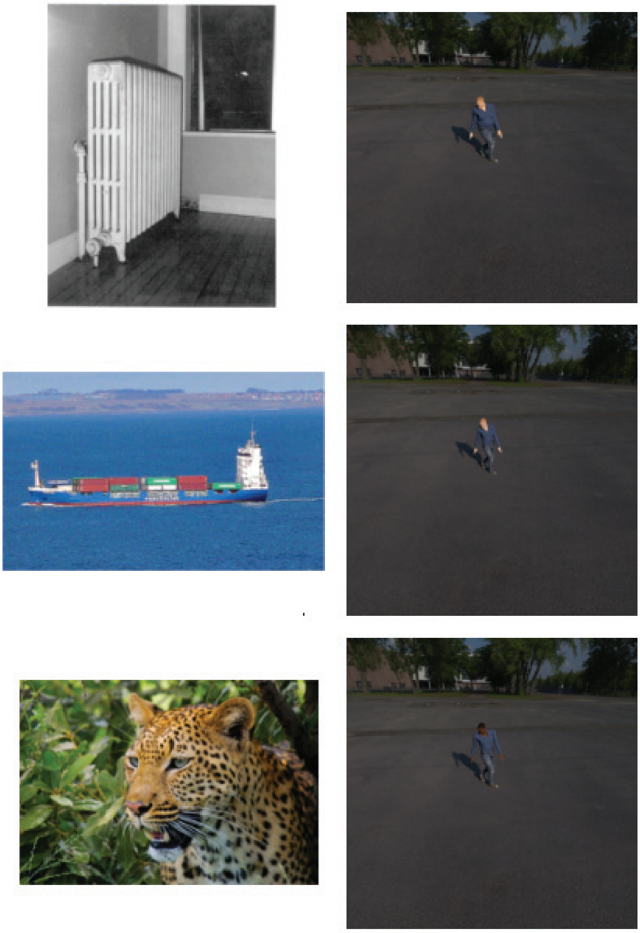


Figure I.1: Sample images showing out-of-class variance. The first column shows three classes from ILSVRC-2013 dataset [44]. These are radiator, container-ship and leopard—the second column shows three classes from the *pre-soma* dataset. One can notice that there is sizeable outer-class variance in ILSVRC-2013 compared to data from the *pre-soma* dataset. Small out-of-class similarity holds for any vision-based biometric classification

regularization (previously performed by Gaussian kernel convolution). The proposed approach consists of three steps:

STEP 1: COMPUTE $\mathcal{S}$ USING A *small* WINDOW SIZE $\mathbf{W}$.

The first step of the proposed visualization approach is implemented by Eq. I.1 with a small window size. The window size $\mathbf{W}$ was defined to be the same as the largest convolution window used in the network. The smaller $\mathbf{W}$ is, the sharper the response.

STEP 2: LAYER BASED CONTRAST ENHANCEMENT.

Do contrast enhancement based on the ratio of activation values in the probed layer L. For the implementation of this step we compute a mask $\mathcal{M} \in \mathbb{N}^{w \times h}$, where w, h are the input image width and height respectively, according to Eq. I.2. The mask performs a contrast enhancement operation that depends on the ratio of the maximum activation value of the probed neuron $(A_{L,N})$ to the maximum activation value across all neurons of its layer, $max(A_L)$. $\mathcal{M}$ also depends on γ, a user-defined parameter that can be tuned to ensure visualization of very deep networks. The γ parameter is tunable by observing the resulting visualization of the final layer. One should increase γ from zero upwards until the neuron with the highest activation produces a visible effect in the final visualization. Tunning of γ is crucial because the saliency maps for deep layers of well-trained ANNs can have very small values; values higher than 255 are clamped.

STEP 3: FILTER OUT HIGH-FREQUENCY NOISE.

The third step is accomplished by a median filter operation
$med(\circ, \mathbf{W})$ where $\mathbf{W}$ is the previously defined filter window size. This operation removes single-pixel activations (noise) and further reduces the high-frequency components in the computed mask.

$$\mathcal{M} = med\left(min\left(\left[\mathcal{S} \cdot \gamma \cdot e^{\frac{max(A_{L,N})}{max(A_L)} - 1} \right], 255 \right), \mathbf{W} \right) \qquad \text{(I.2)}$$

The final visualization is achieved by setting values of the mask $\mathcal{M}$ as the alpha (transparency) channel for the the input image $\mathbf{X}$. Areas with zero alpha are represented as white pixels.

This visualization encodes two pieces of information: it shows which pixels of $\mathbf{X}$ influence the activation of the probed neuron $A_{L,P}$ the most; the contrast of the color components allows the comparison of the activation values across the neurons of layer L. Visualization results are shown in Section 3.3.1.

BIBLIOGRAPHY

[1] Seth B. Agyei, Magnus Holth, F. Ruud van der Weel, and Audrey L. H. van der Meer. "Longitudinal study of perception of structured optic flow and random visual motion in infants using high-density EEG." In: *Developmental Science* (2014). ISSN: 1467-7687. DOI: 10.1111/desc.12221. URL: http://dx.doi.org/10.1111/desc.12221.

[2] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. "Face description with local binary patterns: Application to face recognition." In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12 (2006), pp. 2037–2041.

[3] Igor Barros Barbosa, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen, and Theoharis Theoharis. "Looking beyond appearances: Synthetic training data for deep CNNs in re-identification." In: *Computer Vision and Image Understanding* 167 (2018), pp. 50 –62. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2017.12.002. URL: http://www.sciencedirect.com/science/article/pii/S1077314217302254.

[4] Igor Barros Barbosa, Marco Cristani, Alessio Del Bue, Loris Bazzani, and Vittorio Murino. "Re-identification with rgb-d sensors." In: *Proc. ECCV - Workshops and Demonstrations*. 2012.

[5] Igor Barros Barbosa, Theoharis Theoharis, Christian Schellewald, and Cham Athwal. "Transient biometrics using finger nails." In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. Sept. 2013, pp. 1–6. DOI: 10.1109/BTAS.2013.6712730.

[6] Igor Barros Barbosa, Kenneth Vilhelmsen, Audrey van der Meer, Ruud van der Weel, and Theoharis Theoharis. "EEG Biometrics: On the Use of Occipital Cortex Based Features from Visual Evoked Potentials." In: *28th Norsk Informatikkonferanse, NIK 2015, Høgskolen i Ålesund*. Bibsys Open Journal Systems, Norway, Nov. 2015. URL: http://ojs.bibsys.no/index.php/NIK/article/view/243.

[7] Igor Barros Barbosa, Theoharis Theoharis, and Ali E. Abdallah. “On the use of fingernail images as transient biometric identifiers.” In: *Machine Vision and Applications* 27.1 (Jan. 2016), pp. 65–76. ISSN: 1432-1769. DOI: 10.1007/s00138-015-0721-y. URL: https://doi.org/10.1007/s00138-015-0721-y.

[8] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior. “The relation between the ROC curve and the CMC.” In: *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05)*. Oct. 2005, pp. 15–20. DOI: 10.1109/AUTOID.2005.48.

[9] Andrew Bud. “Facing the future: the impact of Apple FaceID.” In: *Biometric Technology Today* 2018.1 (2018), pp. 5 –7. ISSN: 0969-4765. DOI: https://doi.org/10.1016/S0969-4765(18)30010-9. URL: http://www.sciencedirect.com/science/article/pii/S0969476518300109.

[10] J. Canny. “A Computational Approach to Edge Detection.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (1986), pp. 679–698.

[11] Bkav Corporation. *Bkav's new mask beats Face ID in "twin way": Severity level raised, do not use Face ID in business transactions*. 2017. URL: https://www.bkav.com/top-news/-/view-content/65202/bkav-s-new-mask-beats-face-id-in-twin-way-severity-level-raised-do-not-use-face-id-in-business-transactions (visited on 12/09/2019).

[12] Abir Das, Anirban Chakraborty, and Amit K Roy-Chowdhury. “Consistent Re-identification in a Camera Network.” In: *Proc. ECCV*. 2014.

[13] Marcos Del Pozo-Banos, Jesús B. Alonso, Jaime R. Ticay-Rivas, and Carlos M. Travieso. “Electroencephalogram subject identification: A review.” In: *Expert Systems with Applications* 41 (2014), pp. 6537–6554. ISSN: 09574174. DOI: 10.1016/j.eswa.2014.05.013. URL: http://dx.doi.org/10.1016/j.eswa.2014.05.013.

[14] S. Easwaramoorthy, F. Sophia, and A. Prathik. “Biometric authentication using finger nails.” In: *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*. Feb. 2016, pp. 1–6. DOI: 10.1109/ICETETS.2016.7603054.

[15] Martin A. Fischler and Robert C. Bolles. “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.” In: *Commun. ACM* 24.6 (June 1981), pp. 381–395. ISSN: 0001-0782.

[16] Computer Vision Foundation. *CVPR 2019 - Welcome Slides*. 2019. URL: http://cvpr2019.thecvf.com/files/CVPR2019-WelcomeSlidesFinal.pdf (visited on 12/09/2019).

[17] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. “Synthetic data augmentation using GAN for improved liver lesion classification.” In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Apr. 2018, pp. 289–293. DOI: 10.1109/ISBI.2018.8363576.

[18] Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy. *Person re-identification*. Vol. 1. Springer, 2014.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Nets.” In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger. Curran Associates, Inc., 2014, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[20] Yuval Noah Harari. *Sapiens: A Brief History of Humankind*. Harper, Feb. 2015. ISBN: 0062316095. URL: https://www.xarg.org/ref/a/0062316095/.

[21] K. He, X. Zhang, S. Ren, and J. Sun. “Deep Residual Learning for Image Recognition.” In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.

[22] Alexander Hermans, Lucas Beyer, and Bastian Leibe. “In defense of the triplet loss for person re-identification.” In: *arXiv preprint arXiv:1703.07737* (2017).

[23] Olivier Huynh and Bogdan Stanciulescu. “Person re-identification using the silhouette shape described by a point distribution model.” In: *Proc. WACV*. 2015.

[24] Anil K. Jain, Patrick Flynn, and Arun A. Ross. *Handbook of Biometrics*. Berlin, Heidelberg: Springer-Verlag, 2007. ISBN: 038771040X.

[25] Anil Jain, Ruud Bolle, and Sharath Pankanti. *Introduction to biometrics*. Springer, 1996, pp. 1–41.

[26] K.V. Kale, Y.S. Rode, M.M. Kazi, S.B. Dabhade, and S.V. Chavan. “Multimodal Biometric System Using Fingernail and Finger Knuckle.” In: *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*. Aug. 2013, pp. 279–283. DOI: 10.1109/ISCBI.2013.63.

[27] R.V. Krstic. *Human Microscopic Anatomy: An Atlas for Students of Medicine and Biology*. Springer, 1991. ISBN: 9783540536666.

[28] Amioy Kumar, Shruti Garg, and M. Hanmandlu. “Biometric authentication using finger nail plates.” In: *Expert Systems with Applications* 41.2 (2014), pp. 373 –386. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2013.07.057.

[29] Shih Hsiung Lee and Chu Sing Yang. “Fingernail analysis management system using microscopy sensor and blockchain technology.” In: *International Journal of Distributed Sensor Networks* 14.3 (2018), p. 1550147718767044.

[30] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. “BRISK: Binary Robust invariant scalable keypoints.” In: *Computer Vision, IEEE International Conference on* 0 (2011), pp. 2548–2555. DOI: http://doi.ieeecomputersociety.org/10.1109/ICCV.2011.6126542.

[31] W. Li, R. Zhao, T. Xiao, and X. Wang. “DeepReID: Deep Filter Pairing Neural Network for Person Re-identification.” In: *Proc. CVPR*. 2014.

[32] R. Lienhart and J. Maydt. “An extended set of Haar-like features for rapid object detection.” In: *Image Processing. 2002. Proceedings. 2002 International Conference on*. Vol. 1. 2002, I–900–I–903 vol.1. DOI: 10.1109/ICIP.2002.1038171.

[33] DavidG. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints.” English. In: *International Journal of Computer Vision* 60.2 (2004), pp. 91–110. ISSN: 0920-5691. DOI: 10.1023/B:VISI.0000029664.99615.94.

[34] Avi Marciano. “Reframing biometric surveillance: from a means of inspection to a form of control.” In: *Ethics and Information Technology* 21.2 (June 2019), pp. 127–136. ISSN: 1572-8439. DOI: 10.1007/s10676-018-9493-1. URL: https://doi.org/10.1007/s10676-018-9493-1.

[35] Audrey L H van der Meer, Monica Svantesson, and F.R. Ruud van der Weel. "Longitudinal Study of Looming in Infants with High-Density EEG." In: *Developmental Neuroscience* 34.6 (2012), pp. 488–501. DOI: 10.1159/000345154. URL: http://dx.doi.org/10.1159/000345154.

[36] Marius Muja and David G. Lowe. "Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration." In: *International Conference on Computer Vision Theory and Application VISSAPP'09)*. INSTICC Press, 2009, pp. 331–340.

[37] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam. "Survey on LBP based texture descriptors for image classification." In: *Expert Systems with Applications* 39.3 (2012), pp. 3634–3641.

[38] Timo Ojala, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions." In: *Pattern recognition* 29.1 (1996), pp. 51–59.

[39] George Orwell. *1984*. Centennial. Tandem Library, 1950. ISBN: 0881030368. URL: http://www.amazon.de/1984-Signet-Classics-George-Orwell/dp/0881030368.

[40] Ramaswamy Palaniappan. "Method of identifying individuals using VEP signals and neural network." In: *IEE Proceedings-Science, Measurement and Technology* 151.1 (2004), pp. 16–20.

[41] General Data Protection Regulation. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46." In: *Official Journal of the European Union (OJ)* 59.1-88 (2016), p. 294.

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4.

[43] Jeffrey Rosen. "The right to be forgotten." In: *Stan. L. Rev. Online* 64 (2011), p. 88.

[44] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge." In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: 10.1007/s11263-015-0816-y.

[45] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. "Improved Techniques for Training GANs." In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16. Barcelona, Spain: Curran Associates Inc., 2016, pp. 2234–2242. ISBN: 978-1-5108-3881-9. URL: http://dl.acm.org/citation.cfm?id=3157096.3157346.

[46] J. Shi and C. Tomasi. "Good features to track." In: *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*. June 1994, pp. 593–600. DOI: 10.1109/CVPR.1994.323794.

[47] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps." In: *CoRR* abs/1312.6034 (2013). URL: http://arxiv.org/abs/1312.6034.

[48] Xiaoxiao Sun and Liang Zheng. "Dissecting Person Re-Identification From the Viewpoint of Viewpoint." In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.

[49] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going Deeper With Convolutions." In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2015.

[50] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. "Rethinking the Inception Architecture for Computer Vision." In: *Proc. CVPR*. 2016.

[51] Theoharis Theoharis, Georgios Passalis, George Toderici, and Ioannis A. Kakadiaris. "Unified 3D face and ear recognition using wavelets on geometry images." In: *Pattern Recognition* 41.3 (2008). Part Special issue: Feature Generation and Machine Learning for Robust Multimodal Biometrics, pp. 796 –804. ISSN: 0031-3203. DOI: https://doi.

org/10.1016/j.patcog.2007.06.024. URL: http://www.sciencedirect.com/science/article/pii/S0031320307003214.

[52] Allen Topping, Vladimir Kuperschmidt, and Austin Gormley. *United States Patent US005751835A*. 1998.

[53] Bram Van Ginneken, Alejandro F Frangi, Joes J Staal, Bart M ter Haar Romeny, and Max A Viergever. "Active shape model segmentation with optimal features." In: *IEEE transactions on medical imaging* 21.8 (2002), pp. 924–933.

[54] Gul Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. "BodyNet: Volumetric Inference of 3D Human Body Shapes." In: *The European Conference on Computer Vision (ECCV)*. Sept. 2018.

[55] P. Viola and M. Jones. "Rapid object detection using a boosted cascade of simple features." In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. 2001, I–511–I–518 vol.1. DOI: 10.1109/CVPR.2001.990517.

[56] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang. "Residual Attention Network for Image Classification." In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 6450–6458.

[57] F. Ruud van der Weel and Audrey L. H. van der Meer. "Seeing it coming: infants' brain responses to looming danger." English. In: *Naturwissenschaften* 96.12 (2009), pp. 1385–1391. ISSN: 0028-1042. DOI: 10.1007/s00114-009-0585-y. URL: http://dx.doi.org/10.1007/s00114-009-0585-y.

[58] Yi Xu, True Price, Jan-Michael Frahm, and Fabian Monrose. "Virtual U: Defeating Face Liveness Detection by Building Virtual Models from Your Public Photos." In: *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 497–512. ISBN: 978-1-931971-32-4.

[59] S Yaemsiri, N Hou, MM Slining, and K He. "Growth rate of human fingernails and toenails in healthy American young adults." In: *Journal of the European Academy of Dermatology and Venereology* 24.4 (2010), pp. 420–423.

[60] Jason Yosinski, Jeff Clune, Anh Mai Nguyen, Thomas Fuchs, and Hod Lipson. "Understanding Neural Networks Through Deep Visualization." In: *CoRR* abs/1506.06579 (2015). URL: http://arxiv.org/abs/1506.06579.

[61] X L Zhang, H Begleiter, B Porjesz, W Wang, and A Litke. "Event related potentials during object recognition tasks." In: *Brain Research Bulletin* 38 (1995), pp. 531–538.

[62] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. "Scalable Person Re-identification: A Benchmark." In: *Proc. ICCV*. 2015.

[63] Zhedong Zheng, Liang Zheng, and Yi Yang. "Unlabeled Samples Generated by GAN Improve the Person Re-Identification Baseline in Vitro." In: *The IEEE International Conference on Computer Vision (ICCV)*. Oct. 2017.

Part III

PUBLICATIONS

PUBLICATIONS

A TRANSIENT BIOMETRICS USING FINGER NAILS

Igor Barros Barbosa, Theoharis Theoharis, Christian Schellewald, and Cham Athwal. “Transient biometrics using finger nails.” In: *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. Sept. 2013, pp. 1–6. DOI: 10.1109/BTAS.2013.6712730

Transient Biometrics using Finger Nails

Igor Barros Barbosa Theoharis Theoharis Christian Schellewald
Department of Computer and Information Science
Norwegian University of Science and Technology

Cham Athwal
School of Digital Media Technology
Birmingham City University

Abstract

Transient biometrics, a new concept for biometric recognition, is introduced in this paper. A traditional perspective of biometric recognition systems concentrates on biometric characteristics that are as constant as possible (such as the eye retina), giving accuracy over time but at the same time resulting in resistance to their use for non-critical applications due to the possibility of misuse. In contrast, transient biometrics is based on biometric characteristics that do change over time aiming at increased acceptance in non-critical applications. We show that the fingernail is a transient biometric with a lifetime of approximately two months. Our evaluation datasets are available to the research community.

1. Introduction

Biometric recognition systems offer unique advantage when compared to conventional recognition systems, such as smart cards or passwords. By using a biometric recognition system, the subject does not need to carry or remember any id or password, and there is less risk of loss or disclosure of the recognition token. Biometric recognition is thus gaining support and acceptance in critical recognition situations supported by governments or other large organizations.

Despite the advantages of biometric recognition systems, a major concern of individuals is the possibility of misuse of their biometric data. A card or password can be canceled, but what happens if your biometric data falls into the wrong hands? An individual's *privacy* may be compromised (e.g. through their use for unauthorized recognition purposes) or *discrimination* may be enabled (e.g. through unauthorized use by insurance agents).

Cancelable biometrics [7,8] attempts to answer this concern by pre-transforming (distorting) the biometric data before the biometric signature is extracted. The transformation is non-reversible. Thus, the potential for misuse is limited by the fact that the misuser cannot retrieve the original biometric data, and the transformation can be changed at any time. However, *cancelable* biometrics requires that the subject trusts the biometrics capture point and also that the misuse is detected in order to activate a transform change.

There is plenty of scope for biometric recognition systems to become more *socially acceptable*, in the sense that society could accept and use such systems in day-to-day scenarios. The acceptability issue remains particularly open when dealing with non-critical scenarios and collaborative subjects. For instance, individuals will not happily offer their fingerprints just to have access to their hotel room. The points raised above limits the use of biometric technologies in a multitude of noncritical situations.

In this paper we introduce *transient* biometrics. *Transient* biometrics is defined as biometric recognition technologies which rely on biometric characteristics that are proven to change over time. Thus, they automatically cancel themselves out after a known period of time. A transient biometric approach for the verification task is shown in Fig. 1. In contrast to *cancelable* biometrics, it is the actual biometric data that are naturally changing over time. As a consequence it will presumptively help in the creation of more sociable acceptable recognitions systems. We show that images of the finger nail constitute a *transient* biometric with a lifetime of two months.

The remaining of the paper is organised as follows. Section 2 briefly presents the biometric literature which employs finger nails. Section 3 details our approach followed by Section 4 that shows experimental results. Finally, Section 5 concludes the paper, envisaging some future perspectives.

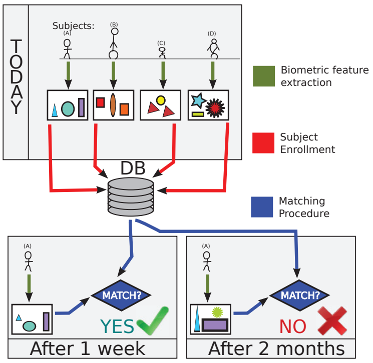


Figure 1. Example of a verification task employing *transient* biometrics.

2. Previous Work

The use of finger nails in biometrics applications has been the topic of a few different lines of research. A complex acquisition system employing tailored lighting equipment has been designed to acquire images of the nail bed, which is the skin under the nail plate [9]. Such images are then used for individual authentication by exploring features from the nail bed grooves. This is possible because the nail bed is unique to each individual [3].

Recently, a *cancelable* biometric approach has developed stickers, which can be placed over finger nails for an identification process [4]. In the presented paper, thumb images were acquired against a black background, for it made it easier to compute the boundaries of the thumb. The stickers glued over the fingernails provide two landmarks which are used in the feature extraction process. The finger outline is pursued, and the distance from the outline to the landmarks creates a distance profile. The final matching procedure is done by computing correlation coefficients between distance profiles. Whenever such sticker is removed or repositioned over the finger nail, the distance profile changes. Therefore, it is possible to cancel the biometric by replacing the landmarks or by shifting the sticker position. The finger nail would only play a role in the recognition system if the nail outgrew the finger. Although it is a valid *cancelable* biometric approach, this work does not truly explore finger nails for biometric recognition.

The information of the nail surface has been explored by a biometric authentication system proposed in [2]. This approach uses images of hands in order to extract information from the finger nails. First, the fingers are segmented by a smart contour segmentation algorithm, then the nails are segmented by grey scale thresholding. This simple segmentation approach is likely to work given that the employed dataset was biased with respect to subject's skin tones. The individual authentication process is built upon the hamming distance of high frequency Haar wavelet coefficients. Experimental results show reasonable recognition rates using three sample images per subject. The first two images are employed in training while the last image is used for testing. Despite the positive results, this work does not explore how the recognition rate behaves with respect to the growth of the finger nails; the authors do not provide the time difference between acquisitions, so we assume that all images were acquired on the same date.

3. The Proposed Approach

The *transient* biometric approach presented in this work addresses the identification problem. Therefore, our objective is to identify a subject by comparing a biometric signature against a dataset of previously collected samples. To this end, the proposed solution can be divided into three phases. The first phase deals with image segmentation and pre-processing. The second phase extracts the biometric signature, while the third phase addresses signature matching.

3.1. Nail image pre-processing

Images of the right index finger are the source of biometric information. Since the images were taken on different days and sometimes with different cameras, the pre-processing of input images is a key step for the overall process. It assures that the images delivered to the signature extraction algorithm fulfil the requisites regarding colour correction, nail plate registration and image size.

Pre-processing starts by segmenting the nail from the finger images. Such segmentation is done by an active shape model (*ASM*) see [11]. The active shape model requires a set of training images where the segmentation has been manually performed (contour drawn). The algorithm employs Principal Component Analysis (*PCA*) to find eigen segmentation contours, with very accurate results. The *ASM* also describes the image around each control point with a grey-level appearance model. This grey-level appearance model is computed using lines perpendicular to each control point, and it is built using the first derivative of grey-level images. This appearance model will be used later in an iterative fashion to correct the position of control points while searching for the best segmentation contour.

The *ASM* requires training data and for this purpose the dataset ***D01*** was used. This training dataset represents the first acquisition day. It contains an image of the right index finger for every enrolled individual, with a total of 32 images (see more information on the dataset in Sec. 4). The ASM was trained with two main landmarks and 20 control points between them. The first landmark is placed at the base of the nail plate just by the intersection with the finger skin. Meanwhile, the other landmark is placed opposite to it, by the end of the nail plate. An input sample is shown as [A] in Fig. 2, and the resulting segmentation is shown as [B].

Image pre-processing continues by computing the bounding box. Next, the bounding box is converted to grey scale, making the input more robust to changes. These changes are likely to happen due to wrong white balance or even due to the use of different cameras. The overall pre-processed image is given by resizing the bounded box to a width and height of 128 pixels. The resulting image is shown as [C] in Fig. 2

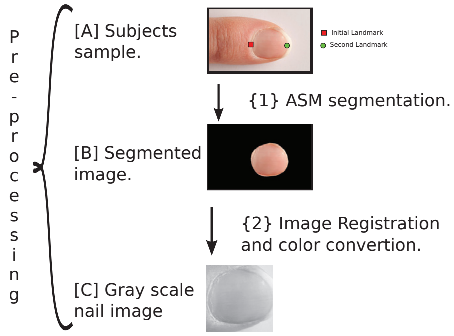


Figure 2. Sample results from the image pre-processing pipeline.

To make the training process more robust each image is used to create multiple variations. These are given through the application of Wiener Filters, by shifting the segmented region-of-interest by a few pixels and by the application of histogram equalisation. When all these images modification are combined in a chain, every input image generates 810 variations.

3.2. Signature extraction based on uniform LBP

It has been observed that a nail plate is categorised by a unique texture which is influenced by patterns in the nail bed [3]. The nail plate texture is also dependent on interaction with external factors. Hence, it is common to notice white spots and marks originating from scratches or bumps. Since the nail plate possesses such rich texture, we have opted to base signature extraction on Local Binary Patterns (*LBP*), which is a successful and robust texture descriptor [5]. *LBPs* are known for their computational efficiency and their capacity to discriminate micro-patterns. They have also been successfully employed in a wide variety of applications, ranging from texture classification [6] (their original purpose), to facial recognition [1]. Thus, *LBP* has been selected for the signature extraction process.

LBP are computed pixel wise, relying on the pixel neighbourhood information. The computation starts by defining a neighbouring circle with a radius of R pixels and P evenly spaced sample points. Bilinear interpolation is used to compute the value of a sample point if it does not fall on a pixel center. Fig. 3 illustrates two possible circular neighbourhoods. *LBP* is computed for the pixel g_c, located in the center of the circle, using the threshold operation of Eq. 1.

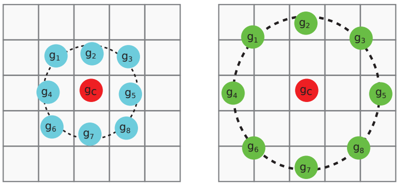


Figure 3. Sample neighbourhoods of $(P, R) = (8, 1)$ and $(P, R) = (8, 2)$.In these examples g_p are the sample points, where p ranges from 1 to P.

$$LPB_{P,R} = \sum_{p=1}^{P} \phi\left(g_p - g_c\right) \times 2^{p-1}$$

$$\phi\left(x\right) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \tag{1}$$

An *LBP* is uniform whenever the coded value is composed of zero, one or two bit-wise transitions. Thus, the patterns 11111111 and 00001111 are uniform since they have zero and two bit-wise transitions, respectively. On the other hand, the pattern 10101010 is non-uniform since it is composed of eight transitions. If a coded pixel uses eight sample points, it is possible to generate 256 patterns, out of which 58 are called uniform LBP patterns. The work of [1,6] confirms that uniform Local Binary Patterns (LBP^{u2}) account for the vast majority of encountered patterns. Therefore, signature extraction employs uniform *LBP* for describing the nail plate texture.

The signature extraction process starts by dividing the pre-processed nail plate image into smaller image blocks. A 4 × 4 grid is used for the division, generating 16 blocks of 32 × 32 pixels. A histogram of the values of $LBP^{u2}_{8,2}$ is then computed for each block. The histogram is composed of 59 bins, 58 of them used for uniform patterns and the last bin for non-uniform ones. The signature is then created by concatenating the 16 histograms, thus forming a global descriptor of the nail plate. This process is illustrated in Fig. 4 and follows the methodology proposed in [1]. A descriptor capable of describing texture and its spatial relationships is thus created, which is very suitable for the nail plate. However, the resulting signature has 944 features. The dimensionality curse is avoided by the use of Karhunen-Loeve

Analysis. The data from ***D01*** (see Sec. 4), is analysed by Karhunen-Loeve decomposition which allows us to map the extracted signature into a subspace of 200 dimensions.

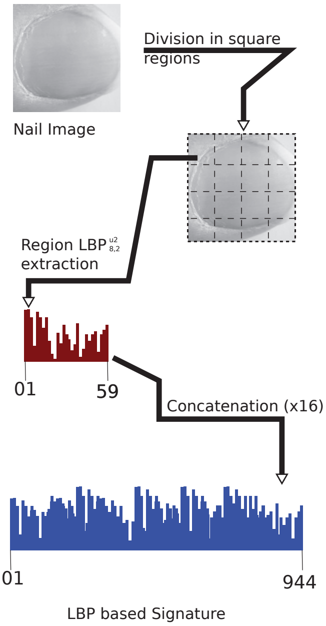

Nail Image
Division in square regions
Region LBP$^{u2}_{8,2}$ extraction
01
59
Concatenation (x16)
01
944
LBP based Signature

Figure 4. Signature extraction pipeline

3.3. Signature matching

Signature matching essentially identifies patterns within a dataset. These patterns should ideally have small variation between objects of the same class (e.g. nail images of the same subject) while they should have large variation across different classes. To identify such patterns Bayesian classification is employed. A Bayesian classifier estimates the boundaries between classes assuring that the Bayes risk/error is minimal. The Bayes rule states that the probability of a subject belonging to a class ω_k given an observation $\boldsymbol{z}$ (signature after dimensionality reduction) is given by the *posterior probability* as shown in Eq. 2.

$$P\left(\omega_k|\boldsymbol{z}\right) = \frac{p\left(\boldsymbol{z}|\omega_k\right)P(\omega_k)}{p(\boldsymbol{z})}. \tag{2}$$

where $P\left(\omega_k|\boldsymbol{z}\right)$ is the *posterior probability*, $p\left(\boldsymbol{z}|\omega_k\right)$ is the probability distribution of $\boldsymbol{z}$ coming from a subject with known class ω_k, $P(\omega_k)$ is the prior probability of having the class ω_k and $p(\boldsymbol{z})$ is the distribution of the observation. If a unitary cost is assumed for every wrong classification, the minimisation of the Bayes risk becomes equivalent to the maximisation of posterior probability [10]. Therefore, the classifier can be re-written as

$$\overset{\wedge}{\omega}_{map}\left(\boldsymbol{z}\right) = \underset{j}{argmax}\left\{p\left(\boldsymbol{z}|\omega_j\right)P(\omega_j)\right\}. \tag{3}$$

Our assumption is that the conditional probability density function, $p\left(\boldsymbol{z}|\omega_j\right)$, can be modelled as normal. Therefore the observations are assumed to have an expectation vector $\boldsymbol{\mu}_k$ and a covariance matrix $\boldsymbol{C}_k$, yielding to the function shown in Eq. 4

$$p\left(\boldsymbol{z}|\omega_j\right) = \frac{1}{\sqrt{(2\pi)^N|C_k|}}e^{\left(\frac{(\boldsymbol{z}\quad\boldsymbol{\mu}_k)^T\boldsymbol{C}_k{}^{1}(\boldsymbol{z}\quad\boldsymbol{\mu}_k)}{2}\right)} \tag{4}$$

Such a conditional probability density function results in a quadratic classifier. It was determined experimentally that if we assume that the covariance does not depend on the class, *e.g.* $\boldsymbol{C}_k = C$ for all possible classes, we end up with a linear Bayes normal classifier which outperformes the quadratic classifier.

The linear Bayes normal classifier is them applied to every computed $\boldsymbol{z}$ and the final classification is given by the conjunction of the 810 images which were artificially created. To get the final classification of each input image, a final distance measure is created by Eq. 5.

$$D\left(\boldsymbol{Z}^{810},\omega_k\right) = \sum_{n=1}^{810}\left|\log_e c\left(\boldsymbol{Z}_n^{810},\omega_k\right)\right| \tag{5}$$

where $\boldsymbol{Z}^{810}$ represents the conjunction of 810 observations derived from a single input image, $c\left(\boldsymbol{Z}_n^{810},\omega_k\right)$ represents the confidence of the nth observation of $\boldsymbol{Z}$ being from class ω_k. When this distance measure is employed, the most probable class is given by the smallest computed distance.

4. Experiments

This section will describe the experimental dataset and the effect of nail plate growth on identification performance.

4.1. Dataset Creation

Our dataset is composed of three different sets of data. All three sets have followed the same acquisition process. First the right index finger of the subject is placed over a white sheet of paper, which is supported by a flat surface. The finger is placed over this surface without putting pressure against it, as pressure changes the colour of the finger nail. Then a diffuse light is placed so the light source points to the top of the finger. Therefore, the finger is virtually pointing to the light source. Such lighting condition avoid highlights and help achieve proper exposure. Finally the image is acquired by framing only the contents of the white paper and maintaining proper focus on the finger nail plate.

Two things differentiate the three sets of data: acquisition date and number of subjects. Set ***D01*** consists of the collection of images from the first acquisition day. Not surprisingly, this represents the largest set in terms of number of subjects, consisting of data from 32 individuals. The second set ***D08*** is composed of images acquired seven days after the initial acquisition. This set is composed of 24 individuals who were also part of ***D01***. The third and smallest set, ***D70***, contains images of 17 subjects, all part of ***D08***. The images of this third set where acquired seventy days after the initial acquisition day. Figure 5 shows samples of four individuals who were represented in all three datasets. The nail dataset and its future extensions will be made available at NTNU's Visual Computing group website [http://www.idi.ntnu.no/grupper/vis/].

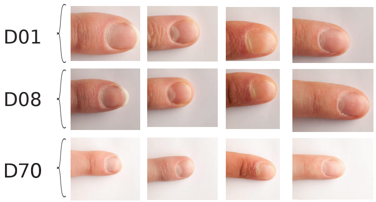


Figure 5. Images samples from available datasets. Each Row represents a dataset while each column represents a different subject.

4.2. Identification performance analysis

In all experiments the classifier was trained using only information from ***D01***. The classifier was then applied to ***D08*** and ***D70*** to evaluate the decay of identification performance, as expected for a *transient* biometric solution. As ***D70*** contains only 17 subjects, the classifier K_{17} is trained on ***D01*** using only the information from subjects available in ***D70***. The classifier is then applied to ***D08*** and ***D70***. The cumulative matching curve (*CMC*) is used as a standardised evaluation graph. It assesses the classification performance in identification problems. *CMC* models the probability of a signature from a test dataset, in this case ***D08*** and ***D70***, being correctly matched in the first P ranked subjects from the training dataset ***D01***. Such rank is derived from the computed distances, as specified in Section 3.3. The CMC curve for both ***D08*** and ***D70*** are plotted in Fig. 6. The performance decay observed in ***D70*** is evidence that the biometric signature extracted from the nail plate biometric is of short persistence. Thus the nail plate is a good candidate for *transient* biometric solutions. If the normalised area under the curve is taken as a performance measure, the changes in the nail plate during the 62 days between the acquisition of ***D70*** and ***D08*** account for a 9.32% decay. If rank one recognition is taken as a measure of performance, the results are even more conclusive: the two month interval represents a decay of 58.82% in the probability of identifying the individual in a first guess.

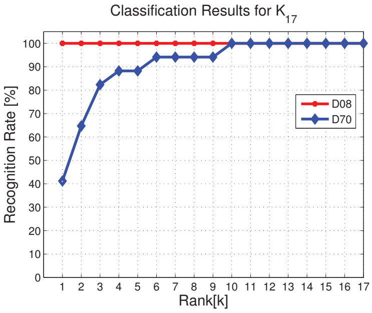


Figure 6. Cumulative matching curves for classifier K_{17} evaluated on ***D08*** and ***D70***. The classifier was trained for 17 subjects and then applied to images acquired 8 days later (D08) and 70 days later (D70). The decay in performance from D08 to D70 arguments in favour that finger nails are *transient* biometrics. Therefore, the biometric information changes during time, making the identification process unreliable after two months.

A second classifier K_{24} was trained on ***D01*** using only the information from subjects available in ***D08***. This classifier is composed of 24 subjects and represents a harder classification task than the one K_{17} was assigned. The objective was to show that positive identification is possible with nail biometrics over a short period of time. The classification results are shown as a *CMC* curve in Fig. 7.

Finally, Table 1 summarises the results of the three classification problems presented.

Classifier K_{17}			
Test Dataset	nAUC	Rank 1	Rank 2
D08	100.00 %	17/17	17/17
D70	90.657 %	7/17	11/17
Classifier K_{24}			
D08	99.479 %	22/24	23/24

Table 1. Classification performance

5. Conclusion

So far, biometrics research has produced significant results in terms of universality, distinctiveness and permanence. Acceptability still remains as an important issue and the main reason behind this is the fear of misuse of one's

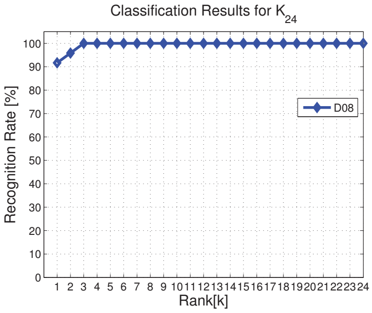


Figure 7. Cumulative matching curves for classifier K_{24} evaluated on ***D08***. The classifier was trained for 24 subjects and then applied to images acquired 8 days later (D08). The overall performance achieved in D08 arguments in favour that finger nails are *transient* biometrics which a week interval has little effect in the recognition capabilities.

permanent biometric data. Individuals are thus reluctant to volunteer their biometric characteristics where possible, and the leap in usability that biometric technology offers (i.e. password- and devicefree access to resources), cannot be realised.

This work introduces a new idea to address the acceptability issue inherent to biometric solutions. This approach, designed for collaborative individuals, instead of recording permanent data, records *transient* data, i.e. data that do change over time and are thus cancelled by nature. Users, who know that the biometric data they offer is going to be useless for recognition purposes after a certain amount of time, are likely to be more willing to offer it, even for day-to-day applications. This approach is termed *transient* biometrics; the idea is to use features with a short permanence, giving a diminutive period of recognition.

A *transient* biometric solution to the identification task was presented, which exploits texture features, extracted from finger nail images, investigating different acquisition intervals. Identification performance was high within a week but degraded considerable after a two month period. This indicates that finger nail images are a valid *transient* biometric solution.

In subsequent work we intend to expand our dataset with more subjects, more realistic capture conditions and different skin tones.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–41, Dec. 2006.

[2] S. Garg, A. Kumar, and M. Hanmandlu. Biometric authentication using finger nail surface. In *2012 12th International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 497–502. IEEE, Nov. 2012.

[3] R. Krstic. *Human Microscopic Anatomy: An Atlas for Students of Medicine and Biology*. Springer, 1991.

[4] N. Nishiuchi and H. Soya. Cancelable Biometric Identification by Combining Biological Data with Artifacts. In *2011 International Conference on Biometrics and Kansei Engineering*, number ii, pages 61–64. IEEE, Sept. 2011.

[5] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59, Jan. 1996.

[6] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, July 2002.

[7] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001.

[8] C. Rathgeb and A. Uhl. A survey on biometric cryptosystems and cancelable biometrics. *EURASIP Journal on Information Security*, 2011(1):3, 2011.

[9] A. Topping, V. Kuperschmidt, and A. Gormley. United States Patent US005751835A, 1998.

[10] F. van der Heijden, R. P. W. Duin, D. de Ridder, and D. M. J. Tax. *Classification, parameter estimation and state estimation - an engineering approach using Matlab*. John Wiley & Sons, Chichester, 2004.

[11] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever. Active shape model segmentation with optimal features. *IEEE transactions on medical imaging*, 21(8):924–33, Aug. 2002.

B ON THE USE OF FINGERNAIL IMAGES AS TRANSIENT BIOMETRIC IDENTIFIERS

Igor Barros Barbosa, Theoharis Theoharis, and Ali E. Abdallah. "On the use of fingernail images as transient biometric identifiers." In: *Machine Vision and Applications* 27.1 (Jan. 2016), pp. 65–76. ISSN: 1432-1769. DOI: 10.1007/s00138-015-0721-y. URL: https://doi.org/10.1007/s00138-015-0721-y [1].

[1] There is a typo in Eq. 9 of this paper: FN should be TN.

Machine Vision and Applications manuscript No.
(will be inserted by the editor)

On the use of fingernail images as transient biometric identifiers

Biometric recognition using fingernail images

Igor Barros Barbosa · Theoharis Theoharis · Ali E. Abdallah

Received: date / Accepted: date

Abstract The significant advantages that biometric recognition technologies offer are in danger of being left aside in everyday life due to concerns over the misuse of such data. The biometric data employed so far focuses on the permanence of the characteristics involved. A concept known as 'the right to be forgotten' is gaining momentum in international law and this should further hamper the adoption of permanent biometric recognition technologies. However, a multitude of common applications are short-term and therefore non-permanent biometric characteristics would suffice for them. In this paper we discuss 'transient biometrics' i.e. recognition via biometric characteristics that *will* change in the short term and show that images of the fingernail plate can be used as a transient biometric with a useful life-span of less than six months. A direct approach is proposed that requires no training and a relevant evaluation dataset is made publicly available.

Keywords Biometric recognition systems · fingernail · non-critical application · transient biometric characteristics · biometric access control · feature extraction · image segmentation

Igor Barros Barbosa
Department of Computer and Information Science,
Norwegian University of Science and Technology,
Sem Sælands vei 7-9
NO-7491 - Trondheim - NORWAY
Tel.: +47 735 93440
E-mail: igor.barbosa@idi.ntnu.no

Theoharis Theoharis
E-mail: theohart@idi.ntnu.no

Ali E. Abdallah
Birmingham City University,
Birmingham B4 7XG
E-mail: ali.abdallah@bcu.ac.uk

1 Introduction

Common non-biometric recognition systems confirm a subject's identity based on what a subject knowns (e.g password) or possesses (e.g access card). Such systems have the inherent risk of disclosure of the recognition token or theft of the possession. Such risks are largely mitigated when biometric recognition systems are employed, as they offer the possibility of confirming a subject's identity based on their own biometric characteristics, rather than what they know or carry. Biometric recognition systems thus offer protection from theft of access data, as well as convenience of use since access data does not have to be remembered or carried.

Recent biometric research has produced compelling results in terms of distinctiveness, universality and performance. However it has also concentrated on permanent biometric features, such as the iris, face or fingerprint. Individuals fearing the misuse of their permanent biometric data and are often unwilling to provide such data to any biometric solution, especially so for noncritical applications. Thus, the benefits granted by biometric technology (i.e. password and device-free access to resources), cannot be fully exploited.

The fear of misuse of biometric data is not unfounded; while an ID-card or password can be canceled, the same cannot be done with one's permanent biometric data. Compromised biometric information may be used for unauthorized recognition purposes while there is also the risk of discrimination via unauthorised use of such data (e.g. by insurance agencies). Cryptography is a plausible solution for the protection of biometric data, but this assumes that the subject trusts the biometric system. Cryptographic solutions are subject to the reliability of the entire computer system, and not just on the cryptographic algorithm used. Furthermore, a subject's trust on a system is not only determined by the quality of the system, but also by the importance and sensitivity of their biometric data. Thus, subjects may be reluctant to offer their biometric information for non-critical applications.

The social acceptability of recognition systems will become increasingly relevant as the 'right to be forgotten' gains momentum in legal systems worldwide [17]. In broad terms, this concept stems from the desire of the individual to determine his or her future without being stigmatized by actions performed in the past.

Research on cancelable biometrics [22,23] concentrates on the acceptability issue. It pre-transforms the biometric information before a biometric signature is extracted. As such a transformation is irreversible, the possibility of exploiting any stolen information is restricted by the fact that the exploiter has no access to the original biometric information. An extra security layer is provided as the transformation can be changed at any given time. Nevertheless, cancelable biometrics has to identify the theft of biometric information in order to change the transformation. Last but not least, a subject still has to entrust the biometric capture point with their permanent biometric information.

This work takes the acceptability matter a step further by proposing the use of biometric data that does change over the short term (i.e. is transient). This concept was discussed in [2] but is extensively explored here. Transient biometrics is defined as the set of biometric recognition technologies which depend on biometric characteristics that are proven to change over time. Thus, such biometric data automatically nullifies itself after a known period of time. In contrast to cancelable biometrics, it is the biometric data itself that changes over time. Transient biometrics is not proposed as a substitute to the cryptographic techniques that should be present in any biometric system, including a transient biometrics system. However, the use of transient data should give the user the assurance that if their data are compromised, it would automatically be rendered useless in a short period of time.

This work discusses the concept of transient biometrics (as a complete version of our initial presentation [2]) and advocates that fingernail plate images constitute a transient biometric characteristic. A set of algorithms for performing biometric recognition using such data are proposed and three methodologies for extracting transient biometric signatures from fingernail plate images are given. It also uses these methodologies in direct approaches (i.e no training or learning phases) for both the verification and the identification tasks. Another contribution is the discussion and selection of a viable signature fusion rule. Finally, a relevant new dataset is presented and made available to the research community to further explore this domain.

This paper is organized as follows. Section 2 presents the biometric literature previous work on biometric recognition based on non-permanent data as well as biometric work embracing fingernails. Section 3 details the technical side of the proposed approach followed by Section 4 that presents the new publicly available fingernail dataset and the experimental results of the proposed method. Finally, Section 5 concludes the paper, envisaging some future perspectives.

2 Previous Work

Transient biometrics is a new area with little previous work to report; we shall thus focus on related fields, the nearest of which is cancelable biometrics [22,23] (see section 1). As mentioned there, the major difference between cancelable biometrics and transient biometrics is that in the case of the former technology, the biometric data are protected via an irreversible transform, while in the case of the latter, the biometric data itself has only temporary recognition value. Individuals are therefore more likely to volunteer such transient data, even for non-critical applications and even when they are not entirely confident on the integrity of the capture point.

Person re-identification is a related transient problem arising in the surveillance area. The objective is to identify if a person has been previously observed in a non-collaborative subject setting using non-invasive techniques. It is therefore usually based on images, from which appearance-based local features are used to re-identify a given subject [3,8]. Such biometric systems produce a transient identification solution since it is only possible to identify a subject until this subject changes clothes or other major appearance characteristics. Appearance is one of the few options to use in re-identification within a surveillance setting where the images are often taken from a distance using a video camera; however it is questionable whether it can be regarded as a biometric trait, since it is rather easy to spoof by knowledgeable subjects.

The Bioelectrical characteristics of a fingernail are used as a biometric signature in [4]. This patent work presents a RFID chip glued over the fingernail. This embedded system measures the subjects' capacitance, which is claimed to be unique, thus creating a biometric solution based over the fingernail region.

The use of fingernail images as biometric data has been the topic of few different lines of research. The skin under a nail plate, called nail bed, is unique for each individual [11]. A special acquisition system has been designed to acquire images of the nail bed. Such images use the grooves of the nail bed for recognition purposes [26]. The fingernail surface has also been explored for a biometric authentication system [12]. This work segments the five fingernails as regions-of-interest (ROI) from a hand image using a contour segmentation algorithm. The hand is photographed while resting on a white surface. This segmentation methodology works but the employed dataset was biased with respect to the subjects' skin tones. Haar wavelet and Independent Component Analysis (ICA) are used to create a biometric signature. From the five fingernail images, three are used for training and two for testing. The methodology yields high recognition rates, yet the paper does not evaluate the effect of the growth of the fingernails on the recognition rates (i.e. no longitudinal analysis is performed). The work of [10] combines biometric information from fingernails and finger knuckles to create a multi-modal biometric system. Mel Frequency cepstral coefficients (MFCC) is employed to create a finger knuckle biometric signature. The fingernail signature is created by using both approximation and detail coefficients of a second level wavelet image decomposition. The final classification is done by a Multilayer Percptron (MLP). Similar to the work of [12], a high classification performance is achieved. Nevertheless, the work does not assess the behaviour of the information over time. In both [10, 12] the final signatures are composed from information of three fingers. There is no special fusion methodology to keep transient characteristics of the data. It is thus likely that the proposed solution learns hard biometric characteristics from the subject's fingers instead of transient information. Thus, none of the above works involving fingernails focus on their transient nature.

The work of [25] shows that the green camera channel and a 3D model of the fingertip can be used for measuring the force exerted by the fingertip. This work explores changes in coloration of the fingernail images to detect the magnitude of the force being applied to/by a specific region. A Bayesian classifier is then used to deduce the relation between force and coloration changes. In their latest work [7], the authors presented an automated calibration for a setup using an adjustable camera, controlled lighting and a magnetic levitation haptic device. Thus they are able to measure forces using only the camera image with higher accuracy. The work also presents three approaches to fingernail registration. The registration results achieved are impressive but depend on a controlled lighting setup.

The work of [6] presents a color based fingernail segmentation. The work found that the third principal component of the RGB color-space can be used to differentiate fingernails from images of skin patches. The work is assessed on a small dataset of five subjects

Our previous work assesses the use of fingernail images to create a transient biometric solution [2]. There, a small dataset was used to evaluate the longitudinal identification rate of fingernails. It was shown that recognition performance decreases to unusable rates after two months . These results are in line with physiological studies showing that healthy fingernails are replaced within three to six months [28]. In the present work we extend and complete the evaluation of fingernail images as transient biometrics by comprehensive testing in terms of algorithms and dataset.

3 Proposed Approach

The transient biometric solution presented in this work is a direct approach to the verification and identification tasks. No training is employed in the task of matching a biometric signature against a database of previously collected signatures.

An image of the right index fingernail is used for the extraction of the biometric signature. The approach will be divided into three parts. The first part outlines the image preprocessing which is necessary in order to make the image suitable for biometric signature extraction. The second part details the extraction of the biometric signature from a fingernail image. The third part describes how signatures are compared and matched.

3.1 Fingernail Plate Image Preprocessing

Image preprocessing ensures that the data delivered to the signature extraction phase fulfills some basic requirements. This should be a square image composed mainly of the fingernail and it eliminates the possibility of using potentially hard biometric information form finger-joints or finger shape [1].

To automatically segment fingernail images, an object detector as proposed in [27] and extended in [15] is employed. In this object detector, a classifier is trained with sample images of manually segmented fingernails, that match the requirements of the signature extractor, generating the so called positive samples. Negative samples are also generated using sample images with no fingernail.

This fingernail image classifier is composed of a cascade of elementary classifiers, also called stages. A given region of interest (ROI) is either rejected by a stage or it proceeds to the next stage. Initial stages are simpler than subsequent ones and focus in rejecting non-positive ROIs, i.e. areas where there is a low chance of detecting a fingernail. As such areas represent larger portions of the images, the overall detection speed is increased. When a stage approves a ROI, this region is passed on to the next stage. If the ROI is approved by every stage, then this region is classified as a fingernail image. Each stage is an Adaboost classifier which relies on haar-like features as input.

A large number of Haar-like features can be computed for every ROI which is significantly larger than the number of pixels of the given region. Thus feature selection is a requirement. Adaboost is employed for both the selection of such features, as well as for the training of the classifiers.

After the input image is converted to grayscale, as required by the object-detection algorithm, a ROI is defined

[1] The used dataset (see Section 4.1) provides already segmented fingernail images using the methodology presented in this Section.

by a scanning window. Robustness to scale variation is due to scaling of the detector, which can be applied at different scales with no extra cost. As the scanning window gazes through the input image, the cascade classifiers detect the fingernail multiple times. To achieve a final detection, as shown in Fig. 1, the detected ROIs are overlaid and merged into a single detection by selecting the average of the bounded regions.

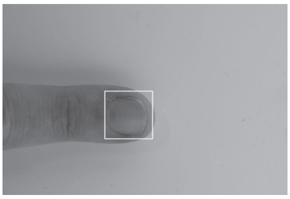


Fig. 1 Result from the object detector proposed by [27,15] trained to detect fingernails. Detection outlined by a white bounding box.

Given the high resolution of the input images, the final ROI is then scaled to $300px \times 300px$. The original color image of the selected ROI is then sent to the signature extraction stage.

3.2 Signature Extraction

Every individual has a unique fingernail bed pattern, similar to a fingerprint, which influences the texture that constitutes the fingernail plate [11]. This texture is also influenced by the day-to-day interaction of the fingernail plate and the environment. Therefore, it is common to find white spots, marks and scratches over the fingernail plate. It is this rich texture region of the fingernail plate that is analyzed in order to create a texture based signature using a grid implementation of Local Binary Patterns [19]. This signature extraction process is described in Sec. 3.2.1.

The fingernail plate boundary and the unique white spots on it can be quite discriminative, and to exploit such characteristics two feature descriptors were employed. Section 3.2.2 explains how the SIFT[16] and BRISK [14] descriptors are used to create a second signature.

Notice that both the fingernail plate texture and its boundary shape have a transient nature since the fingernail plate changes completely over a period of about 6 months [28].

3.2.1 LBP Based Signature

Local Binary Patterns (LBP) was originally proposed as a reliable texture descriptor [19] and is known for its speed, robustness and capacity to differentiate between micro-patterns. The signature extraction uses a GRID extension of the LBP, based on the work of [1].

For every image pixel an LBP value can be computed by comparing the actual pixel value to its neighborhood. The pixel neighborhood is defined by a circle of radius R and a set of P equally spaced sample points. The LBP value for the central pixel is derived out of a binary comparison against the sample points. One bit is assigned to each sample point. The least significant bit value comes from the comparison against the top-left sampling point. It receives 1 when the central pixel is greater than or equal to the sampling point and 0 otherwise. This procedure is then applied to all other sampling points, in a clockwise manner. Therefore when 8 sampling points are used, there are a total of 256 possible LBP values. Fig. 2 illustrates the LBP calculation with a sample neighborhood of $(P,R) = (8,2)$.

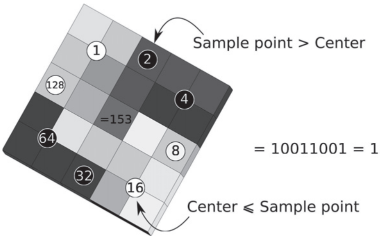


Fig. 2 LBP sample points are shown in red or green circles. The value of a sample point is bilinearly interpolated for sampling points that are not located on the center of a pixel. Dark circles denote sample points that have a lower value than the center pixel. Bright circles denote sample points which have a greater or equal value to the center pixel. Circle numbers indicate the index of the bit position in the LBP code.

LBP values are called uniform if the binary part is composed of one or two bit-wise transitions. Uniform LBP values account for the majority of encountered patterns on natural images [20,1]. For example, in the case of eight sampling points, the patterns 00010000 and 11001111 are uniform since they have two bit-wise transitions. For eight sampling points, a total of 58 out of the 256 possible patterns are uniform.

To extract the LBP signature the input fingernail image is divided into 16 blocks using a 4×4 grid. Each image block is submitted multiple times to a Gaussian smoothing function, creating a Gaussian pyramid image set. This process generates a total of 48 smaller images from each input image. The final signature comes from the computation of uniform LBP values with a sample neighborhood of $(P,R) = (8,2)$ for each color channel. For each small image, 3 histograms of 59 bins are computed, 58 bins employed for the uniform patterns and the last bin for the non-uniform patterns. Thus, for each input image the signature is composed of 3×48 histograms of 59 bins. Although these histograms sum up to 8496 bins, the curse of dimensionality is avoided thanks to the matching technique.

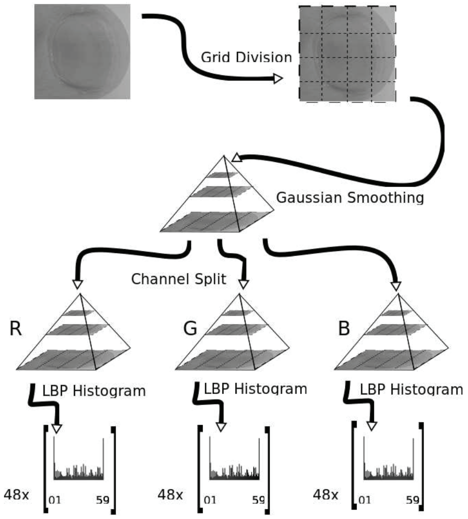


Fig. 3 LBP signature extraction. Each input image generates an LBP based signature comprised of 144 histograms, or 48 histograms per color channel.

3.2.2 *Descriptor Based Signature*

The first step of the descriptor based signature is to decompose the color image into three monochromatic images. Then keypoint detection is performed on each color channel. A keypoint detector that produces an output of low count is fundamental to any matching technique that relies on image descriptors. A low count of keypoints allows the solution to compare only significant points, speeding up the process and yielding a more robust and discriminative set of features. A multitude of different keypoint detectors have been proposed in computer vision while it is common for feature descriptors to propose their own keypoint detectors as for example in [14, 16].

Given that the image pre-processing presented in Sec. 3.1 already yields an image with a decent fingernail alignment and unique orientation, the use of keypoint detectors that are robust to such characteristics would be superfluous. A single fast and simple keypoint detector is shared across different descriptors. The selected keypoint detector is Good Features To Track (GFTT) [24] and represents a modification of the well known Harris corner detector [9]. Fig. 4 shows the result of this keypoint detector. The keypoints concentrate around the fingernail plate boundary as well as fingernail plate scratches and white spots, which are ideal to discriminate across subjects.

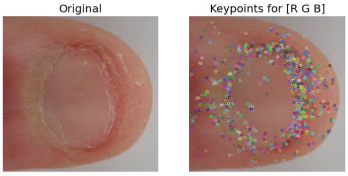


Fig. 4 Keypoints computed for the red, green and blue channels of a fingernail plate image. The keypoints concentrate around the fingernail plate border.

Having defined the keypoints, descriptors must next be computed on them. Two different keypoint descriptors are employed. The Scale Invariant Feature Transform - SIFT [16] was chosen for its success as a robust descriptor while the Binary Robust Invariant Scalable Keypoint - BRISK was selected for its efficiency [14]. Thus two biometric signatures are created for each input image, based on SIFT and BRISK respectively.

The total number of keypoints is reduced since both SIFT and BRISK prune down the keypoints by evaluating their stability (via contrast, distance to other keypoints, neighborhood noise, etc). Final monochromatic images have an average of 800 stable keypoints. Therefore, on average, a total of 2400 keypoints compose each of the two descriptor based signatures, one for SIFT and the other for BRISK.

3.3 Signature Matching

Signature matching defines the metrics used in order to compare different signatures. Ideally there is a small variation across signatures from the same subject, and large variation across signatures from different subjects. It is common to rely on machine learning techniques to discover patterns in signatures and then use such patters for signature matching. Previous work [2] even employed Bayesian classification and dimensionality reduction for signature matching.

Since the proposed approach intends to employ a direct methodology to transient biometrics, it avoids techniques that rely on training and cannot employ dimensionality reduction. Signature matching is thus done in stages; the three signature types are matched independently. This is also convenient for the exploration of different signatures fusion methodologies.

3.3.1 LBP Based Signature Matching

Matching LBP signatures fundamentally depends on obtaining a match score between two histograms, P and Q, with n bins. For this task, Cosine similarity as stated in Eq. 1, is employed:

$$\Phi(P,Q) = \frac{P \cdot Q}{\|P\| \|Q\|} = \frac{\sum_{b=1}^{n} P_b \times Q_b}{\sqrt{\sum_{b=1}^{n} (P_b)^2} \times \sqrt{\sum_{b=1}^{n} (Q_b)^2}} \tag{1}$$

A single LBP signature is composed of 144 histograms. Each histogram is derived from a small image region, and depends on the layer of a Gaussian pyramid and on the color channel. The final matching score is given as the mean score of matching corresponding histograms across different input images. Therefore, a histogram is only compared to its counterpart in another image, which deals with the curse of dimensionality while giving the signature the capability of describing texture and spatial relationships at the same time. If X_i represents the ith histogram of image X, then the LBP matching score L between images A and B is given by Eq. 2:

$$L(A,B) = \frac{1}{144} \sum_{i=1}^{144} \Phi(A_i, B_i) \tag{2}$$

3.3.2 Descriptor Based Signature Matching

A Fast Approximate Nearest Neighbor Search, proposed by [18], is first employed to evaluate the euclidean distance between keypoint combinations of the two images to be matched. The resulting matches are evaluated by a RANSAC [5] algorithm to remove bad matches and to detect a consensus set of plausible matches.

In order to create a unique signature matching score which can be combined with other scores, the RANSAC algorithm is executed multiple times. Each time it returns the percentage Υ of keypoints that are part of the consensus set. [2] The descriptor matching score D of images A and B is then computed as the average of the consensus keypoint percentage:

$$D(A,B) = \frac{1}{n} \sum_{i=1}^{n} |\Upsilon_i| \tag{3}$$

When matching is performed across multiple images using a distance measure, it is usual to normalize its output to yield a consistent matching score across images; however this normalization is only possible after all inter-image distances have been computed. Being an average percentage value, the proposed matching score of Eq. 3 does not require post-normalization and is thus suitable as a direct technique to compute the score between two images.

3.3.3 Signature Fusion

LBP describes the micro-texture that comprises the fingernail plates and their spatial relationships. In contrast to the descriptor based techniques LBP does not focus on discriminating white spots, marks or the border between fingernail plate and skin. A methodology for merging the different techniques is thus necessary. The idea of signature fusion is to generate a final matching score, which combines the properties of LBP, BRISK and SIFT.

The work of [21] shows different methodologies to similarity score fusion. Assuming that S_n is the nth similarity score, it proposes five fusion functions, as shown in (Eq. 4). S_A represents the arithmetic mean of the Manhattan (L_1) metric. S_E computes the root mean square of similarities and performs as a Euclidean (L_2) metric. S_G computes the product of similarities and works as geometric mean metric. S_{max} and S_{min} are simple rules to respectively select as final score the maximum or minimum similarity score:

$$S_A = \frac{1}{n} \sum_{f=1}^{n} S_f \tag{4}$$

$$S_E = \frac{1}{\sqrt{n}} \left(\sum_{f=1}^{n} {S_f}^2 \right)^{1/2} \tag{5}$$

$$S_G = \left(\prod_{f=1}^{n} S_f \right)^{1/n} \tag{6}$$

$$S_{max} = \max_{f=1}^{n} \left(S_f \right) \tag{7}$$

$$S_{min} = \min_{f=1}^{n} \left(S_f \right) \tag{8}$$

Since the proposed matching score functions of Eq. 2 and Eq. 3 have bounded outputs in the range $[0,1]$, all of the proposed methodologies of Eq. 4 could be employed as score fusion techniques. However most of them cannot handle the hidden issue of large variations in the skewing of the distribution scores.

In the case of the Cosine similarity used in the LBP matching technique (Eq. 2), the scores will have a propensity towards high values. While correct matches will present higher matching scores that wrong matches, given the derivation methodology and the fact that similarity is computed on a 59 dimension vector, wrong matches are also likely to give high matching scores.

The opposite is true in the case of descriptor matching where a natural bias towards low values occurs in the matching scores (Eq. 3). Given the low re-projection error accepted by the RANSAC algorithm when computing the

[2] Ransac is run multiple times to ensure change in the seed of the random number generator. Similar results are achieved if RANSAC is executed once for a longer time.

consensus set, the matching score technique will generally produce lower values; of course it is still true that correct matches are likely to yield higher results than wrong matches.

Given the aforementioned matching score value distributions, Eq. 4 and 5 would give a bigger weight to the LBP features, while Eq. 7 would ignore descriptor based features and Eq. 8 would ignore LBP based features. Such issues can be avoided with the normalization of the score distributions, but such a process is unacceptable in a direct approach.

The selected fusion technique is the geometric mean (Eq. 6). Thus the final matching score does not weigh differentially the LBP and descriptor based features. To explore different combinations of features a total of three signature fusions are computed. The first signature fusion F_{LS} relies on fusing LBP and SIFT features. This way the final signature will use the micro-texture capabilities of LBP combined with SIFT's capability of describing the fingernail border and fingernail plate white spots.

A second signature fusion F_{LB} is created using LBP and BRISK, in order to evaluate how the BRISK descriptor compares to the SIFT descriptor in this task. Finally a third signature F_{LBS} is defined, fusing LBP, BRISK and SIFT into a final matching score.

4 Experiments

This section describes the publicly available experimental dataset of fingernail plate images as well as the verification and identification performance of the proposed method on this dataset.

4.1 Publicly Available Dataset of Fingernail Plate Images

An extended version of an experimental dataset called Transient Biometrics Nails Dataset (TBND) was created[3]. TBND is composed of images of the right index finger. During acquisition the subject was instructed to lay their finger over a flat white surface and a simple point-and-shoot camera was used to acquire an image without the the use of a flash. No explicit instructions with respect to force applied were given and thus our results incorporate arbitrary force differences between users and capture sessions. Acquisition was thus done in a semi-controlled environment; apart from the white background and indirect lighting, the images present variation with respect to scale, focal plane and illumination. The dataset consists of three subsets, each one compromising the same 93 subjects, but varying on acquisition date.

The first subset **D01** consists of images acquired on the first acquisition day. The second subset **D02** is composed of images acquired one day later. The third subset **D30** was acquired one month after the first acquisition date. Given acquisition restrictions, the acquisitions of **D30** have up to two days' tolerance. This represents a massive expansion of the originally collected dataset [2], and will also be made available through NTNU Visual Computing lab's website [http://www.idi.ntnu.no/grupper/vis/TBND].

4.2 Verification Performance

To evaluate verification performance a simple direct classifier is used, which thresholds the matching score between two images to determine if they correspond. To asses the verification behavior across time and thus determine how transient the explored fingernail plate biometric really is, images from **D01** are matched against **D02** and **D30**. It is anticipated that the fingernail plate biometric information transforms as the fingernail grows. Therefore, higher verification rates should be expected for matches across a day interval (**D01**x**D02**) than for matches across a month interval (**D01**x**D30**). The difference in verification performance between these two cases will determine how transient the proposed biometric is.

Assuming that each subject represents a class, verification can be treated as a binary classification problem where the proposed solution verifies if a query image is from whom it is claimed to be. This implies that the classification output can yield four types of result: true positive, true negative, false negative and false positive. These outcomes are typically computed by comparing each image from the first dataset (called a query) against every image of the second dataset (called a target). A true positive is the case where the query is correctly classified as a match for the target while a true negative is the case where the query is correctly classified as a non-match for the target. A false negative is when the query is wrongly classified as a non-match for the target while a false positive is when the query is wrongly classified as a match.

Let TP, TN, FN and FP represent the cardinalities of the sets that represent the above four possible classification outcomes. By defining the False Positive Rate (FPR) as shown in Eq. 9 and the True Positive Rate (TPR) as shown in Eq. 10, it is possible to assess the verification performance with the Receiver Operator Characteristics (ROC) curve. This methodology evaluates how different thresholds on FPR (i.e. the risk of the system) impact on TPR (i.e. the convenience in the use of the system), by plotting FPR vs TPR:

$$FPR = \frac{FP}{FP+FN} \tag{9}$$

[3] Thanks to Cham Athwal of the School of Digital Media Technology, Birmingham City University

$$TPR = \frac{TP}{TP+FN} \tag{10}$$

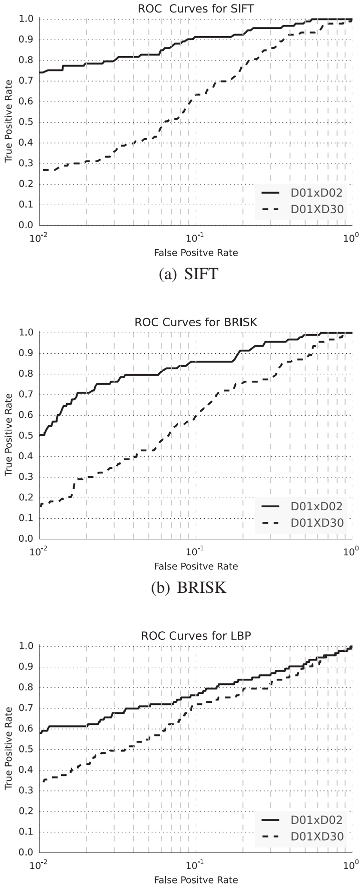
Conventionally when a biometric methodology is evaluated with a ROC curve, the False Positive Rate ranges from 10^{-3} to 1. Given that our datasets compromise 93 subjects (being the first datasets of their kind), evaluating at $FPR = 10^{-3}$ would produce results of low statistical significance since the classification of a single subject would greatly alter the output values. This, in conjunction with the assumption that the proposed methodology is aimed at non-critical biometric applications, led us to compute the ROCs curves in the range 10^{-2} to 1.

We first compute the ROCs for each of the basic features used separately: SIFT, BRISK and LBP. We assess the verification rates for matches done with a day interval (**D01**x**D02**) against matches done with a month interval (**D01**x**D30**). The achieved results are shown in Fig. 5.

These ROC curves shows that all three features undergo an expected performance deterioration within the course of a month. This decay in performance shows that fingernail plate images are a *transient* biometric feature with a short lifetime.

We next present results on the fused signatures in Fig. 6. The ROC curve for F_{LS} shows verification performance for a signature based on the fusion of LBP and SIFT. In theory this signature fusion will have the capability to discriminate subjects' fingernails based on fingernail texture due to LBP, as well as due to fingernail border characteristics and fingernail white spots due to the SIFT descriptor. The computed ROCs are shown in Fig. 6 (a). The achieved results indicate that the proposed signature fusion technique of Eq. 6 on the fused signature F_{LS} outperforms the two best performing individual features shown in Fig 5 [(a) & (c)]. Therefore, the signature fusion technique was successful. This is further demonstrated in the fusion that results in the F_{LB} signature; in this case the LBP signature is combined with the efficient BRISK descriptor. The ROC curves for this fusion are shown in Fig. 6 (b). The final signature fusion F_{LBS} employs all three features, LBP, BRISK and SIFT. The idea is similar to before; use LBP to describe texture and SIFT/BRISK to describe fingernail borders and discriminating points. This time the fusion will also exploit any complementary information hidden in the combination of BRISK and SIFT. The results are shown in Fig. 6 (c).

Table 1 gives verification data for the proposed signatures. It shows the True Positive Rates achieved at an FPR of 0.01. The results indicate that the final signature fusion F_{LBS} for fingernail plate images is a transient biometric with a lifetime of less than 6 months. This is based on the assumption that the TPR of 0.247 after one month is already at an unacceptable level for practical recognition purposes (i.e. the biometric feature has been invalidated) and that the TPR value of 0.774 after one day is acceptable, at least for non-critical applications. We sextuple the invalidation period (from one to six months) to allow for possible future algorithmic improvements that could improve these figures and also taking into account physiological knowledge indicating that human fingernails totally outgrow within a period between 3 and 6 months [28].

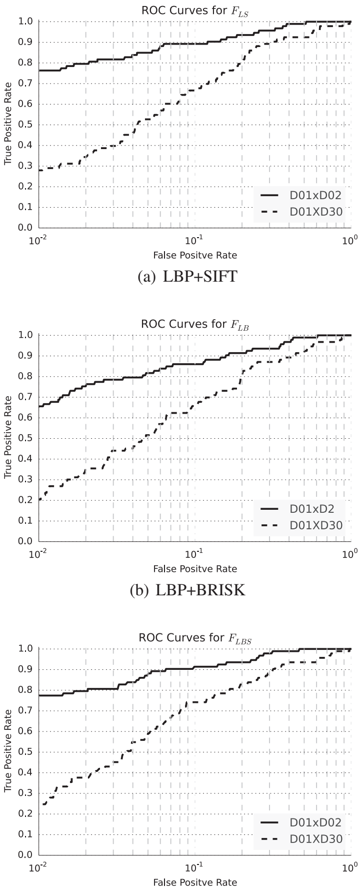


(a) SIFT

(b) BRISK

(c) LBP

Fig. 5 ROC curves for SIFT **(a)**, BRISK **(b)** and LBP **(c)**. Curves labeled **D01**x**D02** show the verification performance for matches done with a day interval, while curves labeled **D01**x**D30** show the verification performance across an interval of a month. The performance decay between the two intervals shows that fingernail plate images are a *transient* biometric feature.



(a) LBP+SIFT

(b) LBP+BRISK

(c) LBP+BRISK+SIFT

Fig. 6 ROC curves for combinations of [LBP+SIFT] F_{LS} (**a**), [LBP+BRISK] F_{LB} (**b**) and [LBP+BRISK+SIFT] F_{LS} (**c**). The performance improvement against the individual features of Fig. 5 indicates that a successful feature fusion methodology was found. Curves labeled **D01**x**D02** show the verification performance for matches done with a day interval, while curves labeled **D01**x**D30** show the performances for a month interval. The larger performance decay between the two intervals compared to single features, further supports the case that fingernail plate images are a *transient* biometric feature.

4.3 Identification Performance

The identification task is a multi class problem, where a query biometric signature is compared against a list of collected signatures (the target set) with the objective of finding if the query matches any of the collected signatures. To evaluate identification performance the cumulative match curve (*CMC*) will be used. *CMC* gauges the probability of a signature from a query dataset, in this case **D02** or **D30**, being correctly matched in the first *k* ranked subjects from the target set, in this case **D01**. The subjects of the target set are ranked using F_{LBS} (Fusion of LBP, BRISK and SIFT by Eq. 6). The abscissa in the CMC graph shows the rank while the ordinate shows the probability of a correct match up to that rank.

Table 1 Verification data (TPR) at 0.01 FPR across datasets captured with a day interval (**D01**x**D02**) and a month interval (**D01**x**D30**).

	True Positive Rate for different intervals	
Signature	One day	One Month
SIFT	0.742	0.269
BRISK	0.505	0.151
LBP	0.581	0.333
F_{LS}	0.763	0.279
F_{LB}	0.656	0.204
F_{LBS}	0.774	0.247

In the identification task, a simple threshold classifier plays no role on performance and the results show how reliable a computed matching score is for finding a correct match from an entire dataset. Figure 7 shows the CMC for F_{LBS} and gives another evaluation of the transient nature of the proposed biometric approach.

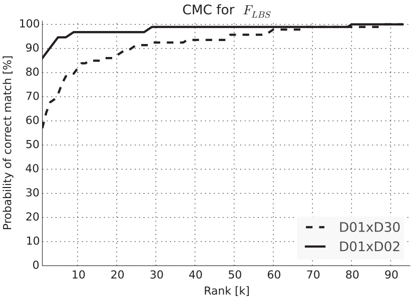


Fig. 7 Cumulative match curves for F_{LBS}. The scores are computed by comparing images from two query sets, respectively acquired within a day interval **D02** and within a month interval **D30**, against images of the target set **D01**. The decay in performance from **D02** to **D30** further supports the presumption that fingernail plate images constitute a transient biometric characteristic.

It is interesting to compare the identification results against our previous study which involved only 24 subjects [2]. Although the present method is significantly more robust, it achieves 86.022% Rank one identification on 93 subjects compared to 99.479% Rank one for the previous method on 24 subjects. [4]

4.4 Matching Score Distribution

In this section the results are analyzed using two charts showing matching score distributions. The score distribution charts give us an unbiased view of the behavior of a transient biometric across time that eliminates any bias introduced by classifier selection.

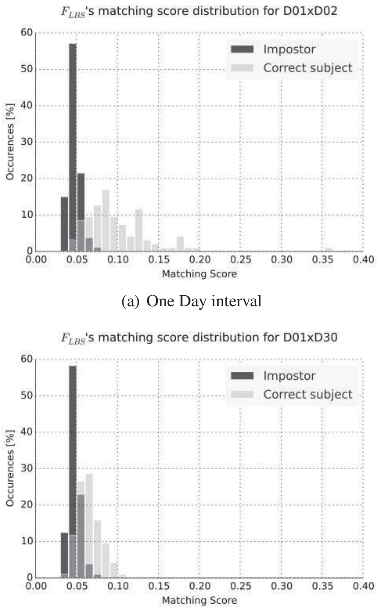
A histogram is used to approximate the probability density function (PDF) for matches done within a day interval (**D01**x**D02**) and within a month interval(**D01**x**D30**). In both scenarios the matching scores were divided into two different PDFs. The first PDF, called 'Impostor', accounts for the cases where the matching score is computed between a query subject and an impostor. The second PDF, named 'Correct Subject', accounts for the cases where the matching score is computed between a query subject and itself.

By separating the matching scores into these two categories, impostors and correct subjects, it is possible to observe the effect of time, and thus get an idea of the decay in recognition performance. These matching scores are computed using F_{LBS} (Fusion of LBP, BRISK and SIFT using Eq. 6). The abscissa represents matching score values while the ordinate represents frequency expressed as a percentage.

5 Conclusion

Transient biometrics are introduced as a plausible solution to the acceptability issue of biometric recognition systems. It presents a methodology which reduces the risk of misuse of biometric information; instead of relying on permanent biometric data it uses biometric data that changes within the short term and thus nullifies itself. Therefore, it is an engaging solution for collaborative individuals which are reluctant to volunteer hard biometric information (e.g. fingerprints, retina images) for non-critical biometric recognition tasks. Given the knowledge that the collected biometric data has an expiration date and becomes useless for recognition after this, individuals are more likely to volunteer such biometric data for day-to-day recognition tasks.

A transient biometric feature and methodology for verification and identification tasks was presented. This builds and completes previous work [2] and uses fingernail plate images. A new dataset is presented and will be made publicly available; it consists of a larger number of subjects, more realistic (and challenging) capture conditions as well as subjects with different skin tones.



(a) One Day interval

(b) One Month interval

Fig. 8 Probability Distribution Functions (PDF) for matches against impostors and correct subjects for a one day interval **(a)** and a month interval **(b)**. The change observed in the correct subject PDF from (a) to (b) further indicates that fingernail plate images are a plausible *transient* biometric. Notice that the score distribution for correct matches changes in such a way, that in case (b) it is hard to differentiate matching scores between correct subjects and impostors. This analysis eliminates any influence that may have been added from the choice of classifier. There is no post normalization after the matching score of Eq. 3; as this score comes from the percentage of RANSAC inliers, using a rather strict threshold, it is natural to have low matching scores.

The proposed methodology exploits both texture features and (descriptor based) information extracted from discriminant fingernail keypoints. No training or machine learning techniques are employed in the computation of the biometric signatures making this a direct approach.

Both verification and identification performance was high within a day interval but degrades considerably after a month, indicating that fingernail plate images are a valid *transient* biometric feature. Here we consider the performance of 80% to be high, given the novelty of the explored biometric trait. Nevertheless, some more traditional (non-transient) biometric technologies currently offer significantly higher recognition rates. It would therefore be important to explore improvements in the recognition rate, so that the proposed transient biometric could become commensurate with currently 'acceptable' recognition levels. One such improvement would

[4] Note that in the current study we compare day 1 to day 2 while in the previous study the comparison was across day 1 and day 8.

be the use multi-biometrics of the same class, by exploiting more than one fingernail per subject. In this case it is important to adopt a suitable fusion technique and to maintain the transient nature of the solution when extracting information from multiple fingernails, which could potentially be derived from a single image of the hand (that potentially also contains non-transient data). Another possibility for improving the recognition rate would be to use machine learning techniques instead of the current direct approach.

If one was willing to sacrifice the transient nature of the proposed approach, e.g. in order to create a multi-biometric solution using fingerprints and fingernails, the entire images of the fingers could be used. This would fit well with the works of [13], where finger knuckle images are used for biometric identification. A multi-biometric approach would also relate to the work of [12] where both fingernails and finger knuckles are used as biometrics.

The current size of the dataset does not allow for a realistic scalability study (e.g. it is not possible to compute a meaningful FPR of 10^{-3}). In further work, it would be interesting to expand the size of the dataset in order to allow for such a study.

Acknowledgements We would like to thank Cham Athwal of the School of Digital Media Technology, Birmingham City University, for co-acquiring the largest part of the fingernail datasets.

References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: application to face recognition. IEEE transactions on pattern analysis and machine intelligence **28**(12), 2037–41 (2006). DOI 10.1109/TPAMI.2006.244
2. Barbosa, I.B., Theoharis, T., Schellewald, C., Athwal, C.: Transient biometrics using finger nails. In: Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on, pp. 1–6 (2013). DOI 10.1109/BTAS.2013.6712730
3. Bazzani, L., Cristani, M., Murino, V.: Symmetry-driven accumulation of local features for human characterization and re-identification. Comput. Vis. Image Underst. **117**(2), 130–144 (2013)
4. Chornenky, T.: United states patent us20030098774 (2001). URL https://www.google.no/patents/US20030098774
5. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Commun. ACM **24**(6), 381–395 (1981)
6. Fujishima, N., Hoshino, K.: Fingernail detection system using differences of the distribution of the nail-color pixels. JACIII **17**(5), 739–745 (2013)
7. Grieve, T., Lincoln, L., Sun, Y., Hollerbach, J., Mascaro, S.: 3d force prediction using fingernail imaging with automated calibration. In: Haptics Symposium, 2010 IEEE, pp. 113–120 (2010). DOI 10.1109/HAPTIC.2010.5444669
8. Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B.: Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In: 2nd ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC-08). Stanford, Palo Alto, États-Unis (2008)
9. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey vision conference, vol. 15, p. 50. Manchester, UK (1988)
10. Kale, K., Rode, Y., Kazi, M., Dabhade, S., Chavan, S.: Multimodal biometric system using fingernail and finger knuckle. In: Computational and Business Intelligence (ISCBI), 2013 International Symposium on, pp. 279–283 (2013). DOI 10.1109/ISCBI.2013.63
11. Krstic, R.: Human Microscopic Anatomy: An Atlas for Students of Medicine and Biology. Springer (1991)
12. Kumar, A., Garg, S., Hanmandlu, M.: Biometric authentication using finger nail plates. Expert Systems with Applications **41**(2), 373 – 386 (2014). DOI 10.1016/j.eswa.2013.07.057
13. Kumar, A., Ravikanth, C.: Personal authentication using finger knuckle surface. Information Forensics and Security, IEEE Transactions on **4**(1), 98–110 (2009). DOI 10.1109/TIFS.2008.2011089
14. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: Binary robust invariant scalable keypoints. Computer Vision, IEEE International Conference on **0**, 2548–2555 (2011). DOI http://doi.ieeecomputersociety.org/10.1109/ICCV.2011.6126542
15. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: Image Processing. 2002. Proceedings. 2002 International Conference on, vol. 1, pp. I–900–I–903 vol.1 (2002). DOI 10.1109/ICIP.2002.1038171
16. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004). DOI 10.1023/B:VISI.0000029664.99615.94
17. Mantelero, A.: The eu proposal for a general data protection regulation and the roots of the 'right to be forgotten'. Computer Law and Security Review **29**(3), 229 – 235 (2013)
18. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: International Conference on Computer Vision Theory and Application VISSAPP'09), pp. 331–340. INSTICC Press (2009)
19. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition **29**(1), 51–59 (1996). DOI 10.1016/0031-3203(95)00067-4
20. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7), 971–987 (2002)
21. Perakis, P., Theoharis, T., Kakadiaris, I.A.: Feature fusion for facial landmark detection. Pattern Recognition **47**(9), 2783 – 2793 (2014). DOI http://dx.doi.org/10.1016/j.patcog.2014.03.007
22. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. IBM Systems Journal **40**(3), 614–634 (2001). DOI 10.1147/sj.403.0614
23. Rathgeb, C., Uhl, A.: A survey on biometric cryptosystems and cancelable biometrics. EURASIP Journal on Information Security **2011**(1), 3 (2011). DOI 10.1186/1687-417X-2011-3
24. Shi, J., Tomasi, C.: Good features to track. In: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, pp. 593–600 (1994). DOI 10.1109/CVPR.1994.323794
25. Sun, Y., Hollerbach, J.M., Mascaro, S.A.: Measuring fingertip forces by imaging the fingernail. p. 20. IEEE Computer Society, Los Alamitos, CA, USA (2006). DOI http://doi.ieeecomputersociety.org/10.1109/VR.2006.97
26. Topping, A., Kuperschmidt, V., Gormley, A.: United States Patent US005751835A (1998)
27. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. I–511–I–518 vol.1 (2001). DOI 10.1109/CVPR.2001.990517
28. Yaemsiri, S., Hou, N., Slining, M., He, K.: Growth rate of human fingernails and toenails in healthy american young adults. Journal of the European Academy of Dermatology and Venereology **24**(4), 420–423 (2010). DOI 10.1111/j.1468-3083.2009.03426.x

C EEG BIOMETRICS: ON THE USE OF OCCIPITAL CORTEX BASED FEATURES FROM VISUAL EVOKED POTENTIALS

Igor Barros Barbosa, Kenneth Vilhelmsen, Audrey van der Meer, Ruud van der Weel, and Theoharis Theoharis. "EEG Biometrics: On the Use of Occipital Cortex Based Features from Visual Evoked Potentials." In: *28th Norsk Informatikkonferanse, NIK 2015, Høgskolen i Ålesund*. Bibsys Open Journal Systems, Norway, Nov. 2015. URL: http://ojs.bibsys.no/index.php/NIK/article/view/243

EEG Biometrics: On the Use of Occipital Cortex Based Features from Visual Evoked Potentials

Igor Barros Barbosa[1], Kenneth Vilhelmsen[2], Audrey van der Meer[2], Ruud van der Weel[2] and Theoharis Theoharis[1]

[1]Department of Computer and Information Science,
[2]Department of Psychology,
Norwegian University of Science and Technology

Abstract

The potential of using Electro-Encephalo-Gram (EEG) data as a biometric identifier is studied. This is the first study that assesses looming stimuli for the creation of biometrically useful Visual Evoked Potentials (VEP), i.e. EEG responses due to visual stimuli. A novel method for the detection of VEP responses with minimal expert interaction is introduced. The EEG data, segmented based on the VEP, are used to create a reliable feature vector. In contrast to previous studies, we provide a publicly available evaluation dataset based on infants which is therefore not biased due to unhealthy individuals. Only data from the occipital cortex are used (i.e. about 3 of the many possible electrode positions in the scalp), making the potential EEG biometric capture devices relatively simpler.

1 Introduction

The quest for reliable biometric identifiers has dominated biometrics research over the last decades. Spoofing is a major concern and in this respect many biometrics fail and have to resort to liveness testing, which further complicates matters. One biometric identifier that is potentially very hard to spoof is the activity of a person's brain. The question, of course, is how reliable a biometric identifier such brain data would be and it is this question that the present paper addresses by providing new results on an *unbiased* dataset of brain activity, which will be made publicly available.

Most work on biometric recognition from brain activity concentrates on Visual Evoked Potentials (VEPs) because of the relatively clear response and the fact that a publicly available dataset is available [3, 18]. However, this dataset is considered biased because the have been acquired from alcoholic individuals [5]. We addressed this issue by creating a new dataset based on electroencephalographic (EEG) recordings from infants' VEPs. Infants have considerably thinner skulls than adults and hardly any hair. This can explain why we are able to identify looming-related brain electrical responses on a trial-by-trial basis in the raw data from the EEG recordings in infants. Another aspect of our

This paper was presented at the NIK-2015 conference; see http://www.nik.no/.

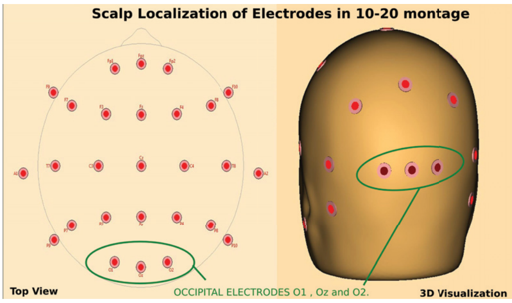

Scalp Localization of Electrodes in 10-20 montage
Top View
OCCIPITAL ELECTRODES O1 , Oz and O2.
3D Visualization

Figure 1: Head-drawing (nose up) and 3D head-model showing scalp localization of the standard 27 electrodes, with green circle showing electrodes O1, Oz, and O2 used in the current study. These electrodes capture EEG activity of the human visual cortex.

method is that it uses just 3 channels (O1, Oz, and O2 as shown in Fig. 1), since we know that VEPs reflect a response in the occipital area, elicited by visual stimuli; this would practically allow the acquisition device for such a biometric to be a far simpler version of the cumbersome tens of electrodes that are usually used to acquire a general brain EEG.

When a subject is exposed to specific external sensory stimulation, it is possible to register on EEG recordings an evoked response. Visual evoked potentials (VEPs) are electrical potentials, recorded by EEG, reflecting activity evoked by visual stimuli. In EEG research these VEPs are elicited by any type of visual stimuli, and are often found over visual areas (the occipital cortex). Looming stimuli in infants will elicit a negative component in the visual cortex.

In this work we show the potential usage of visual evoked responses in biometrics, without the averaging of EEG activity. More specifically, we show how visual evoked potentials (VEP) evoked by looming stimuli can be employed for biometrics.

Related Work

Biometric recognition based on brain electrical activity is a developing field of research. A recent review of existing work on the use of EEG for person recognition is [5]. As the review shows, a substantial part of the efforts to use EEG in biometrics focus on Event Related Potentials (ERP); VEP is the only ERP biometric analysed so far.

Most EEG-related biometrics research has been using the publicly available dataset of [3, 18], which we shall refer to as Zhang-DDBB. This dataset consists of 125 subjects, out of which 77 were alcoholics, and was not created for biometrics research. *Palaniappan and Raveendram* started investigating the use of EEG VEPs for biometrics in [10]. In this work 61 electrodes were used to record brain activity after a visual stimulus and it uses information on the gamma frequency band to classify a dataset of 10 subjects which are not made publicly available. The method achieves a classification performance of 90.95%. The same research group published in 2007 [9] a follow-up work where a Multiple Signal Classification Algorithm computes dominant frequencies of EEGs. The dominant frequencies are the features used in classification. This time, a nearest neighbor approach achieved 97.61% classification rate on 102 subjects from Zhang-DDBB. The work of [8] uses 8 electrodes for VEP analysis, creating a large feature vector by an ensemble of voice processing techniques. This large feature vector is reduced using a correlation-based feature selector and a Support Vector Machine (SVM) is used for classification. In VEP analysis the work achieves two different classification results. The first classification performance is 92.80% for a dataset of 20 subjects from Zhang-DDBB and the second is 61.70% for 122 subjects from Zhang-DDBB.

As mentioned above, the problem with the Zhang-DDBB dataset is that it mainly consists of data from unhealthy (alcoholic) subjects. To quote [5], *'the problem with this practice is that alcohol has been proven to affect various aspects of the EEG (Begleiter & Porjesz, 2006; Zietsch et al., 2007), including the so much used alpha rhythm (Vogel, 2000). Therefore, it is likely that such datasets were biased, overrating the systems' performances. In order to clarify this, a statistical study probing that the parameters used are not affect by the disorder, i.e. assuring they are uncorrelated, must be adjoined.'*

The work of [4] uses Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM). VEP recordings are done by 64 electrodes placed over the scalp. The work collected 1000 VEP trials for 20 subjects (non-public dataset). Classification performance shows that SVM outperforms LDA for VEP analysis. On a two-fold SVM it achieves 91.56% classification accuracy. The works of [8, 4, 9, 10] place electrodes over the entire scalp. All the electrodes are then used for biometric recognition.

The above works assume that all recorded activity on the scalp is a result of the visual stimulus input. This is in contrast to current VEP and ERP literature [7, 2] which focuses on identified responses, e.g. the N2, P300, N170. The brain activation area will vary according to the stimulus. Therefore, proper analysis of ERP responses should be carried out by investigating clearly identified areas in the brain. The brain is divided into different regions. Each region processes a different stimulus input and is responsible for different ongoing activities [2]. Thus, averaging whole brain activation across multiple areas will most likely result in a mix of signals. The research of [4] shows that occipital [1] electrodes produce the most discriminative signals. This finding seems to enforce the methodological points of [7, 2].

The works of [13, 14] do promising EEG VEP analyses while focusing solely on the occipital cortex. Both methods employ checkerboard patterns to generate VEP responses. The first work, [13], has a classification performance of 86.54% on a private dataset composed of 13 subjects. Classification is done by a Linear Discriminant Classifier (LDC). The classifier uses as features well-known ERP responses (P100, N75). The work of [14] focuses on the occipital 'Oz' electrode. The work proposes the use of wave landmarks and a two dimensional Gaussian kernel classifier and achieves 78% classification accuracy on a private dataset composed of 10 subjects.

Contributions

The first contribution is to assess whether a new stimulus can be used to generated VEPs with biometric capabilities. In this work we present the first use of looming stimuli in a biometric study.

Second, a novel method for the detection VEP responses with minimal expert interaction is introduced. EEG data, segmented based on the VEP, are used to create a reliable VEP feature extractor.

Third, we contribute to reproducible research by making our dataset publicly available at `http://www.idi.ntnu.no/grupper/vis/eeg/`. Note that our dataset is not biased by individuals with any known issue; instead it is based on infant subjects who should respond to the stimuli as purely as possible. We thus expect that our dataset (and its

[1] The occipital lobe is the brain region where the visual cortex is located.

planned future extension to hundreds of subjects) could become a standard in evaluation of biometric work using the EEG.

The rest of the paper is organized as follows. Section 2 presents the looming stimulus. Section 3 discusses the pre-processing steps that are useful for EEG analysis. Section 4 presents the segmentation and feature extraction methods of the proposed method. Section 5 discusses the classifier selection. Experimental results are given in Section 6 and conclusions in Section 7.

2 Looming Visual Evoked Potentials

Looming stimuli are defined as approaching (virtual) objects on a collision course with the observer. Such stimuli can be artificially created by displaying an object looming towards the observer, thus allowing for control over the variables and making it easier for the data (psychophysical or behavioural) to be recorded.

Literature shows that a negative VEP component is prevalent in electrodes O1, Oz, and O2 in the visual cortex in relation to a looming stimulus [16]. Brain activity following impending collision of an approaching object can be seen in the visual cortex using EEG [17].

Infants presented with a looming stimulus display VEP components in occipital areas approximately 800 ms before loom hit [16].

In this work 16 infants were exposed to looming stimuli. The EEG recordings are then processed for VEP segmentation and feature extraction. For more details on the dataset see Section 6.

3 Processing of Raw EEG

This Section will present magnitude frequency responses and the reasoning behind the selection of preprocessing filters for EEG recordings.

Digital EEG recordings are preprocessed and filtered differently depending on acquisition procedures and purpose. The main ambition of this work is to assess the biometric capabilities of VEPs. Thus, signal preprocessing needs to keep VEPs intact while removing noise and unwanted data.

The first issue to address is signal corruption caused by utility/mains frequency. It is common for EEG recordings to be dominated by the frequency of AC currents, F_{AC}, which is either $50Hz$ or $60Hz$ depending on region. This signal corruption is visible in the unprocessed EEG readings of Fig. 2. Therefore, the filtering of F_{AC} and its harmonics is vital for subsequent data analysis.

Another issue is artifacts generated by bio-electric flowing potentials [15]. Small body movements, breathing and other similar behaviours can distort EEGs. These distortions create an amplitude modulation in low frequencies. As a consequence, a high pass filter (HPF) with a small cutoff frequency becomes essential.

Literature shows that VEPs have a concise spectrum [17, 16, 1]. This means that information related to the signal is not spread over all frequencies. One would ideally like to filter out frequencies carrying unwanted data. Thus, a low pass filter (LPF) must be employed in addition to the HPF. This filter needs to have a well selected cutoff frequency so as not to impair VEPs.

To further specify filters and respective cutoff frequencies, the present work assessed VEP characteristics. Literature shows [17, 1, 16] that looming VEPs are found between 1.8Hz and 60Hz. To further restrain operational frequencies, we computed the time

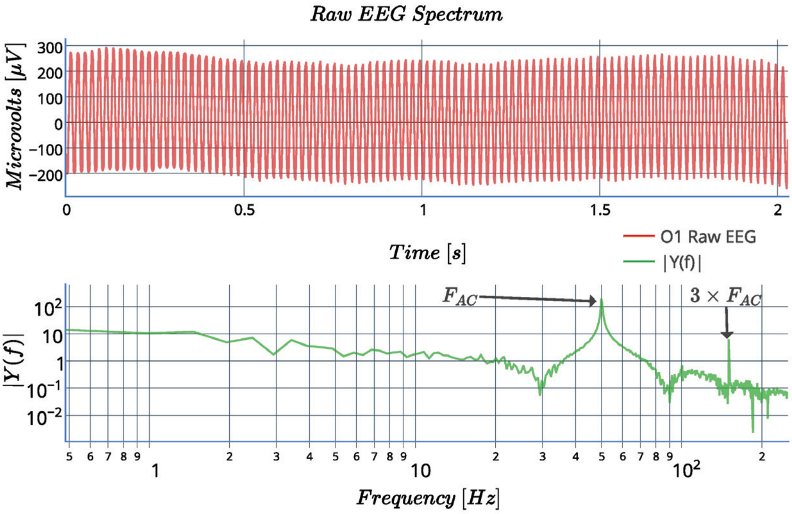


Figure 2: A sample of Raw EEG readings for Electrode 'O1' and its spectrum analysis. The EEG readings are dominated by F_{AC} as shown in both the time and frequency domains.

between peaks and valleys on typical VEP responses of [17]. This showed that VEP frequencies do not exceed 20Hz. Consequently, all harmonic contributions of F_{AC} would be filtered out by a LPF at 20Hz, implying no need for any further filters. However, given the high noise amplitude at 50Hz, a filter at F_{AC} is also needed, to minimise the effect of noise on VEP analysis.

As per the above analysis, three filters are implemented. The three filters are designed as Infinite Impulse Response (IIR), allowing us to create digital counterparts of well established analogs filters [12]. The filter design starts with a first order Butterworth HPF with a cutoff frequency of 1.6Hz; this is a safe cutoff frequency as it is below the $1.8Hz$ lowest VEP frequency reported in the literature. The second filter used is a fourth order Butterworth LPF. This time, the selected cutoff frequency is 20Hz as estimated above. The third filter is a sixth order notch filter. This is a band rejection Butterworth. The rejected band ranges from $45Hz$ to $55Hz$. The magnitude responses of the three designed filters are shown in Fig. 3.

The resulting filtered EEG as well as its spectral analysis are shown in Fig. 4. The processed signal is used for feature extraction, as demonstrated in Section 4.

4 Segmentation and Feature Extraction

Research shows that looming stimuli create distinctive VEP responses. An example of such responses is shown in Fig. 5, which shows VEP responses on 'O1', 'Oz' and 'O2'.

Unfortunately, VEP responses do not always take place consistently. Their timing occurrence varies even within subjects. It is even common for a subject to produce multiple VEPs for a single looming stimulus. We would ideally like to segment out VEP responses before feature extraction, in order to increase the reliability of the latter. In this

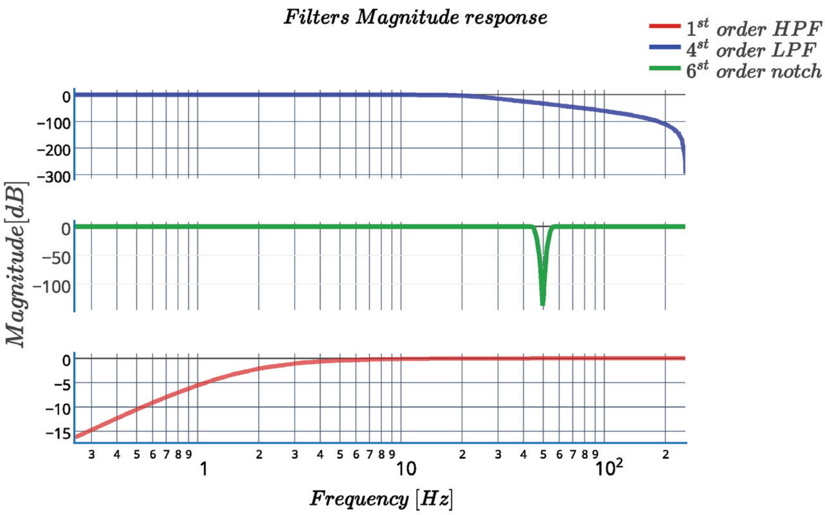


Figure 3: A combination of three filters is used to filter out noise and artifacts from Raw EEG recordings. The magnitude responses of such filters are shown here.

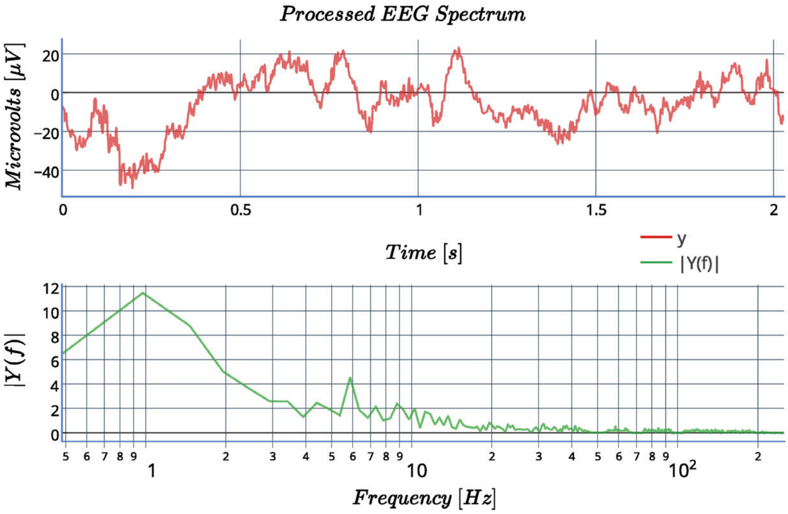


Figure 4: A sample of filtered EEG data for Electrode ’O1’ and its spectrum analysis, after the application of three filters to remove noise and artifacts from the raw signal.

section we present a reliable VEP segmentation technique followed by an inexpensive and dependable feature extraction method.

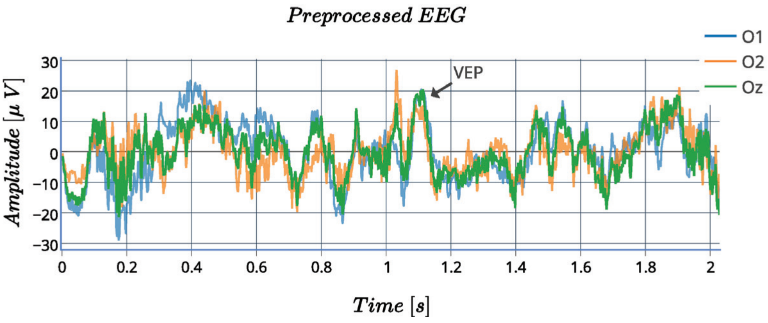


Figure 5: A sample of EEG looming VEP. The graph show readings for electrodes ’O1’, ’Oz’, and ’O2’. These are the electrodes over the occipital cortex.

VEP Segmentation

The potential segmentation algorithm should first list VEP candidates. This is arduous because a single trial can have multiple VEP responses, or none at all, for a given stimulus.

Our detection technique explores characteristics of typical VEP responses. As one can notice in Fig. 5, sharp transitions delineate VEPs. Thus, we can select potential candidates by searching for high frequencies. For this we propose a uni-dimensional edge detector to detect the sharp transitions. This edge detector is instantiated by Eq. 1:

$$\mathcal{E}(x) = \left[\frac{\delta}{\delta x_{ma}} \Phi_{ma}(x, 30)\right]^3 \tag{1}$$

where $\Phi_{ma}(x, N)$ is a moving average of the input EEG signal x with N elements and x_{ma} is the output of the moving average function. Edge detectors are sensitive to noise in input signals. Thus, we first apply the moving average filter Φ_{ma} to the input. This above edge detector can be approximated in the discrete case by the difference Eq. 2:

$$\mathcal{E}[k] = (x_{ma}[k+1] \quad x_{ma}[k])^3 \tag{2}$$

In this discrete version, the computed edge $\mathcal{E}$ is one element smaller than x. This small difference can be ignored in the following steps. We next compute a search vector $\mathcal{S}$ which will allow us to circumvent issues common to looming stimuli. An initial $\mathcal{S}$ is computed by Eq.3:

$$\mathcal{S}(\mathcal{E}_{O1} + \mathcal{E}_{O2} + \mathcal{E}_{Oz}) = \mathcal{G}\left(\left|\frac{\mathcal{E}_{O1} + \mathcal{E}_{O2} + \mathcal{E}_{Oz}}{3}\right|, \sigma, N\right) \tag{3}$$

where $\mathcal{E}$ represents a computed edge in the electrode given in the respective index. $\mathcal{G}(x, \sigma, N)$ represents the convolution of a Gaussian filter with the signal x with standard deviation σ and N represents the size of the convolution filter. For this work a tentative $sigma = 5$ and $N = 15$ were employed.

No looming related activity happens in the initial 500ms. We thus replace the first half second of the search vector by its average value. Finally the search vector $\mathcal{S}$ is normalized to the interval $[0, 1]$. An example of a resulting search vector is shown in Fig. 6.

We can now look for VEP candidates by searching for peaks (local maxima) in $\mathcal{S}$. Peak candidates need to have a minimal distance of $200ms$ to each other and we only accept peaks higher than 0.95; if these criteria are fulfilled the VEP candidates are treated as VEP indicators.

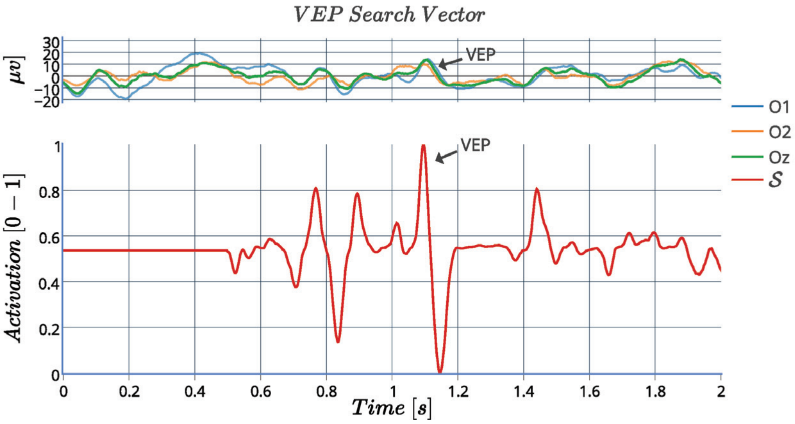


Figure 6: The first graph shows the EEG readings. These readings were processed with a moving average filter. The second graph shows the search vector. In $\mathcal{S}$ the peaks indicate VEP candidates.

With a list of VEP indicators it is possible to segment and extract features. Segmentation extracts a 200ms long signal sample. The segmented signal starts 80ms before the VEP indicator and ends 120ms after it. Thus, all information from the VEP is represented in the segmented signal. This segmented data is next used for feature extraction.

Feature Extraction

For feature extraction we create a signal 400ms long. Two slices of 200ms compose this signal. The first slice is the preprocessed EEG recording and the second is the computed edge detector $\mathcal{E}$. For concise results we normalise to the range $[0, 1]$ for each 200ms slice independently. Fig. 7 shows the two slices that compose the feature vector.

Using the above signals, the feature extraction process is computationally inexpensive, while it concisely inherits information on VEPs. The result is an efficient and robust feature extractor.

5 Matching / Classification

For classification we use a simple feed-forward Neural Network. A one-hidden-layer Multi Layer Perceptron (MLP) is trained with the standard back-propagation algorithm [6]. This work employs stochastic gradient descent with mini batches for training, ensuring a faster training time. The MLP uses the sigmoid function for neuron activations and soft-max for the output neurons.

The EEG recording apparatus had a sampling frequency of 500Hz. Therefore, the 400ms feature vector is in fact a 200 dimension vector. We thus create a neural network with 200 input neurons. These neurons are fully connected to a hidden layer with 15 neurons. These 15 neurons are connect to the output layer. The number of output neurons

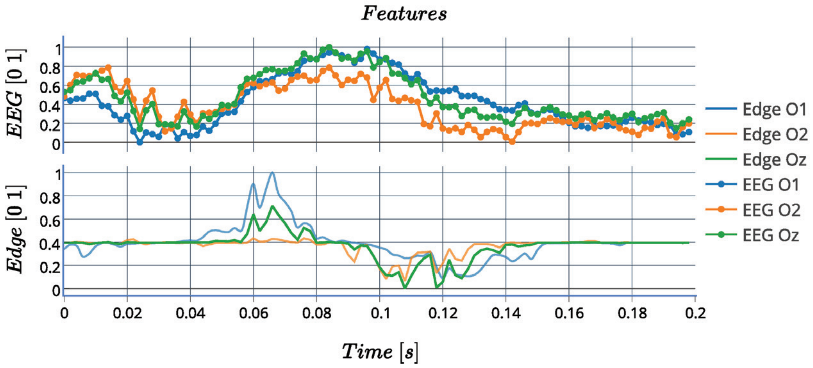


Figure 7: The first graph shows normalized EEG readings; these 200ms signals are the result of the proposed VEP detection and segmentation. The second graph shows edge features computed over the EEG signals. Together these two 200ms slices compose a Feature Vector. The Feature Vector can be computed for any electrode. Here we use electrodes over the visual cortex: 'O1','Oz', and 'Oz'.

is the same as the number of subjects in the dataset. Thus, for this experiment we use 16 output neurons. The MLP was implemented with the help of a publicly available library [11].

As a single subject generates many feature vectors, a final classification is done by majority voting the MLP outputs.

6 Experimental Results

This Section presents the experimental results as well as details of the acquired dataset.

Dataset and Acquisition Procedure

The data were initially acquired from 16 infants. The infants would arrive with their parent(s) and be familiarised with the surroundings. Meanwhile, the parent(s) would be informed about the experiment and signed a consent form. The EEG net was prepared in electrolyte solution to ensure good impedance; the recording equipment was prepared before the participant arrived. The net would be applied and the infants would be seated in front of the screen; a parent would always be with the infant in the test room. The experiment would be aborted if an infant became too fussy or started crying.

Infants were presented with a black approaching rotating circle consisting of four smaller circles $\frac{1}{3}$ of the size (red, green, blue, yellow) rotating within the black circle on a white background. The loom rotated with a constant angular velocity of $300°/s$ and started at a virtual distance of 43.1 m giving a visual angle of 5° (6cm diameter); it then grew to a maximum size of 131° (350cm diameter). The loom would approach the infants at constant accelerations at three different speeds with a duration of 2 seconds ($21.1\frac{m}{s^2}$), 3 seconds ($9.4\frac{m}{s^2}$), and 4 seconds ($5.3\frac{m}{s^2}$). The looming object would move the same distance in all three cases, as well as in reverse.

The order of the presented stimuli was randomly generated. The only constraint was that consecutive stimuli should be different. There was a break of 1.5 seconds between two stimuli. In this work looming with two second durations are used for classification. We only employ information from the occipital electrodes 'O1', 'O2' and 'Oz'.

In the acquired dataset we have a varying number of looming stimuli with 2 second duration. Some infants were presented up to 20 stimuli and some just 7. The number of trials depended on the willingness of each baby.

Classification Results

The pipeline described in this paper is followed, i.e. pre-processing the EEG recordings as described in Section 3, segmentation of the information related to looming with a 2 second duration and feature extraction as described in Section 4. The Feature Vector is then used in the MLP Voting classifier described in Section 5 to obtain the classification results.

We use half of the available data for training, and the other half for testing, obtaining a recognition (classification) accuracy of 62.50%. This result is comparable to the state of the art occipital VEP EEG processing [13, 14]. Different from these previous works, the dataset assessed in this paper has more subjects and is composed of a much smaller number of trials, making it a harder evaluation scenario. This explains why the evaluation performance achieved here is smaller than [13, 14], which used private datasets. Table 1 shows classification performance as well as details of the assessed datasets.

Method	Subjects	VEP trials	Acc'cy	N- Folds
Proposed	16	7 to 20	62.50%	2
[13]	13	120	86.54%	4
[14]	10	200	78.00%	2

Table 1: Classification accuracy

7 Conclusion

The first attempt to assess whether looming stimuli can be used to generate occipital Visual Evoked Potential (VEP) in EEG signals with biometric capabilities, is presented. At the same time the first public dataset of EEG responses to looming stimuli is being made available. This dataset is based on infant subjects that should respond to the stimuli as purely as possible. We expect that this dataset and its planned extensions can become a standard for biometric studies of VEP.

We show that our VEP segmentation methodology, as well as our proposed feature extraction, can produce biometrically capable EEG based features. The achieved classification performance is comparable to the state of the art biometric studies using occipital VEPs. However, these results were achieved using a dataset with more subjects and with fewer samples (VEP trials) per subject.

Further work could address automatically learned features (e.g. with a deep learning approach), but this would require at least two orders of magnitude more data. Another possibility is to investigate other machine learning techniques capable of learning while using a small numper of samples (VEP Trials).

References

[1] S. B. Agyei, M. Holth, F. R. van der Weel, and A. L. H. van der Meer. Longitudinal study of perception of structured optic flow and random visual motion in infants using high-density EEG. *Developmental Science*, 2014.

[2] M. F. Bear, B. Connors, and M. Paradiso. *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, fourth international edition, March 2015.

[3] H. Begleiter. Eeg database data set. `http://archive.ics.uci.edu/ml/datasets/EEG+Database`. Accessed: 2015-03-23.

[4] K. Das, S. Zhang, B. Giesbrecht, and M. P. Eckstein. Using rapid visually evoked EEG activity for person identification. *IEEE Engineering in Medicine and Biology Society. Conference*, 2009:2490–2493, 2009.

[5] M. Del Pozo-Banos, J. B. Alonso, J. R. Ticay-Rivas, and C. M. Travieso. Electroencephalogram subject identification: A review. *Expert Systems with Applications*, 41:6537–6554, 2014.

[6] Y. Le Cun, D. Touresky, G. Hinton, and T. Sejnowski. A theoretical framework for back-propagation. In *The Connectionist Models Summer School*, volume 1, pages 21–28, 1988.

[7] S. J. Luck and E. S. Kappenman, editors. *The Oxford Handbook of Event-Related Potential Components (Oxford Library of Psychology)*. Oxford University Press, 1 edition, December 2011.

[8] P. Nguyen, D. Tran, X. Huang, and D. Sharma. A Proposed Feature Extraction Method for EEG-based Person Identification. In *The International Conference on Artificial Intelligence*, 2012.

[9] R. Palaniappan and D. P. Mandic. Biometrics from brain electrical activity: A machine learning approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):738–742, 2007.

[10] R. Palaniappan and P. Raveendran. Individual identification technique using visual evoked potential signals. *Electronics Letters*, 38(25):1634–1635, Dec 2002.

[11] R. B. Palm. Prediction as a candidate for learning deep hierarchical models of data. Master's thesis, 2012.

[12] T. W. Parks and C. S. Burrus. *Digital Filter Design*. Wiley-Interscience, New York, NY, USA, 1987.

[13] A. Power, E. Lalor, and R. Reilly. Can visual evoked potentials be used in biometric identification? In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 5575–5578, Aug 2006.

[14] G. K. Singhal and P. Ramkumar. Person identification using evoked potentials and peak matching. *2007 Biometrics Symposium, BSYM*, 2007.

[15] M. Teplan. Fundamentals of eeg measurement. *Measurement Science Review*, 2(2):1–11, 2002.

[16] A. L. H. van der Meer, M. Svantesson, and F. R. van der Weel. Longitudinal study of looming in infants with high-density EEG. *Developmental Neuroscience*, 34(6):488–501, 2012.

[17] F. R. van der Weel and A. L. H. van der Meer. Seeing it coming: infants brain responses to looming danger. *Naturwissenschaften*, 96(12):1385–1391, 2009.

[18] X. L. Zhang, H. Begleiter, B. Porjesz, W. Wang, and A. Litke. Event related potentials during object recognition tasks. *Brain Research Bulletin*, 38:531–538, 1995.

D LOOKING BEYOND APPEARANCES: SYNTHETIC TRAINING DATA FOR DEEP CNNS IN RE-IDENTIFICATION

Igor Barros Barbosa, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen, and Theoharis Theoharis. “Looking beyond appearances: Synthetic training data for deep CNNs in re-identification.” In: *Computer Vision and Image Understanding* 167 (2018), pp. 50 –62. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2017.12.002. URL: http://www.sciencedirect.com/science/article/pii/S1077314217302254

Looking Beyond Appearances: Synthetic Training Data for Deep CNNs in Re-identification

Igor Barros Barbosa[a], Marco Cristani[b], Barbara Caputo[c], Aleksander Rognhaugen[a], Theoharis Theoharis[a]

[a] *Norwegian University of Science and Technology (Norway),*
[b] *University of Verona (Italy)*
[c] *Sapienza Rome University(Italy)*

Abstract

Re-identification is generally carried out by encoding the appearance of a subject in terms of outfit, suggesting scenarios where people do not change their attire. In this paper we overcome this restriction, by proposing a framework based on a deep convolutional neural network, SOMAnet, that additionally models other discriminative aspects, namely, structural attributes of the human figure (e.g. height, obesity, gender). Our method is unique in many respects. First, SOMAnet is based on the Inception architecture, departing from the usual siamese framework. This spares expensive data preparation (pairing images across cameras) and allows the understanding of what the network learned. Second, and most notably, the training data consists of a synthetic 100K instance dataset, SOMAset, created by photorealistic human body generation software. SOMAset will be released with a open source license to enable further developments in re-identification. Synthetic data represents a cost-effective way of acquiring semi-realistic imagery (full realism is usually not required in re-identification since surveillance cameras capture low-resolution silhouettes), while at the same time providing complete control of the samples in terms of ground truth. Thus it is relatively easy to customize the data w.r.t. the surveillance scenario at-hand, *e.g.* ethnicity. SOMAnet, trained on SOMAset and fine-tuned on recent re-identification benchmarks, matches subjects even with different apparel.

Keywords: Re-identification, deep learning, training set, automated training dataset generation, re-identification photorealistic dataset

1. Introduction

Re-identification (re-id) aims at matching instances of the same person across non-overlapping camera views in multi-camera surveillance systems [1]. Initially a niche application, re-id has attracted huge research interest and has been the focus of thousands of publications in the last five years, although current solutions are still far from what a human can achieve [2].

Recently, deep learning approaches have been customized for re-identification, notably with the so-called *siamese* architectures [3, 4, 5, 6, 7, 8]. In a siamese network, a pair of instances is fed into the network, with a positive label when the instances refer to the same identity, negative otherwise. This causes the network to learn persistent visual aspects that are stable across camera views. An issue with this setting is the setup of the training data: positive and negative pairs should be prepared beforehand, with a significant increase in complexity.

The majority of re-id approaches focuses on modeling the appearance of people in terms of their apparel, with the obvious limitation that changing clothes between camera acquisitions seriously degrades recognition performance. The RGB-D data provide significantly more information, which explains why there has been considerable progress in this case [9, 10], but, on the other side, current RGB-D sensors cannot operate at the same distance as typical surveillance cameras; therefore, focusing on RGB does remain an important challenge.

In this paper, we present a re-identification framework based on a convolutional neural network, with the aim of facing the above issues. The framework exhibits several advantageous characteristics.

First, the structure of the network is simpler than a siamese setup. It is based on the Inception architecture [11], and is used as a feature extractor. This is similar to a recent approach proposed by [12], which also opted for an Inception-based network architecture. As a by-product, probing inner neurons of deep layers to understand what is learnt by the network is easier than in siamese-like designs. In particular, we show that the network is able to capture structural aspects of the human body, related to the somatotype (gender, being fat or lean, etc.), in addition to clothing information. For this reason, we dubbed the network SOMAnet.

The second unique characteristic of our framework is the data used to train SOMAnet: for the first time we employ a completely synthetic dataset, SOMAset, to train our net-

Email address: igor.barbosa@ntnu.no (Igor Barros Barbosa)

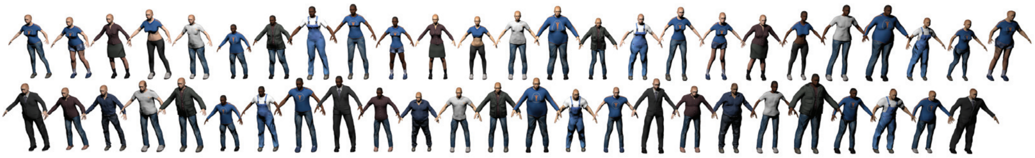


Figure 1: Renderings of the 50 human prototypes in SOMAset, each one of them wearing one of the 8 sets of clothing available. The top row shows the 25 female subjects and the bottom row the 25 male subjects.

work from scratch. SOMAset consists of 100K 2D images of 50 human prototypes (25 female and 25 male, Fig. 1), created by mixing three somatotypical "seeds" [13]: *ectomorph* (long and lean), *mesomorph* (athletic, small waist) and *endomorph* (soft and round body, large frame), and accounting for different ethnicities. Each of these prototypes wears 11 sets of clothes and assumes 250 different poses, over an outdoor background scene, ray-traced for lifelike illumination. Training networks with synthetic data is not a totally new concept, with pioneering works in 3D object recognition such as [14, 15, 16]. However, rendering images of human avatars as a proxy for dealing with real images of people has no precedent in the re-identification literature. We show in our experiments that this choice is effective: when SOMAnet is trained with SOMAset and fine-tuned with other datasets, it achieves state-of-the-art performance on four popular benchmarks: CUHK03 [3], Market-1501 [17], RAiD [18] and RGBD-ID [10].

Third, on the RGBD-ID dataset, we are able to show the capability of SOMAnet in recognizing people independently of their clothing, based only on RGB data. This is the first such attempt in the literature, surpassing previous approaches that additionally used depth features.

The rest of the paper is organized as follows: after reviewing relevant previous work (Sec. 2), we describe SOMAset (Sec. 3) and the SOMAnet architecture (Sec. 4). Sec. 5.2.2 describes our strategy for probing the inner neurons of the deep layers. Sec. 5 reports on an exhaustive set of experiments, illustrating the power of the SOMA framework. Sec. 6 concludes with a summary and by sketching future work.

2. Related literature

In this section we review the recent literature on re-identification, focusing in particular on deep learning techniques. We also discuss the recent trend of creating training sets for recognition, with emphasis on the re-id task.

2.1. Re-identification

Strategies for re-identification are many and diverse, with brand new techniques being presented at a vertiginous pace: at the time of writing, Google Scholar gives more than a thousand papers published since January 2016; therefore, it is very hard to recommend an updated survey, the most recent dating back to 2014 [19].

The early works on re-id were designed to work on single images [20]; batches of multiple images have been considered afterwards [21]. On this input, discriminative signatures are extracted as manually crafted patterns [21] or low-dimensional discriminative embeddings [22]. While most of the signatures focus on the appearance of the single individuals independently on the camera setting, the study of the inter-camera variations of color (and illumination) gives also convincing results [23, 18, 24, 25]. Signatures can be matched by exploiting specific similarity metrics [26], which are learned beforehand, thus casting re-id as a metric learning problem [27, 28, 29, 30, 31].

Traditionally, re-id assumes that people do not change their clothes between camera acquisitions. The common motivation is that re-id is a short-term operation, thought to cover a time span of minutes/few hours max, that is, the time necessary for a person to walk between cameras in an indoor environment (an airport, a station etc.). In reality, even in such a time-span, an individual can change his appearance, for example taking off a jacket due to the heating, wearing a backpack etc. Few approaches cover the clothing-change scenario [9, 10], all of them relying on RGB-D data. This work is the first that captures structural characteristics of the human figure, in addition to clothing information, exploiting mere RGB data.

The very recent re-id approaches incorporate deep network technology. Typically, they consider image pairs as basic input, where each image comes from a different camera view: when the two images portray the same individual, a positive label is assigned, negative otherwise. These pairs are fed into the so-called *siamese* or *pseudo-siamese* networks, which learn the differences of appearance between camera acquisitions [3, 4, 5, 6, 7, 8]. A very recent alternative is the use of triplet loss, where three or more images are compared at the same time [32, 33]. One image is selected as anchor, while the remaining two images are divided into a positive (having the same identity as the anchor) and a negative one. The objective function over triplets correlates the anchors and the positive images, minimizing their distance. Conversely, the distance from the anchors to the negative images is maximized. Triplet loss has also been used in non-deep learning methods to learn an ensemble of distance functions that can minimize the rank for perfect re-identification [34]. One disadvantage of the siamese and triplet loss methods is that they require the dataset to be prearranged in terms of labels. This is cumbersome, may cause highly unbalanced target class distributions and even increase computational complexity.

We argue that re-identification can be carried out by simpler network architectures combined with similarity measurements. This idea was also successfully demonstrated by Sun et al. in 2014 for face verification [35] where they extract a descriptor using a single path network.

The basis of the proposed approach is to employ a simple (single path) network to learn a descriptor, using a synthetic dataset for training, before fine-tuning on the training partition of a specific dataset. In addition, a probing approach allows the investigation of the characteristics of the network.

2.2. Training data generation

In the last years, the design of training sets for recognition has changed from a mostly human-driven operation (crowdsearching data in the most advanced attempts [36]) to a proper research field aiming at automatically producing samples spanning the whole visual semantics of a category, with numbers sufficient to deal with deep learning requirements [37, 38, 39, 40]. Two main paradigms do exist: the former assumes that good training data is available on the Internet, and aims at creating retrieval techniques that bridge lexical resources (as *Wordnet*) with the visual realm [37, 38, 39]. The latter, most recent, direction assumes that web data is too noisy or insufficient (particularly in the 3D domain) and relies on the generation of photorealistic synthetic data [40]. In this case, the trained classifiers should be adapted to the testing situation by attribute learning [41], domain adaptation [42] or transfer learning [43]. This direction seems to be very promising, especially in conjuction with deep architectures [14, 15, 16].

In the re-identification field, the only work that considers the augmentation of a training set by synthetic data is that of [44], substituting the background scene of the training images with different types of 2D environments. This has been shown to help in reducing the dataset bias, favouring cross-dataset performance. Unfortunately, illumination is not natural in the synthesized samples, and the strategy cannot easily be applied to any dataset (foreground/background segmentation is necessary). Our work goes in the opposite direction, focusing on photorealistic images of the foreground subjects instead of the scenery (which in our case consists in a single, large, outdoor scenario).

3. The SOMAset dataset

In this section we present SOMAset[1], describe the protocol followed for creating it and discuss the features that make it unique compared to other existing re-id collections.

The human figure is normally defined as a mixture of three main somatotypes [13]: *ectomorph* (long and lean), *mesomorph* (athletic, small waist) and *endomorph* (soft and round body, large frame). We account for these facets using an open-source program for 3D photo-realistic human design, *Makehuman*, and a rendering engine, *Blender*. Starting from a generic 3D human model we created 25 male and 25 female subjects, by manually varying the height, weight and body proportions for each subject so as to represent mixtures of the three aforementioned somatotypes. In order to further improve the similarity to real acquisitions, we also slightly varied parameters like symmetry and the size of legs and/or arms, so as to better simulate natural body variations.

In almost all previous re-identification scenarios, it is assumed that subjects do not change their clothes between camera acquisitions. Re-identification datasets adhere to this assumption, associating identity to appearance (a particular apparel represents a single subject). With SOMAset, we relax this constraint, rendering each of the 50 subjects with 8 different sets of clothing: 5 of these were shared across the sexes while 3 each were exclusive for males / females (thus in total there are 11 types of outfit). In this way, we stimulate the network to focus on morphological cues, other than mere appearance. Experiments with the RGBD-ID dataset (Sec. 5.2.3) confirm this, having people wearing different clothing between acquisitions.

In more detail, the 3 clothing variations dedicated to females are: T-shirt with shorts; blouse with skirt; sport top with leggings. The 3 male clothing variations are: suit; striped shirt with jeans; shirt with black trousers.

The shared clothing category includes the following 5 variations: white t-shirt with jeans; long sleeve shirt with jeans; blue T-shirt with jeans; jacket over shirt with jeans; overalls. Fig. 1 shows renderings of the 50 subjects, with female and male subjects in the top and bottom row respectively. The first 8 columns show the 8 clothing possibilities for each gender. To account for ethnicity variations, different skin colors were mapped onto the subjects. Out of the 50 subjects, 16 received Caucasian skin, 16 have darker skin tones, while the remaining 18 have beige skin tones to model Asian types. We did not include further variations (*e.g.* structural) of the faces and we did omit hair styles, to bound the number of possible variations. Notably, adopting more types of garments does not seem to affect the performance drastically, after some preliminary experiments, not reported here for the lack of space.

Each of the 400 subject-clothing combinations assumed 250 different poses. These poses are extracted

[1] SOMAset will be released with a open source license to enable further developments in re-identification.

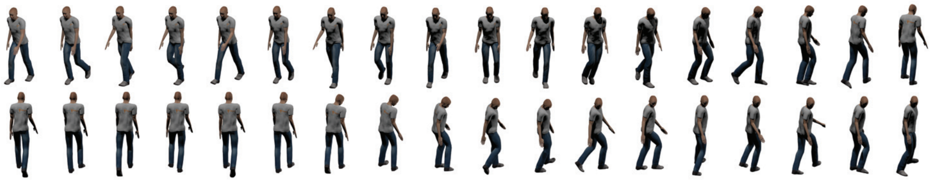


Figure 2: Renderings of a specific subject-clothing assuming 36 out of 250 possible poses. Note the change of the orientation w.r.t. to the camera.

from professionally-captured human motion recordings, provided by the CMU Graphics Lab Motion Capture Database [45]. We opted for extracting poses from a recording titled 'navigate', where the subject walks forwards, backwards and sideways. A sample of 36 poses from a specific subject-clothing combination is shown in Fig. 2.

Each of the resulting $N = 100K$ subject-clothing-pose combinations ($N = 50$ *subjects* $\times$ 8 *clothing sets* $\times$ 250 *poses*) is placed in a realistic scene (see below) and captured by a virtual camera with a randomly chosen viewpoint, following a uniform distribution. Specifically, we place the subject in a random location over the floor of the scene, and we take 250 different viewpoints uniformly spanning a hemisphere centered 8 meters away from the subject's initial position. This induces a distance varying from 6 to 10 meters between the camera and the rendered subject-clothing-pose.

The different camera viewpoints generate people with diverse image occupancy, different lighting patterns and relative pose w.r.t. the observer. A structured outdoor scene was created for rendering, which covers an area of approximately 900 m^2, where each of the 100K instances was located. The scene includes trees, buildings, pavement, grass and a vehicle, giving a certain variability as the viewpoint changes. A small collage of images from male subjects of SOMAset is shown in Fig. 3.

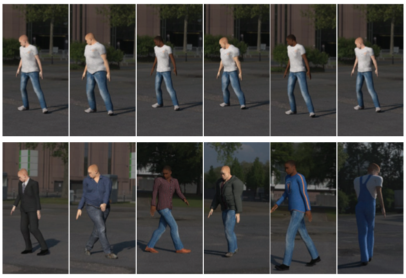


Figure 3: Images sampled from SOMAset. Different male subjects are represented in each column. The second row shows examples where the subjects' pose and clothing vary at the same time. We see that the single 3D environment does yield background variability.

4. The SOMAnet architecture

SOMAnet is a deep neural network that can compute a concise and expressive representation of high level features of an individual, portrayed in an RGB image. This representation enables simple yet effective similarity calculations.

SOMAnet is based on the Inception V3 modules [46], that proved to be well-suited to work on synthetic data [47]. Experiments conducted using other frameworks such as Alexnet [48], VGG16/VGG19 [49], Inception V1 [11] and Inception V2 [50] confirmed this. The architecture of SOMAnet is described in Sec. 4; the motivation for our architectural choices are discussed in Sec. 4.1.1 and Sec. 4.1.2. Subsequently, we present the pipeline for training SOMAnet from scratch in Sec. 4.2, and the fine-tuning strategy to customize it to diverse testing scenarios, together with the re-identification algorithm, in Sec. 4.3.

4.1. Architecture

Our architecture follows closely the Inception V3 model [46] (Fig. 4): the initial sequence of convolutions and max pooling replicates the original architecture. These are followed by two cascading Inception modules and a modified Inception module (Reduced Inception Module) that reduces the input data size by a half by using larger strides in the 3×3 convolution and in the pooling layer. Moreover, it drops the 1×1 convolution windows that would be used as output of the inception module. The network proceeds to a fourth inception module providing data to our last layers; a max pooling layer followed by a convolution layer that feeds the fully connected layer leading to the output softmax layer.

The use of 3×3 windows is preferred over other window sizes, because they are more computationally efficient than larger convolutions used in previous works. A cascade of 3×3 convolution windows can provide a proxy for the analysis derived by 5×5 and 7×7, which were used in [11, 50]. The convolution layers in our network uses rectified activation units (ReLUs) [51] which have sparse activation and efficient gradient propagation as they are less affected by vanishing or exploding gradients. Unlike previous Inception networks [11, 50, 46] our fully connected layers employ the hyperbolic tangent as activation unit.

We performed a toy experiment on SOMAset + SOMAnet to sense the complexity of a re-id task where a

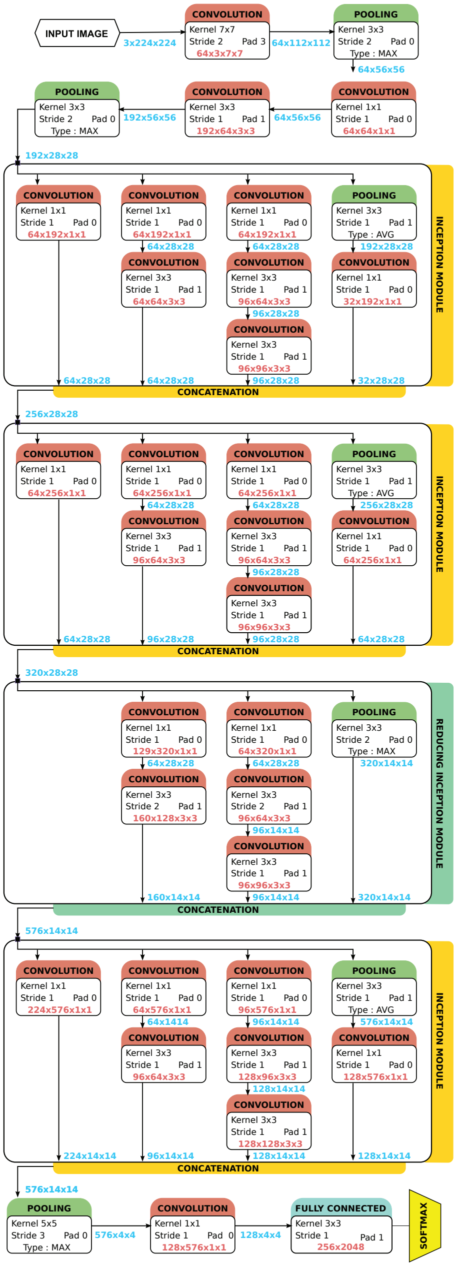


Figure 4: SOMAnet architecture; each layer has its respective activation outputs presented in blue text.

given synthetic subject (out of the gallery of the 50 different identities) can wear different clothes. The 100,000 images have been partitioned into training (70% of the total images), validation and testing sets (15% each); the validation and testing sets were provided so that users of our dataset can assess problems in training, such as overfitting and underfitting. The images have been sampled so to have a proportional number of instances for each subject within each partition. The re-identification results, following the algorithm explained in Sec. 4.3, in terms of Cumulative Matching Characteristic (CMC) curve recognition rate at predefined ranks, are reported in Table 1. We see that this training strategy allows to get an adequate descriptor for the somatotypical characteristics of the subjects, giving 79.69% rank 1 success rate on the testing set.

SOMAset		
Set	Rank 1	Rank 5
Validation	99.77%	100.00%
Testing	79.69%	99.75%

Table 1: SOMAnet classification performance on the rendered SOMAset.

4.1.1. Difference to GoogLeNet

The GoogLeNet inception network [11, 50, 46] was designed for the Large Scale Visual Recognition challenge [52]. Hence, it needed to be deep enough to learn abstract features able to differentiate up to a thousand different classes. Such deep architecture might be unnecessary for more specific image recognition tasks. The original design of GoogLeNet also presents three objective functions, conceived to help with gradient propagation as the network becomes deeper.

To assess the appropriate depth of the Inception architecture in the case of the re-id task, we used the rendered SOMAset. We designed a classification task, where the original GoogLeNet network must correctly classify all the subjects' identities. We used the experimental setting described above.

The original GoogLeNet was trained until the validation set reached a plateau for all its three objective functions. Results showed that, for the specific task of re-id using SOMAset, there was no performance gain by using the deeper classification stages. The network was thus re-designed to use only four Inception V3 modules. Consequently, the network did not need multiple outputs to help in gradient propagation (Fig. 4).

SOMAnet also differs from previous versions of GoogLeNet in the fully connected layer, where the hyperbolic tangent is used as the activation function, because it is zero-centered and has a bounded output space. The output of the fully connected layer of SOMAnet produces a vector $\mathcal{X} \in \mathbb{R}^{256}$ within $[-1, 1]$. This enforces a new embedding computation, with a dimensionality reduction from 2048 to 256 dimensions.

4.1.2. Difference to Siamese networks

Siamese networks have been successfully employed in re-identification by reformulating the task as a binary classification problem [3, 4]. Because the input space of siamese networks is expanded from one image to two, complexity challenges arise when training such networks on large datasets. Space requirements increase as the square of the input images. Thus, it becomes infeasible to process the complete set of combinations during training time and one needs to select which image pair samples to use in order to have a balanced training set. In the case of pseudo-siamese networks, training resources must be spent in learning the convolution weights of each different input branch. This effectively makes the network wider in shallow layers.

Our architecture does not suffer from these issues (Fig. 4). It computes a compact latent embedding space (in our case a $\mathbb{R}^{256}$ descriptor) and we thus use the network as a feature extractor, with linear space and time requirements. Given the descriptor, similarity distances are setup and evaluated to perform re-id, as explained in Sec. 4.3 .

4.2. Training Phase

SOMAnet is trained using backpropagation [53] to minimize the cross-entropy objective function.

Parameters are optimized using a mini-batch gradient descent method with momentum and weight decay [54]. This type of training strategy has been shown to be effective [48, 11, 49]. Here, 32 images are used per mini-batch. The SOMAnet model uses the Xavier initialization of weights, which is a good starting point for deep neural networks [55].

The cross-entropy objective function expects that both the target and predictive outcomes are probability distributions. This constraint can be achieved by encoding target outcomes as one-hot vectors, while the predictive outcomes produced by the neural network can be transformed into a distribution by using the softmax function as seen in Fig. 4. The learning rate is initialised as $\alpha = 0.1$ and is reduced by a factor of 10 whenever the objective function reaches a plateau. We can successfully use high learning rates because the proposed model uses batch normalization, as presented in [50]. SOMAnet was trained using the Caffe package [56] on an NVIDIA GeForce GTX TITAN X GPU. Training the architecture on the full SOMAset took 3 hours.

4.3. Adapting SOMAnet to Real Data and Re-identification

To deal with testing sets made up of real people, domain adaptation strategies have to be included, that in the case of deep neural networks amounts to fine-tuning [57]. Specifically, we force the fine-tuning to focus on the actual classification task (softmax layer). We expect this to help avoid over-fitting of shallower layers, while at the same time giving the chance to obtain strong results with little target data. We also want to avoid the layer specificity problem [58]. Therefore, we allow fine-tuning to take place in shallower layers but with smaller learning rates. This forces the transferred SOMAnet to comply with the new re-id task by only changing the initial layers a little. The fine-tuning protocol is summarized here:

1. **Transfer SOMAnet to a new task** by replicating all layers except for the final softmax layer.
2. **Set the learning rates** of all layers preceding the softmax layer to be ten times smaller than that of the final layer.

After fine-tuning, the output of the penultimate layer is used as feature descriptor. This vector individuates a $\mathbb{R}^{256}$ latent space which is bounded within $[-1, 1]$ and represents an embedding suitable for efficient distance computations. To add more invariance to the descriptor, we mirror the input image, extract another 256-dimensional feature vector and concatenate it with the original one, obtaining a 512-dimension descriptor.

To compute the distance between descriptor $\mathbb{F}_Q$ and $\mathbb{F}_G$ of a query image Q and a gallery image G, we opted for the cosine distance. In preliminary tests this distance function was shown to be effective. The distance between descriptors gives the output of the re-id, that is, the rank of the gallery images w.r.t. the distance to the query sample.

5. Experiments

In this section we explore the potential of SOMAnet and SOMAset, focusing on different aspects of the network and analysing the contribution of the dataset, by performing quantitative and qualitative experiments. After describing the benchmarks used (Sec. 5.1), we focus on SOMAnet (Sec. 5.2), we describe experiments illustrating its performance against other deep architectures (Sec. 5.2.1), we show how some neurons encode specific features of humans (Sec. 5.2.2) and how a synthetic training dataset has a positive impact on a deep architecture (Sec. 5.2.3). Then, we consider SOMAset (Sec. 5.3), illustrating its role in increasing re-id performance (Sec.5.3.1), and exploring how different versions of SOMAset (different number of subjects and poses) change the recognition scores (Sec. 5.3.2 and Sec. 5.3.3, respectively).

5.1. Datasets

We briefly present here the four datasets that we focus on, highlighting the different challenges they represent in terms of re-id. For comparative purposes, for each dataset we consider state-of-the-art peer-reviewed methods in terms of the Cumulative Matching Characteristic (CMC) curve, and the mean Average Precision (mAP).

5.1.1. CUHK03

CUHK03 is the first person re-id dataset large enough for deep learning [3], with an overall 13164 images. It consists of 1467 identities, taken from five cameras with different acquisition settings. Each identity is observed by at least two disjoint cameras.

The images are obtained from a series of videos recorded over months, thus incorporating drastic illumination changes caused by weather, sun directions, and shadow distributions (even considering a single camera view). As usual in the literature, we consider here the dataset version where pedestrians in the raw images are manually cropped to ease the re-id.

The evaluation of re-id performance on this dataset follows two protocols, one for single-shot and another for multi-shot. In the single-shot case, we follow the protocol of [3], commonly adopted in the literature: the dataset is partitioned into a training set of 1367 identities and a test set of 100 identities; during evaluation, we randomly take one of the test set images from each identity of a camera view as probe, using another camera for the corresponding gallery set images (where there exists one image of the same identity as the probe). In the multi-shot case, there are multiple images of each identity in both the probe and gallery sets. We thus compute the average distance from all the probe images w.r.t. all images of the gallery set, producing a ranking. All the experiments are evaluated with 10 cross validations using random training/test set partitions.

5.1.2. Market-1501

Market-1501 is the largest real-image dataset for re-id so far, containing 1501 identities over a set of 32668 images, where each image portrays a single identity [17]. Five high-resolution and one low-resolution camera were used in the dataset acquisition. Each identity is present in at least two cameras. The dataset is partitioned as follows: the training set consists of 750 identities and 12936 images; these are the images used for training/fine-tuning SOMAnet. The remaining 751 identities are contained in a test set of 19732 images, i.e. 3368 query images which are matched against a gallery set of 16364 images $(19732 - 3368)$.

The testing protocol has been specified in [17], and the code for the perfomance evaluation has been provided by the authors. In the single-shot re-id modality, each query image is compared against the gallery images, excluding those that refer to the subject captured by the same camera view (for each query image, there are an average of 14.8 cross-camera ground truths). The mean average precision (mAP) metric is employed, since it is capable of measuring the performance with multiple ground truths. The dataset also contains extra sets of images for each of the 3368 identities in order to allow testing a multi-shot scenario.

5.1.3. RAiD

RAiD (Re-identification Across indoor-outdoor Dataset) is a 4-camera dataset where a limited number of identities (41) is seen in a wide area camera network [18]. The images of RAiD have large illumination variations as they were collected using both indoor (cameras 1 and 2) and outdoor cameras (cameras 3 and 4). The protocol for re-id is the following: the subjects are randomly divided in two sets, training (21 identities) and testing (20 identities). In total there are 6920 images, for an average of around 161 images per identity. The training set (around 3300 images in total, depending on the chosen subjects) is used to fine-tune SOMAnet; this represents a challenge due to the small number of data, when compared to the other, more recent, repositories.

For evaluating the multi-shot modality, 10 images for each test identity are picked as query from a single camera and the images associated with a different camera are used as gallery set. Specifically, we evaluate the camera pairs 1-2, 1-3, 1-4, where the latter two configurations have large inter camera illumination variations Evaluation is done using five cross-validation rounds. For each round, a new random identity partition is made for creating the training and test set, always keeping the proportion of 21/20 subjects for training/testing.

5.1.4. RGBD-ID

The RGBD-ID dataset has been originally crafted to explore depth data in a re-id scenario. It contains four different groups of data, all from the same 79 people (identities): 14 female and 65 male. The first “Collaborative” group has been obtained by recording, in an indoor scenario, with a Kinect camera (RGB + depth data), a frontal view of the people, 2 meters away from the camera, walking slowly, avoiding occlusions and with stretched arms. The second group (“Backwards”) consists of back-view acquisitions of the people while walking away from the camera. The third (“Walking1”) and fourth (“Walking2”) groups of data are composed by frontal recordings of the people walking normally while entering a room in front of the camera. There are in average 5 frames of RGBD data per person per group. It is important to note that people in general changed their clothes between the acquisitions related to the four groups of data; most cloth changes occur between groups “Walking1” and “Walking2” (59 cases out of 79). Additionally, in the “Walking2” group 45 out of the 79 people have the same t-shirt, in order to simulate a work environment where people wear the same attire.

In the experiments, we use the “Collaborative” and the “Backwards” groups for fine-tuning, keeping “Walking2” as probe set and “Walking1” as gallery set.

Note that we introduce here a new way to use the RGBD-ID data. Previous studies mostly focus only on the depth data to obtain reasonable results, thereby ignoring the RGB imagery. This is because the change of outfit between acquisitions makes the re-id problem harder.

Preliminary studies carried out in [10] reported low performance when using RGB data only. In contrast, we only use the RGB data here to see whether SOMAnet can extract structural aspects of the human silhouette from them.

5.2. Analysis of SOMAnet

5.2.1. Comparing SOMAnet to other deep architectures

The first experiments show the advantage of SOMAnet w.r.t. other deep network architectures, such as siamese-inspired architectures. To this aim, we use the CUHK03 and Market-1501 datasets. We train and test on the respective partitions of each dataset, obtaining $\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$ and $\underbrace{\text{SOMAnet}}_{\text{Market-1501}}$.

Comparative results for the CUHK03 dataset against other recent deep network architectures are reported in Table 2.

	Rank				
	1	**2**	**5**	**10**	**20**
FPNN [3]	20.65	32.53	50.94	67.01	83.00
JointRe-id [4]	54.74	70.04	86.50	93.88	98.10
LSTM-re-id [7]	57.30	-	80.10	88.20	-
Personnet [8]	64.80	73.55	89.40	94.92	98.20
MB-DML [5]	65.04	-	-	-	-
Gated [6]	68.10	-	88.10	94.60	-
DGD-CNN [12]	72.60	-	-	-	-
MTDnet [33]	**74.68**	-	**95.99**	**97.47**	-
$\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$	68.90	82.10	91.00	95.60	98.30

Table 2: Analysis of the performance of SOMAnet against other deep network architectures when trained and tested exclusively on CUHK03, in the *single-shot* modality

The top three performers when trained exclusively on CUHK03 are DGD-CNN, $\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$ and MTDnet; note that two out of these three networks are Inception-based while the third is a multitask network trained with triplet loss.

With respect to the multi-shot modality, the only approach we can compare to is MB-DML [5, 59], a siamese architecture based on bilinear convolutional neural networks and deep metric learning. Results are shown in Table 3, with $\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$ achieving the best score.

	Rank				
	1	**2**	**5**	**10**	**20**
MB-DML [5]	80.60	-	-	-	-
$\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$	**83.60**	**93.40**	**97.50**	**99.20**	**99.70**

Table 3: Analysis of the performance of SOMAnet against another deep network architecture when trained and tested exclusively on CUHK03, in the *multi-shot* modality. (Note that only rank-1 performance in the multi-shot modality is provided by the MB-DML paper [5]).

Single- and multi-shot results on the Market-1501 dataset are shown on Table 4. For the singe-shot modality, SOMAnet surpasses the other methods in terms of CMC ranks and mean average precision.

Single-shot	Rank						**mAP**
	1	**5**	**10**	**20**	**30**	**50**	
Personnet [8]	37.21	-	-	-	-	-	18.57
SSDAL [32]	39.40	-	-	-	-	-	19.60
MB-DML [5]	45.58	-	-	-	-	-	26.11
Gated [6]	65.88	-	-	-	-	-	39.55
$\underbrace{\text{SOMAnet}}_{\text{Market-1501}}$	**70.28**	**87.53**	**91.69**	**94.57**	**95.64**	**96.85**	**45.05**
Multi-shot	Rank						**mAP**
	1	**5**	**10**	**20**	**30**	**50**	
SSDAL [32]	49.00	-	-	-	-	-	25.80
MB-DML [5]	56.59	-	-	-	-	-	32.26
LSTM-re-id [7]	61.60	-	-	-	-	-	35.3
Gated [6]	76.04	-	-	-	-	-	48.45
$\underbrace{\text{SOMAnet}}_{\text{Market-1501}}$	**77.49**	**91.81**	**94.69**	**96.56**	**97.27**	**98.25**	**53.50**

Table 4: Analysis of the performance of SOMAnet against other deep network architectures when trained and tested exclusively on Market-1501, in both *single-shot* and *multi-shot* modalities.

These first experiments show that SOMAnet, independently from the training dataset, leads to state-of-the-art re-id results. As we will see in Sec. 5.2.3, the adoption of SOMAset as training data gives even higher scores.

5.2.2. Probing specialized neurons in SOMAnet

Several insights have come from attempts at visualising what specific neurons respond to, for a given convolutional neural network [60, 61, 62]. Some approaches pose the problem of understanding neuron behavior as an optimization problem, where images are propagated through the network to find which image region maximizes the activation of a particular neuron [63]. Other visualization techniques have been used to identify neurons that respond to specific visual stimuli; for example, [64] individuates a neuron responsive to face patterns in a CNN trained for the Large Scale Visual Recognition challenge [52]. Although visualization methods can be used for finding specialized neurons, it can be a slow task since the analysis of results is still manual.

Here we propose a different approach: the goal is to find a specialized neuron N_S over a given set of neurons which gives the highest response to a given stimuli or characteristic (e.g. gender, obesity, a particular type of clothing) and is unlikely to respond to other characteristics. The search can be formulated as solving the optimization problem:

$$N_S = \underset{N}{\operatorname{argmax}}\, \mathcal{D}(\mathbf{C}, \mathbf{R}, N), \tag{1}$$

where $\mathcal{D}$ is a *discernibility* measurement between two sets of images $\mathbf{C}$ and $\mathbf{R}$, given a neuron N. The first set, $\mathbf{C}$, consists of the images that carry the *characteristics* that the specialized neuron N_S should respond to. The second set, $\mathbf{R}$, consists of all the remaining images that do not. The discernibility score is composed of two score functions. The first one, called *fire rate score*, indicates the tendency of a neuron to fire only for the set $\mathbf{C}$. The second score, called *activation score*, highlights how strong this tendency appears to be. By averaging the two scores, $\mathcal{D}$

indicates both the tendency and the strength of a neuron response for the set **C**. The score has to be applied on each neuron under analysis; other than finding the most involved neuron as in Eq. 1, the score can be used to sort the neurons with respect to their sensitivity to **C**.

We first define the fire rate score $\mathcal{F}$ as the difference in mean neuron activity over the sets **C** and **R**, as in Eq. 2, where #**C** and #**R** represent the cardinalities of the aforementioned sets:

$$\begin{aligned}\mathcal{F}(\mathbf{C},\mathbf{R},N) = & \left(\frac{1}{\#\mathbf{C}}\sum_{\mathbf{X}\in\mathbf{C}} T(A_N(\mathbf{X}))\right) \\ & - \left(\frac{1}{\#\mathbf{R}}\sum_{\mathbf{X}\in\mathbf{R}} T(A_N(\mathbf{X}))\right),\end{aligned} \tag{2}$$

where $\mathbf{X}$ is the input image, A_N is the activation of neuron N in a given layer and T is a threshold function here selected to be the Heaviside step function:

$$T(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

In our case, N is a neuron in the fully connected layer preceding the softmax layer. T was set to the Heaviside step function because A_N of the probed layer is zero-centered.

The activation score $\mathcal{A}$ is the normalized difference between the mean activations of subsets C and R as defined in Eq. 4. The normalizing factors of $\frac{1}{2}$ are due to the use of hyperbolic tangent as activation unit. These will ensure a maximum activation score $\mathcal{A}$ of 1:

$$\begin{aligned}\mathcal{A}(\mathbf{C},\mathbf{R},N) = & \frac{1}{2}\left(\frac{1}{\#\mathbf{C}}\sum_{\mathbf{X}\in\mathbf{C}} A_N(\mathbf{X})\right) \\ & - \frac{1}{2}\left(\frac{1}{\#\mathbf{R}}\sum_{\mathbf{X}\in\mathbf{R}} A_N(\mathbf{X})\right).\end{aligned} \tag{4}$$

Finally, discernibility $\mathcal{D}$ is defined as the mean of the fire rate and activation scores:

$$\mathcal{D}(\mathbf{C},\mathbf{R},N) = \frac{1}{2}\cdot(\mathcal{F}(\mathbf{C},\mathbf{R},N) + \mathcal{A}(\mathbf{C},\mathbf{R},N)). \tag{5}$$

To investigate the role that the neurons of SOMAnet play in encoding the human figure, we use the discernability measure defined in Sec. 5.2.2 on two datasets: the synthetic SOMAset and the real RGBD-ID. Given a dataset, we partition it into two groups, the *localization* **L** and the *exploration* **E**. The images in **L** are used to localize the specialized neurons w.r.t a structural characteristic (as being obese); in particular, the set **L** is manually subdivided in **C** (with images of subjects with that characteristic) and **R** (absence of that characteristic), in order to compute the discernability measure (see Eq. 5) . Subsequently, the images in **E** triggering the specialized neurons the most are analyzed, looking for analogies with the images in **C** . In general, we focus on visual characteristics that are present in a sufficient number of samples of a dataset: for SOMAset, we analyze obesity and gender, while for the RGBD-ID we analyse ectomorphism (being long and lean) and a particular kind of clothing, independently on color information.

In the case of SOMAset, **L** contains 64,000 images from 32 randomly selected subjects, 16 female and 16 male, while **E** contains 36000 images from the remaining 18 subjects - 9 female and 9 male. For the obesity trait, **C** contains 4,000 images from two obese subjects and **R** the remaining 60,000 ones from the other 30 subjects. Using Eq. 2 and 4 we compute the fire rate $\mathcal{F}$ and the activation $\mathcal{A}$, averaging them to get the discernability score $\mathcal{D}$ (Eq. 5). This process is carried out for each neuron, producing at the end a ranking of the most responsive neurons. Heuristically, we select the top 10 of them, as giving good results when it comes to the analysis of **E**; their values for $\mathcal{F}$, $\mathcal{A}$ and $\mathcal{D}$ are shown in Fig. 6a. As visible, the ranking shows the neurons reacting in a similar way, and this could mean that they are cooperating together to explain the data, in line with the distributed representation theory [65, 66, 67]. An automatic selection of the number of neurons required to represent a visual characteristic is still an open topic planned for future research.

Subsequently, on the set **E**, we extract those images that cause the network to have as most discerning neurons the same 10 found on the set **L**. In the majority of the cases, obese subjects pop out. A random sampling of the images is shown in Fig. 5.

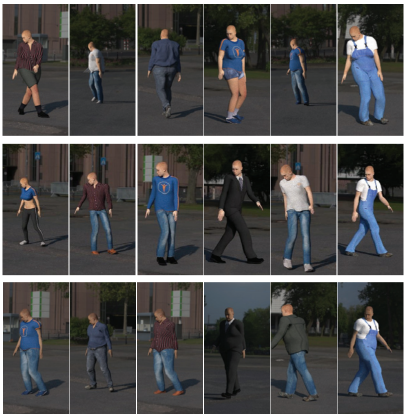
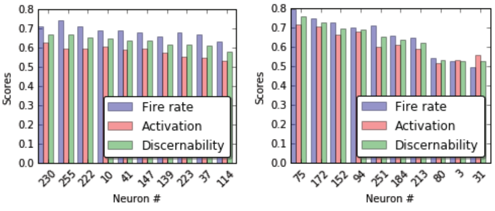


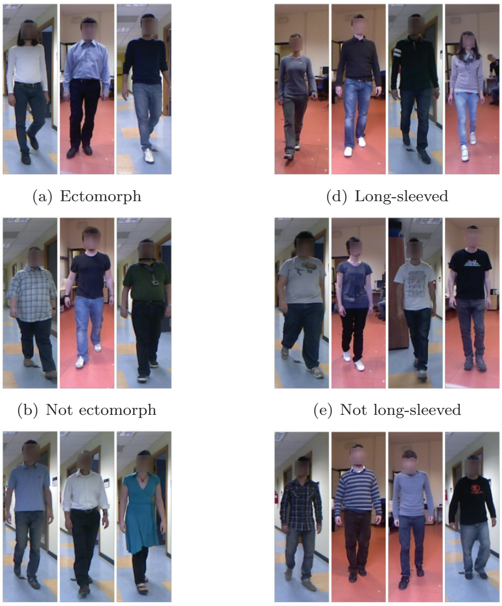
Figure 5: The first two rows show images used to find specialized neurons: the first row has images of obese subjects (∈ **C**), the second shows subjects without such characteristic (∈ **R**). The third row shows the test images ∈ **E** which responded to the specialized neurons. As visible, all of them portray obese subjects.

In the case of gender, **C** contains all the images from **L** where the subject is female (32000 elements), and **R** contains the male subjects (the other 32000). Even in this case, the top 10 neurons in terms of discernability score are kept (see Fig. 6b). Furthermore, $\mathcal{D}$ shows to decrease more rapidly than for the obesity case; this could mean that the gender trait is more easily detectable, requiring less neurons to focus on it. In fact, on the exploration set **E**, all the female subjects have been detected correctly.



(a) Obesity (b) Gender

Figure 6: Top 10 neurons ranked by their discernability score $\mathcal{D}$.



(a) Ectomorph

(d) Long-sleeved

(b) Not ectomorph

(e) Not long-sleeved

(c) Images triggering *ectomorph* neurons

(f) Images triggering *long sleeve* neurons

Figure 7: The first two rows show images used to find specialized neurons: the first row has images of subjects with the visual characteristic ($\in$ **C**), the second has images that do not have it ($\in$ **R**). The third row shows random test images $\in$ **E** which responded to the specialized neurons.

Concerning the analysis on the real RGBD-ID dataset, we use SOMAnet, trained on SOMAset and fine-tuned on the RGBD-ID, called here $\underbrace{\text{SOMAnet}}_{\text{SOMAset+RGBD-ID}}$. We first find neurons that respond to *ectomorph* (long and lean) subjects and subjects with long-sleeved shirts, respectively. In particular, the localization set **L** is composed of 59 subjects, and the exploration set **E** of the remaining 20; both partitions contain subjects that possess or not the characteristic. In the same way as for the previous experiments, we find the 10 top neurons that respond to the presence of the characteristic (high positive $\mathcal{D}$), and subsequently we check those images of the exploration set which trigger the same 10 neurons. Results are shown in Fig. 7.

These experiments show that SOMAnet sees beyond the appearance of the human silhouette, capturing structural aspects, which are capable of boosting the re-id performance.

5.2.3. SOMAnet + SOMAset

We next analyze the re-id performance when SOMAnet is trained from scratch with SOMAset, and fine-tuned on the training partition of another dataset, whose testing partition is used to calculate the re-id figures. In this case, we analyze the performance on all the four datasets, comparing against the approaches that, at the time of writing, exhibit the best performance.

For CUHK03, results of the single- and multi-shot modalities are reported in terms of CMC curves and mAP in Table 5.

Here SOMAnet has been trained from scratch on SOMAset and fine-tuned on CUHK03, labelled $\underbrace{\text{SOMAnet}}_{\text{SOMAset+CUHK03}}$. The resulting classifier is competitive against the state-of-the-art. Concerning the multi-shot modality, the only approach that operates on the CUHK03 dataset is MB-DML, which provides results just for rank 1 of the CMC curve.

We also report the scores obtained with SOMAnet, trained from scratch on CUHK03 (these are the results reported in Sec. 5.2.1), to show the advantage of bringing in SOMAset into play.

Results on Market-1501 are reported in Table 6 along with the competitive approaches. We also report here the scores obtained with SOMAnet, trained from scratch on Market-1501. For the single-shot evaluation $\underbrace{\text{SOMAnet}}_{\text{SOMAset+Market-1501}}$ provides a mAP of 47.89%.

The third dataset under analysis is RAiD, useful for evaluating the behavior of the SOMA approach when few data are available to fine-tune the network. The dataset has so far just been employed for the multi-shot modality. The CMC scores on RAiD are reported in Table 7

SOMA saturates CMC performance very soon for all camera combinations, in several cases starting as early as rank 2. This supports the fact that fine-tuning SOMAnet on a small dataset worked appropriately.

Single-shot	Rank				
	1	2	5	10	20
KISSME [27]	14.17	22.30	37.46	52.20	69.38
FPNN [3]	20.65	32.53	50.94	67.01	83.00
LOMO+XQDA [28]	52.20	66.74	82.23	92.14	96.25
JointRe-id [4]	54.74	70.04	86.50	93.88	98.10
LSTM-re-id [7]	57.30	-	80.10	88.20	-
LOMO+MLAPG [29]	57.96	-	-	-	-
Ensemble [34]	62.10	76.60	89.10	94.30	97.80
Null space [31]	62.55	-	90.05	94.80	98.10
Personnet [8]	64.80	73.55	89.40	94.92	98.20
MB-DML [5]	65.04	-	-	-	-
Gated [6]	68.10	-	88.10	94.60	-
DGD-CNN [12]	72.60	-	-	-	-
MTDnet [33]	**74.68**	-	**95.99**	**97.47**	-
$\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$	68.90	82.10	91.00	95.60	98.30
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+CUHK03}}$	72.40	81.90	92.10	95.80	98.50

Multi-shot	Rank				
	1	2	5	10	20
MB-DML [5]	80.60	-	-	-	-
$\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$	83.60	93.40	97.50	99.20	**99.70**
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+CUHK03}}$	**85.90**	**94.00**	**98.10**	**99.30**	99.60

Table 5: Analysis of the performance of SOMAnet against other methodologies when trained on SOMAset, fine-tuned on the training partition of the CUHK03 dataset, and tested on the test partition of the CUHK03 dataset, in both *single-shot* and *multi-shot* modalities.

Single-shot	Rank						mAP
	1	5	10	20	30	50	
BoW [17]	35.84	52.40	60.33	67.64	71.88	75.80	14.75
BoW,LMNN [17]	34.00	-	-	-	-	-	15.66
BoW,ITML [17]	38.21	-	-	-	-	-	17.05
Personnet [8]	37.21	-	-	-	-	-	18.57
SSDAL [32]	39.40	-	-	-	-	-	19.60
BoW,KISSME [17]	44.42	63.90	72.18	78.95	82.51	87.05	20.76
SCSP [30]	51.90	-	-	-	-	-	26.35
Null space [31]	61.02	-	-	-	-	-	35.68
Gated [6]	65.88	-	-	-	-	-	39.55
$\underbrace{\text{SOMAnet}}_{\text{Market-1501}}$	70.28	87.53	91.69	94.57	95.64	96.85	45.05
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+Market-1501}}$	**73.87**	**88.03**	**92.22**	**95.07**	**96.20**	**97.39**	**47.89**

Multi-shot	Rank						mAP
	1	5	10	20	30	50	
BoW [17]	44.36	60.24	66.48	73.25	76.19	76.69	19.42
SSDAL [32]	49.00	-	-	-	-	-	25.80
LSTM-re-id [7]	61.60	-	-	-	-	-	35.30
Null space [31]	71.56	-	-	-	-	-	46.03
Gated [6]	76.04	-	-	-	-	-	48.45
$\underbrace{\text{SOMAnet}}_{\text{Market-1501}}$	77.49	91.81	94.69	96.56	97.27	98.25	53.50
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+Market-1501}}$	**81.29**	**92.61**	**95.31**	**97.12**	**97.68**	**98.43**	**56.98**

Table 6: Analysis of the performance of SOMAnet against other methodologies when trained on SOMAset, fine-tuned on the training partition of the Market-1501 dataset, and tested on the test partition of the same dataset, in the *single-shot* and *multi-shot* modalities.

	Rank				
cam1-cam3	1	2	5	10	20
Double-view [24]	46.67	90.00	96.67	98.33	100.00
NCR on ICT [18]	60.00	82.00	95.00	100.00	100.00
Multi-view [24]	61.67	91.67	96.67	100.00	100.00
NCR on FT [18]	67.00	83.00	93.00	98.00	100.00
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+RAID}}$	**69.00**	**99.00**	**100.00**	100.00	100.00

	Rank				
cam1-cam2	1	2	5	10	20
Multi-view [24]	78.33	98.33	100.00	100.00	100.00
NCR on FT [18]	86.00	97.00	100.00	100.00	100.00
Double-view [24]	88.33	100.00	100.00	100.00	100.00
NCR on ICT [18]	89.00	98.00	100.00	100.00	100.00
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+RAID}}$	**95.00**	100.00	100.00	100.00	100.00

	Rank				
cam1-cam4	1	2	5	10	20
NCR on ICT [18]	66.00	84.00	94.00	100.00	100.00
Multi-view [24]	66.67	98.33	100.00	100.00	100.00
NCR on FT [18]	68.00	86.00	99.00	99.00	100.00
Double-view [24]	76.67	100.00	100.00	100.00	100.00
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+RAID}}$	**90.00**	100.00	100.00	100.00	100.00

Table 7: Analysis of the performance of SOMAnet against other architectures when trained from scratch with the training RAID dataset, and tested on the test partition of the same dataset, in the *multi-shot* modality.

	Rank					
	1	5	10	20	30	50
RGBD-ID [10]	12.66	43.04	53.16	84.81	96.20	100.00
PDM [68]	17.72	36.71	40.51	59.49	77.22	91.14
$\underbrace{\text{SOMAnet}}_{\text{SOMAset+RGBD-ID}}$	**63.29**	**82.28**	**88.61**	**94.94**	**96.23**	**98.73**
Average Human Performance	65.00	95	-	-	-	-

Table 8: Analysis of the performance of SOMAnet against other architectures when trained on SOMAset, fine-tuned with the training RGBD-ID dataset, and tested on the test partition of RGBD-ID, in the *multi-shot* modality. Average human performance in Rank 1 and Rank 5 also reported for reference.

Finally, the last dataset we take into account is RGBD-ID. Results on the "Walking1" vs "Walking2" setting are reported in Table 8. In order to compare against human performance, we also conducted an experiment with 20 human annotators who were asked to select the top 5 subjects to a given query image (thus producing results for Rank 1 and Rank 5); we report the average performance of the 20 annotators in the same Table. The annotators complained that the task was tedious, time consuming and challenging, especially when blurred faces were involved. They also reported that their selection was based not only on body shapes but also on detecting common accessories between query and gallery.

Notably, the competing approaches work either with the silhouette [68] or with depth images associated to the RGB images (which are not used in the case of SOMA) [10]. As mentioned in Sec. 5.1.4, the reason is that people change their clothing between different camera acquisitions. In our case, we do the opposite, discarding depth information while retaining the RGB images only; still the results are well above the state-of-the-art. $\underbrace{\text{SOMAnet}}_{\text{SOMAset+RGBD-ID}}$ has almost the same rank 1 performance as the average human but SOMAnet is much faster.

Fig. 8 illustrates some probe images in the left column and the corresponding ranked gallery images provided by our approach in the rest of the columns, where the correct match is framed in green. As visible, in most of the cases, the correct individual is in the early ranks *even if he/she does not wear the same clothing*. Interestingly, in the second row of Fig. 8, the probe image of the woman produced two images of women in the top 2 ranks, with another woman in seventh position. In the case of the other probe male subjects, no female subjects appeared in the top ranked gallery images. In the first row, the probe subject is wearing a jacket which is abundant on the belly. Consequently, many of the top ranked gallery images are of endomorph subjects. In contrast, the probe subjects on the third and fourth rows are ectomorph and so are the majority of the retrieved images in the top ranks.

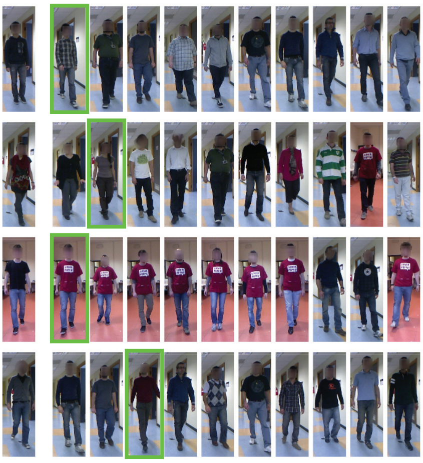


Figure 8: Ranking results of RGBD-ID; probe images are shown in left column. The top 10 ranked gallery images are shown on the right. The ground-truth match is highlighted with a green frame.

5.3. *Analysis of SOMAset*

This section explores different characteristics of SOMAset, clarifying their role in the re-identification task. First, we evaluate the importance of training SOMAnet on SOMAset from scratch, *independently of the dataset used to perform fine-tuning and testing*. To this end, we evaluate the training from scratch with diverse datasets (in addition to SOMAset), choosing different datasets for the fine-tuning and testing (for example, we train SOMAnet from scratch on CUHK03, fine-tuning and testing on Market-1501). Second, we evaluate the effect of reducing the number of different subjects and the number of poses.

5.3.1. *Training from scratch on different datasets*

The datasets that are suitable for training deep networks from scratch are CUHK03 and Market-1501, due to their size (see Sec. 5.2). In particular, we calculate re-id scores (rank 1 of CMC curve and mAP for brevity) where SOMAnet is trained with CUHK03, fine-tuned and tested with Market-1501, labelled $\underbrace{\text{SOMAnet}}_{\text{CUHK03+Market-1501}}$; then we consider the network trained on SOMAset, fine-tuned and tested on Market-1501, labelled $\underbrace{\text{SOMAnet}}_{\text{SOMAset+Market-1501}}$. To show the effect of this cross-dataset learning, we also present the results where SOMAnet is trained with Market-1501 from scratch and tested on it, labelled $\underbrace{\text{SOMAnet}}_{\text{Market-1501}}$ (these latter are the results already presented in the experiments of Sec. 5.2).

We next invert the roles of CUHK03 and Market-1501, that is, CUHK03 is employed as evaluation dataset, giving rise to $\underbrace{\text{SOMAnet}}_{\text{Market-1501+CUHK03}}$, $\underbrace{\text{SOMAnet}}_{\text{SOMAset+CUHK03}}$ and $\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$. All of these setups are evaluated in both the single and multi-shot modalities. Results are reported in Table 9 and Table 10, respectively.

	Mean Average Precision		
	$\underbrace{\text{SOMAnet}}_{\text{Market1501}}$	$\underbrace{\text{SOMAnet}}_{\text{CUHK03+Market1501}}$	$\underbrace{\text{SOMAnet}}_{\text{SOMAset+Market1501}}$
Single-shot	45.05	45.97	**47.89**
Multi-shot	53.50	54.20	**56.98**
	Rank 1		
	$\underbrace{\text{SOMAnet}}_{\text{Market1501}}$	$\underbrace{\text{SOMAnet}}_{\text{CUHK03+Market1501}}$	$\underbrace{\text{SOMAnet}}_{\text{SOMAset+Market1501}}$
Single-shot	70.28	73.22	**73.87**
Multi-shot	77.49	79.81	**81.29**

Table 9: Analysis of the role of SOMAset as learning data for the training from scratch step of SOMAnet. For the same testing dataset, Market-1501, different repositories are used for the training from scratch, namely, Market-1501 itself, CUHK03 and SOMAset, respectively.

	Mean Average Precision		
	$\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$	$\underbrace{\text{SOMAnet}}_{\text{Market1501+CUHK03}}$	$\underbrace{\text{SOMAnet}}_{\text{SOMAset+CUHK03}}$
Single-shot	73.92	73.91	**76.65**
Multi-shot	86.79	87.49	**88.60**
	Rank 1		
	$\underbrace{\text{SOMAnet}}_{\text{CUHK03}}$	$\underbrace{\text{SOMAnet}}_{\text{Market1501+CUHK03}}$	$\underbrace{\text{SOMAnet}}_{\text{SOMAset+CUHK03}}$
Single-shot	68.90	68.90	**72.40**
Multi-shot	83.60	84.40	**85.90**

Table 10: Analysis of the role of SOMAset as training data for the training from scratch step of SOMAnet. For the same testing dataset, CUHK03, different datasets are used for the training from scratch, namely, the CUHK03 itself, Market-1501 and SOMAset, respectively.

By observing the two tables, some useful facts emerge.

First, cross-dataset learning seems to be beneficial in general, except for the single-shot modality when testing on the CUHK03 dataset, where the performance essentially does not change. Notably, fine-tuning gives better results when it is carried out with Market-1501 on the network trained from scratch on CUHK03, than vice-versa. This is possibly due to the larger size of Market-1501 w.r.t CUHK03. When SOMAset is used for the training from scratch, the improvement is systematically very significant.

This is an interesting result, since it indicates that, other than being an economic and effective proxy for real data, the SOMA framework appears to produce a nice general optimization of the network, that later can be properly specialized using the data where the classifier will be applied.

5.3.2. Changing the number of subjects

In these experiments, we analyze the effect of reducing the number of subjects of SOMAset. We recall here that each subject (that is, a mixture of somatotypes) gives rise to 2000 images (250 human poses × 8 sets of clothes). The original SOMAset has 50 subjects, and we evaluate the effect of having 32, 16 and 8. These numbers have been obtained by *randomly* removing people from the dataset, repeating the experiments twice. When we go to fewer than 8 subjects (in particular, we tried 4) the training of SOMAnet produces several dead/deactivated neurons.

The evaluation of the reduced SOMAsets is carried out with fine-tuning and testing on the Market-1501 dataset. The results are given in terms of CMC ranks and mAP in Table 11.

Single-shot			
#Images in Dataset	**#Subjects in SOMAset**	**Rank 1**	**mAP**
100000	50	73.87	47.89
64000	32	73.13	46.70
32000	16	72.12	46.23
16000	8	71.70	45.77
Multi-shot			
#Images in SOMAset	**#Subjects in SOMAset**	**Rank 1**	**mAP**
100000	50	81.29	56.98
64000	32	80.70	55.46
32000	16	80.14	55.35
16000	8	78.79	54.68

Table 11: Analysis of the role of the *size* of SOMAset as training data. Here SOMAset was rendered in original and reduced versions by *changing the number of rendered subjects*. The different versions of SOMAset were fine-tuned with the training partition of the Market-1501 dataset, and tested on the test partition of the same dataset.

As one can expect, adding subjects leads to increased performance. The curious aspect is that the increase is very mild, both in terms of rank 1 and mAP. A roughly linear relation between number of subjects and the performance seems to hold.

We should highlight two points: Market-1501 has 750 subjects in the testing set, and having just 1% of performance increase does impact substantially the re-identification capabilities (an increase of 7.5 subjects matched correctly in the first rank); secondly, in the deep network literature it is widely known that the role of fine-tuning is absolutely crucial, much more than the role of the training from scratch.

5.3.3. Changing the number of poses

In the final experiment, we investigate the impact of reducing SOMAset by *randomly* removing poses from the rendering protocol. To compare with Sec. 5.3.2, and understand if it is more important to have more poses or more subjects into play, we select a number of poses that result in the same number of images as in the previous study.

Specifically, we create reduced datasets with 250, 160, 80 and 40 poses, giving rise to 100K, 64K,32K and 16K images, corresponding to what we obtained with 50, 32, 16 and 8 subjects, respectively.

Single-shot			
#Images in Dataset	**#Poses in SOMAset**	**Rank 1**	**mAP**
100000	250	73.87	47.89
64000	160	72.39	46.08
32000	80	71.44	45.18
16000	40	70.19	44.58
Multi-shot			
#Images in SOMAset	**#Poses in SOMAset**	**Rank 1**	**mAP**
100000	250	81.29	56.98
64000	160	79.16	54.90
32000	80	78.65	53.72
16000	40	78.65	53.56

Table 12: Analysis of the role of the *size* of SOMAset as training data. Here SOMAset was rendered in original and reduced versions by *changing the number of poses*. The different versions of SOMAset were fine-tuned with the training partition of Market-1501, and tested on the test partition of the same dataset.

The comparison of Tables 11 and 12 indicates that having more subjects than poses is more auspicable, and this is meaningful, since the intraclass variance of a dataset is intuitively higher when having different subjects instead of different poses, in terms of visual variability (consider the rows starting from the second one, since the first row shows the performance of the full SOMAset, which is the same in both tables).

5.4. Effects of Illumination, poses and camera viewpoints

It is interesting to attempt to quantitatively assess the individual effect of illumination, poses and camera viewpoints on performance. Then one could determine which variable to prioritize while modeling and rendering a synthetic dataset for re-id. To this end, we have isolated 4 variants of SOMAset with 16000 images: the first consists of a manual selection of 16000 images where the subject appears dark (bad illumination), the second consists of 16000 images in which the number of rendered poses has been reduced following the procedure of Section 5.3.3 (restricted poses), the third consists of a manual selection of 16000 images where the subject is seen from the back (bad viewpoint), while the fourth is a balanced random selection of 16000 images called the control group, for comparison. We have repeated the experiment with 4 similar variants of 32000 images in order to see how the dataset size change influences these factors. The results can be seen in Table 13.

Multi-shot				
SOMAset variant	Rank 1	mAP	Rank 1	mAP
Balanced Control Group	80.14	55.35	78.89	54.68
Bad Illumination	79.19	54.77	71.73	54.33
Bad Viewpoint	79.13	53.84	71.56	54.56
Restricted Poses	78.65	53.72	70.19	53.56
	32000 Images		16000 Images	

Table 13: Comparitve analysis of rendering factors of SOMAset on SOMAnet performance. The effect of a balanced control group is compared against similarly sized datasets with bad illumination, restricted number of poses and bad camera viewpoints. The experiment was performed for 16000 and 32000 images giving a total of 8 variants of SOMAset. All variants were fine-tuned with the training partition of Market-1501 and tested on the test partition of the same dataset.

As one would expect, the Balanced Control Group performs best across both dataset sizes. Looking at Rank 1 performance, in both dataset sizes, the most degrading factor compared to Balanced Control Group performance is restricting the number of poses, followed by bad viewpoint and bad illumination. The mAP performance generally follows the same pattern, except for the case of bad viewpoint where, paradoxically, mAP performance drops when going from 16000 to 32000 images. The Rank 1 difference between the Balanced Control Group and the ’degraded’ variants at 16000 images is over 7% while the equivalent figure for 32000 images is less than 2%; this is likely to be due to the fact that overfitting to a ’degraded’ dataset is easier the smaller its size is.

6. Conclusions

Synthetic training data can greatly help to initialise deep networks. Tasks such as re-identification should not be faced exclusively by siamese architectures; instead, single-path networks can be employed as successful feature extractors. A by-product is that these networks can be easily probed, investigating the semantics being captured by the neurons.

In this work, we find that such networks can see beyond apparel, capturing structural aspects of the human body, such as their somatotype. This can be fully exploited with an appropriate dataset; in this respect, we introduce, for the first time in the re-identification field, the strategy of using synthetic data as proxy for real data. In particular, having synthetic datasets for training a network from scratch seems to be a very effective manoeuvre, producing successive fine-tuned architectures with a very high recognition rate. The proposed inception-based network, SOMAnet, trained on the synthetic dataset SOMAset[2] can match people even if they change apparel between camera acquisitions.

Various future directions are intriguing and promising. First, the nature of the synthetic dataset needs to be explored under different respects: an obvious question is, what is the behavior of the network when the number of subjects contained in the dataset tends to infinity. Specifically, we show a somewhat linear increase in performance with respect to the addition of diverse subjects. Certainly, at a given point, a plateau should be reached, and finding this point is a key open issue.

Another question regards the importance of the background in the images: to bound the degree of freedom of our analysis, we decided to place our synthetic pedestrians in a single scene that, even if arbitrarily large, does not offer the variability contained in other datasets. Our intuition is that having a fixed background forces the network to focus on the foreground objects. At the same time, a single scene may help the network in understanding differences among individuals, acting as a frame of reference to capture, for example, different sizes among individuals. In a preliminary experiment, not reported here intentionally, we omit the background leaving a grey homogeneous flat area behind the subjects. Results in recognition are definitely worse, but we did not investigate this point further. The importance of having realistic images is another question that we would like to explore. As already mentioned, the usual re-identification setup produces individuals at a certain low resolution, so that fine details such as the face cannot be processed. It could be nice to have an advanced re-id setting, where high-resolution cameras are employed, collecting high frequency cues. In that case it would be reasonable to expect a difference in recognition rates, depending on the realism of the training data.

Finally, a wider, conceptual question pops out: with such a framework, capable of understanding bodily cues of human beings, going beyond the mere appearance of the outfit, is it still reasonable to talk about re-identification, or does it make more sense to call for non-collaborative person recognition at a distance? In that case, a brand new biometric field is opening up.

[1] S. Gong, M. Cristani, S. Yan, C. C. Loy, Person re-identification, Vol. 1, Springer, 2014. 1

[2] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, V. Murino, Custom pictorial structures for re-identification., in: Proc. BMVC, 2011. 1

[3] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: Deep filter pairing neural network for person re-identification, in: Proc. CVPR, 2014. 1, 2, 6, 7, 8, 11

[4] E. Ahmed, M. Jones, T. K. Marks, An improved deep learning architecture for person re-identification, in: Proc. CVPR, 2015. 1, 2, 6, 8, 11

[5] E. Ustinova, Y. Ganin, V. S. Lempitsky, Multiregion bilinear convolutional neural networks for person re-identification, CoRR abs/1512.05300. 1, 2, 8, 11

[6] R. Rama Varior, M. Haloi, G. Wang, Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification, in: Proc. ECCV, 2016. 1, 2, 8, 11

[7] R. Rama Varior, B. Shuai, J. Lu, D. Xu, G. Wang, A Siamese Long Short-Term Memory Architecture for Human Re-Identification, in: Proc. ECCV, 2016. 1, 2, 8, 11

[8] L. Wu, C. Shen, A. van den Hengel, Personnet: Person re-

[2] SOMAset will be released with a open source license to enable further developments in re-identification.

identification with deep convolutional neural networks, CoRR abs/1601.07255. 1, 2, 8, 11

[9] M. Munaro, A. Fossati, A. Basso, E. Menegatti, L. Van Gool, One-shot person re-identification with a consumer depth camera, in: Person Re-Identification, Springer, 2014, pp. 161–181. 1, 2

[10] I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, V. Murino, Re-identification with rgb-d sensors, in: Proc. ECCV - Workshops and Demonstrations, 2012. 1, 2, 8, 11

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proc. CVPR, 2015. 1, 4, 5, 6

[12] T. Xiao, H. Li, W. Ouyang, X. Wang, Learning deep feature representations with domain guided dropout for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. 1, 8, 11

[13] W. H. Sheldon, S. S. Stevens, W. B. Tucker, The varieties of human physique., Harper, 1940. 2, 3

[14] X. Peng, B. Sun, K. Ali, K. Saenko, Learning deep object detectors from 3d models, in: Proc. ICCV, 2015. 2, 3

[15] X. Zhang, Y. Fu, A. Zang, L. Sigal, G. Agam, Learning classifiers from synthetic data using a multichannel autoencoder, arXiv preprint arXiv:1503.03163. 2, 3

[16] A. Borji, S. Izadi, L. Itti, ilab-20m: A large-scale controlled object dataset to investigate deep learning, in: Proc. CVPR, 2016. 2, 3

[17] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: Proc. ICCV, 2015. 2, 7, 11

[18] A. Das, A. Chakraborty, A. K. Roy-Chowdhury, Consistent re-identification in a camera network, in: Proc. ECCV, 2014. 2, 7, 11

[19] A. Bedagkar-Gala, S. K. Shah, A survey of approaches and trends in person re-identification, Image and Vision Computing 32 (4) (2014) 270–286. 2

[20] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: Proc. ECCV, 2008. 2

[21] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: Proc. CVPR, 2010. 2

[22] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: Proc. ECCV, Springer, 2014. 2

[23] N. Martinel, A. Das, C. Micheloni, A. K. Roy-Chowdhury, Re-identification in the function space of feature warps, Transactions PAMI 37 (8) (2015) 1656–1669. doi:10.1109/TPAMI.2014.2377748. 2

[24] Z. Zhang, Y. Chen, V. Saligrama, Group membership prediction, in: Proc. ICCV, 2015. 2, 11

[25] A. Chakraborty, A. Das, A. K. Roy-Chowdhury, Network consistent data association, Transactions PAMI 38 (9) (2016) 1859–1871. doi:10.1109/TPAMI.2015.2491922. 2

[26] S. Bak, F. Brémond, Re-identification by covariance descriptors, in: Person Re-Identification, Springer, 2014, pp. 71–91. 2

[27] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Proc. CVPR, 2012. 2, 11

[28] S. Liao, Y. Hu, X. Zhu, S. Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proc. CVPR, 2015. 2, 11

[29] S. Liao, S. Z. Li, Efficient psd constrained asymmetric metric learning for person re-identification, in: Proc. ICCV, 2015. 2, 11

[30] D. Chen, Z. Yuan, B. Chen, N. Zheng, Similarity learning with spatial constraints for person re-identification, in: Proc. CVPR, 2016. 2, 11

[31] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: Proc. CVPR, 2016. 2, 11

[32] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Deep attributes driven multi-camera person re-identification, in: Proc. ECCV, 2016. 2, 8, 11

[33] W. Chen, X. Chen, J. Zhang, K. Huang, A multi-task deep network for person re-identification., in: The Association for the Advancement of Artificial Intelligence (AAAI), 2017, pp. 3988–3994. 2, 8, 11

[34] S. Paisitkriangkrai, C. Shen, A. van den Hengel, Learning to rank in person re-identification with metric ensembles, in: Proc. CVPR, 2015. 3, 11

[35] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27, Curran Associates, Inc., 2014, pp. 1988–1996. 3

[36] M. Buhrmester, T. Kwang, S. D. Gosling, Amazon's mechanical turk a new source of inexpensive, yet high-quality, data?, Perspectives on psychological science 6 (1) (2011) 3–5. 3

[37] Y. Xia, X. Cao, F. Wen, J. Sun, Well begun is half done: Generating high-quality seeds for automatic image dataset construction from web, in: Proc. ECCV, 2014. 3

[38] X. Chen, A. Gupta, Webly supervised learning of convolutional networks, arXiv preprint arXiv:1505.01554. 3

[39] D. S. Cheng, F. Setti, N. Zeni, R. Ferrario, M. Cristani, Semantically-driven automatic creation of training sets for object recognition, Computer Vision and Image Understanding 131 (2015) 56–71. 3

[40] B. Sun, K. Saenko, From virtual to reality: Fast adaptation of virtual object detectors to real domains, in: Proc. BMVC, 2014. 3

[41] C. H. Lampert, H. Nickisch, S. Harmeling, Learning to detect unseen object classes by between-class attribute transfer, in: Proc. CVPR, 2009. 3

[42] X. Glorot, A. Bordes, Y. Bengio, Domain adaptation for large-scale sentiment classification: A deep learning approach, in: Proc. ICML, 2011. 3

[43] S. J. Pan, Q. Yang, A survey on transfer learning, Knowledge and Data Engineering, IEEE Transactions on 22 (10) (2010) 1345–1359. 3

[44] N. McLaughlin, J. M. Del Rincon, P. Miller, Data-augmentation for reducing dataset bias in person re-identification, in: Proc. AVSS, 2015. 3

[45] Cmu graphics lab motion capture database, http://mocap.cs.cmu.edu/, accessed: 2015-09-30. 4

[46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proc. CVPR, 2016. 4, 5

[47] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, J. Xiao, 3dshapenets: a deep representation for volumetric shapes, in: Proc. CVPR, 2015. 4

[48] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 1097–1105. 4, 6

[49] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556. 4, 6

[50] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift., in: Proc. ICML, 2015. 4, 5, 6

[51] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks., in: Aistats, 2011. 4

[52] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, IJCV 115 (3) (2015) 211–252. 5, 8

[53] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, in: Neurocomputing: Foundations of Research, 1988, Ch. 8, pp. 696–699. 6

[54] I. Sutskever, J. Martens, G. E. Dahl, G. E. Hinton, On the importance of initialization and momentum in deep learning, in: Proc. ICML, 2013. 6

[55] X. Glorot, Y. Bengio, Understanding the difficulty of training

deep feedforward neural networks., in: Aistats, Vol. 9, 2010, pp. 249–256. 6

[56] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the ACM International Conference on Multimedia, ACM, 2014, pp. 675–678. 6

[57] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, T. Darrell, Decaf: A deep convolutional activation feature for generic visual recognition, in: Proceedings of the 31th International Conference on Machine Learning, ICML, 2014. 6

[58] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks?, in: Advances in Neural Information Processing Systems, 2014, pp. 3320–3328. 6

[59] T.-Y. Lin, A. RoyChowdhury, S. Maji, Bilinear cnn models for fine-grained visual recognition, in: Proc. ICCV, 2015. 8

[60] D. Erhan, Y. Bengio, A. Courville, P. Vincent, Visualizing higher-layer features of a deep network, Dept. IRO, Université de Montréal, Tech. Rep 4323. 8

[61] K. Simonyan, A. Vedaldi, A. Zisserman, Deep inside convolutional networks: Visualising image classification models and saliency maps, arXiv preprint arXiv:1312.6034. 8

[62] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: Proc. ECCV, 2014. 8

[63] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proc. CVPR, 2014. 8

[64] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson, Understanding neural networks through deep visualization, in: Deep Learning Workshop, International Conference on Machine Learning (ICML), 2015. 8

[65] G. Hinton, J. L. McClelland, D. E. Rumelhart, Distributed representations, in parallel distributed processing: Explorations in the microstructure of cognition (1986). 9

[66] T. Plate, Distributed representations, Encyclopedia of Cognitive Science. 9

[67] A. B. Valdez, M. H. Papesh, D. M. Treiman, K. A. Smith, S. D. Goldinger, P. N. Steinmetz, Distributed representation of visual objects by single neurons in the human brain, The Journal of Neuroscience 35 (13) (2015) 5180–5186. 9

[68] O. Huynh, B. Stanciulescu, Person re-identification using the silhouette shape described by a point distribution model, in: Proc. WACV, 2015. 11

ISBN 978-82-326-4872-6 (printed ver.)
ISBN 978-82-326-4873-3 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology