EURECOM
Department of Multimedia
2229, route des Crêtes
B.P. 193
06904 Sophia-Antipolis
FRANCE

Research Report RR-12-262

Improving Identification by Pruning: a Case Study on Face Recognition and Body Soft Biometric

Last update January 4th, 2012

Carmelo Velardo, and Jean-Luc Dugelay

Tel : (+33) 4 93 00 81 00
Fax : (+33) 4 93 00 82 00
Email : {velardo,dugelay}@eurecom.fr

[1]EURECOM's research is partially supported by its industrial members: BMW Group, Cisco, Monaco Telecom, Orange, SAP, SFR, Sharp, STEricsson, Swisscom, Symantec, Thales.

Improving Identification by Pruning: a Case Study on Face Recognition and Body Soft Biometric

Carmelo Velardo, and Jean-Luc Dugelay

Abstract

We investigate body soft biometrics capabilities to perform pruning of a hard biometrics database improving both retrieval speed and accuracy. Our pre-classification scheme based on anthropometric measures is elaborated on a large scale medical dataset to guarantee statistical meaning of the results, and tested in conjunction with a face recognition algorithm. Our assumptions are verified building and testing our system on a chimera dataset. We clearly identify the trade off among pruning, accuracy, and mensuration error of an anthropomeasure based system. Our results show that even in the worst case of $\pm 10\%$ error magnitude in the anthropometric measures, our pruning scheme improves the accuracy performances guaranteeing a speedup of $2\times$ factor.

Index Terms

Soft biometrics, pruning, anthropometric measurements, face recognition

Contents

List of Figures

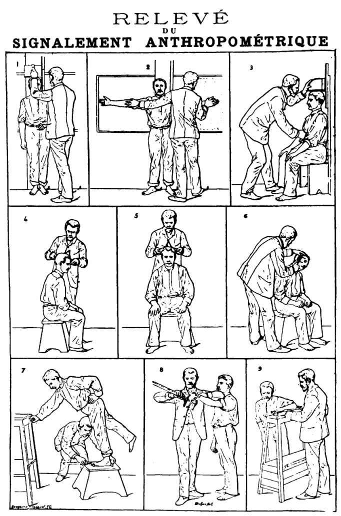


Figure 1: The image shows the methodology applied to gather information from the suspect in the *bertillonage* system. The procedure was standardized by Bertillon in his book .

1 Introduction

Soft biometrics are a new trend in biometric studies which exploit the information coming from non-reliable, non-discriminative human traits (height, eye color, ...). They were firstly investigated by Jain et al.to improve multimodal fusion [1,2] and later they were exploited to perform identification [3, 4], or to simply extract information related to the user [5].

Anthropometric measures fulfill the definition of soft biometrics [3]: they do not provide a specific pattern for identification (*e.g.*fingerprint and iris), they are human-compliant, and they are available without user cooperation. Anthropometry is the science that uses human body measurements to study human variation and differences. This research field is particularly useful in case of medicine, industrial design, fashion, and other areas. A particular area where anthropometric studies were flourishing at the beginning of the biometric history was the identification of people: Alphonse Bertillon first derived a mensuration methodology (see fig. 1) which involved several body measures to classify (and then identify) criminals. Nevertheless, anthropometry was early replaced by the use of fingerprint identification which was found more reliable. Indeed, hard biometrics [3] contrarily to soft biometrics, provide a higher distinctiveness that eases the identification task, although their elaboration is often quite expensive in terms of resources required.

In this paper we explore the use of body soft biometric, under the form of anthropometric measurements, to create a pre-classification scheme that is able to prune the search space of a subsequent hard biometrics (face recognition) module. We demonstrate that a hard biometric based face recognition system, properly preprocessed by our soft biometric anthropometry features, improves its performance in both accuracy and recognition speed. Using our anthropometric signature to prune the face database, we exploit the independent complementary information provided by the body soft biometrics that makes this accuracy and speed gain possible.

However, the anthropometric mensuration process is inaccurate due to several factors. Sources of error are mainly due to the sensing device (tape meter, 3D scanner, ...), to the human operator, and to the inner variability of human measures [6]. Even if the sensing procedure could be eased in a near future by a new generation of specialized devices like body scanners or 3D video sensors like Microsoft Kinect, errors will eventually affect the recognition performance of an anthropometric system. Given this observation, we study how an increasing error during the mensuration step can affect the retrieval process and we show that even in presence of strong error magnitude (10% of the real measure) anthropometric measures could still be used for pruning the search space improving the performance of other biometrics (like face).

To evaluate our system and to verify that the results are statistically meaningful, we perform our analysis on a large-scale anthropometric medical dataset available through the U.S. Center for Disease Control and prevention (CDC). The dataset includes more than 28000 subjects with the corresponding anthropometric measurements, and to the best of our knowledge this is the first time an anthropometry-based system includes this amount of users. Furthermore, in order to verify the performance of a complete system, we exploit the results of the pruning with a hard biometrics database. Therefore we check how the performance varies accordingly when we perform pruning with anthropometric measures and identification with a face recognition algorithm.

The article is structured as follow: we review works on anthropometric systems and search in biometric databases in Section 2. In Section 3 we present the anthropometric dataset used and we introduce the methodology of the study proposed to couple the pruning system with a face recognition algorithm. Finally in Section 4 we analyze the results obtained by our statistical analysis, and we show the gain in performance.

2 Previous work

2.1 Anthropometry-based recognition

After the historical example of the "Bertillonage" system, used in the 1882 for profiling and identifying prisoners (see figure 1), the first example of anthropometric study that involves identification of people is the one proposed by Daniels in [7]. This study presents the quest for the "Average man", *i.e.*a person with all the analyzed characteristics falling into average values (up to a given accuracy range). The author studies the possibility that such an individual could exists. The experiment involves a database of 4064 men of the Air Force flying personnel from which

131 measurements are extracted. Through an elimination process, the study shows how it is impossible that a person could belong to average classes in all his/her measures.

Later on, the study of Daniels was exploited by [8] to implement a people recognition system based on multiple biometrics. The article presents the result on fusing an anthropometric signature with the output of a gait analysis system. They reach 90% of accuracy in a database consisting of 48 individuals.

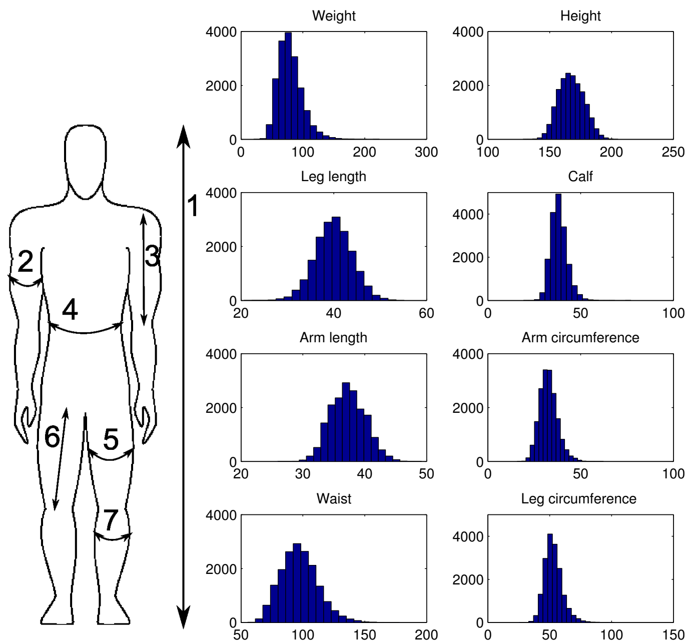
To the best of our knowledge, the latest work involving people recognition based on anthropometric measures is the one presented by Ober et al.in [4]. The authors exploit the CAESAR 3D dataset that contains approximately the 3D scans and 2D measures of 4400 individuals (divided by classes of gender, age, and weight). Ober et al.analysis involves 27 different anthropometric measures. Their approach is based on the dimensionality reduction of these 27 measures through Linear Discriminant Analysis (LDA) analysis. They reach 97% accuracy in a population of 2000 subjects.

Another interesting work [9] uses the 3D scans of the CAESAR dataset to analyze the effect of the error over the measures. They study the possible performance of an anthropometric recognition system as the reliability of the capture degrades. The original measures are altered adding error to the 3D landmarks to recreate different capturing conditions. Afterwards, they analyze the performance loss when looking for the altered data in the dataset.

All these methods start from the assumption that such a quantity of measures (27 in the case of [4]) should be easily available to the system in order to perform the identification. However, for some applications it would be challenging to obtain some of them (*e.g.*the foot breadth). Moreover, we believe that a real case scenario would be affected by errors similar to what studied by [9]. Thus making seriously challenging to create a system that could easily perform identification with noised input data.

2.2 Soft-biometrics based database pruning

A discussion on the possibility of using pruning via soft biometrics is already present in [10]; a later formalization of this idea is present in [3]. In [11] by Samangooei et al.semantic information are coupled with a gait signature to retrieve corresponding people from a database of individuals. Our case study differs from all these works as we do not consider the quantization step applied to each single feature. We believe that an error in the classification of one the features can prejudice the result of the pre-classification scheme and, subsequently, the performance of the recognition. Such a drawback does not apply in our experiment as the considered feature vector is composed of continuous values.



(a) (b)

Figure 2: Image (a) shows the measures included in the NHANES dataset. Additionally NHANES dataset include body weight that we used as feature. On the right, image (b) depicts the histograms of the 8 anthropometric measures.

3 Proposed case study

Each year the U.S. CDC office promotes the National Health and Nutrition Examination Survey (NHANES) [1], a study on American citizens devoted to understand health statistics behind a representative sample of the population. The importance of this database lays on its size, that makes it a suitable statistical source of information. If we consider the period taken into account in our study, we benefit of more than 28000 individuals recorded over a period that goes from 1999-2008.

This dataset is rich of heterogeneous information regarding everything related to health status of the subjects (medical analysis, demographics, and so on). A prevalent part of this dataset is composed of a set of anthropometric measures that is taken in controlled conditions. The measures are the following: (1) height, (2) arm circumference, (3) arm length, (4) waist circumference, (5) leg circumference, (6) leg length, (7) calf circumference, (8) body weight. The measures are shown in figure 2, along with the distributions regarding each of them along the entire dataset.

[1] http://www.cdc.gov/nchs/nhanes.htm

As all the databases of large dimension, NHANES dataset suffers of a considerable amount of missing data. Since in our case we cannot afford missing data in any of the previously mentioned measures, we discarded each subject that presented one of more missing measures. We discarded as well the subjects below 20 years old, as they had prevalence of missing data and became less represented in the dataset.

Afterward, we store the measures available in the dataset in a feature vector to be used in our pre-classification scheme. Prior to the normal identification module, we use those features to pre-classify the subject to be identified. In our case all the features are used as continuous value, and each of the features are considered of the same importance and then of the same discriminative power. One could argue that in order to select the features with most discriminative power, a study similar to [4, 5] should be conducted. In the first case LDA is used over a training set to extract the maximum information while discarding non useful information. In the second case, a brute force search is performed to check which feature combination performs best in estimating weight of a person. Nevertheless, to perform an analysis similar to the one of [4] we should consider a plethora of measures which is out of the scope of this paper as we aim at easily define the physical shape of each user.

The pruning capabilities offered by our body soft biometric signature are used to prune a hard biometric database. We use face recognition against a complete face dataset to show the gain in performance provided by our pruning scheme. In this case, the results are obtained using the well known FERET biometric face dataset that consists of 1195 subjects. We thus select a subset of people from the NHANES dataset so as to build a "chimera" database where each identity (face) is associated with a randomly chosen anthropometric feature vector [3]. Exploiting the code provided by [12] we create a baseline recognition algorithm that exploits eigenfaces. Another well known face dataset (ARFD) is employed to create the eigenspace that will be used to recover the identity of the test subjects. The recognition results are compared as cumulative matching characteristic (CMC) curves. We demonstrate how the pruning can be used not only to speed up the hard biometric algorithm, but how it can be exploited to identify the best performing parameters that allow an increase of performance without any loss of accuracy.

4 Experimental results

Our experimental results can be divided into two sections. In the first part we will analyze the pruning capabilities of an anthropomeasures based system as the mensuration error increases. Similar to [5, 9] each anthropometric measure will be biased by an error of increasing intensity to simulate a mensuration system. The results will discuss the amount of pruning acceptable as the mensuration error progressively increases. We will see that a trade off is possible so as to not interfere with the hard biometric recognition performance. The second part of this section

Measure	Mean	Std	Min	Max
Weight	7.9	1.9	2.5	21.8
Height	16.7	1.0	13.0	20.3
Leg length	3.9	0.4	2.2	5.4
Calf	3.8	0.4	2.1	7.5
Arm Length	3.7	0.3	2.6	4.8
Arm circ	3.2	0.5	1.7	6.1
Waist	9.7	1.5	5.9	17.5
Leg circ	5.2	0.7	2.7	10.0

Table 1: We summarize the statistics of 10% error magnitude. The units are expressed in kilograms for body weight, and centimeters for all the other measures.

will be devoted to the use of the pruning capabilities in a full system where the identification is performed using a face recognition algorithm. This algorithm is built and considered as baseline. This second part will also be composed of two different analysis: firstly we will discuss the achievable performance increase in accuracy that our anthropometric based pruning permit, and secondly, the trade off between accuracy, mensuration error, and penetration rate will be shown.

4.1 Anthropometry system performance analysis

Large scale systems of size comparable to the challenging Aadhaar indian project [13] will soon face problems due to the considerably high number of users to control. Exploiting the NHANES dataset we conduct our simulation over the entire dataset population, so as to guarantee statistical significance of the results. This provide us more than 17500 subjects from an original population size of almost 28000 individuals.

The literature of search and pruning of biometric databases proposes the *penetration rate* as measure to compare the performance of a pruning algorithm. It consists on the fraction of database we are able to isolate to perform the identification. The higher is the penetration rate via the analysis of our feature vector, the larger is the portion of database considered in the identification rate [2]. However, the pre-classification step (binning) can also be affected by an error. The binning error can impair the performance of the recognition algorithm performed afterwards.

In our case we consider our feature vector as the ensemble of the anthropometric measures. To rank our database we employ an euclidean distance metric as suggested in [9]. To simulate a real mensuration system we consider a varying error, and in order to increase the randomness of this bias, an approach similar to

[2] A penetration rate of 1 means that the full database is analyzed (*i.e.*no pruning is done), while a value of 0.5 means that half database is considered.

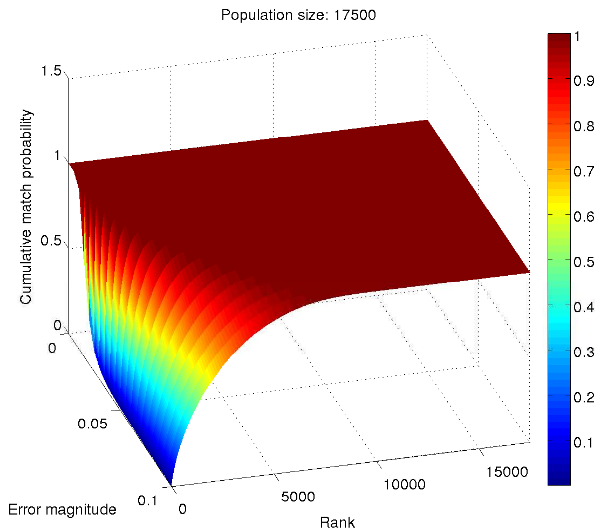


Figure 3: The plot shows the CMC curves as a 3D surface.

the one proposed by [5] is used, where the error is applied varying randomly the error for each measure.

Considering our feature vector $F = [f_1, \ldots, f_n]$ we add to each measure separately an error proportional to the magnitude of the original feature and with random sign, so as to obtain its biased version F_ϵ:

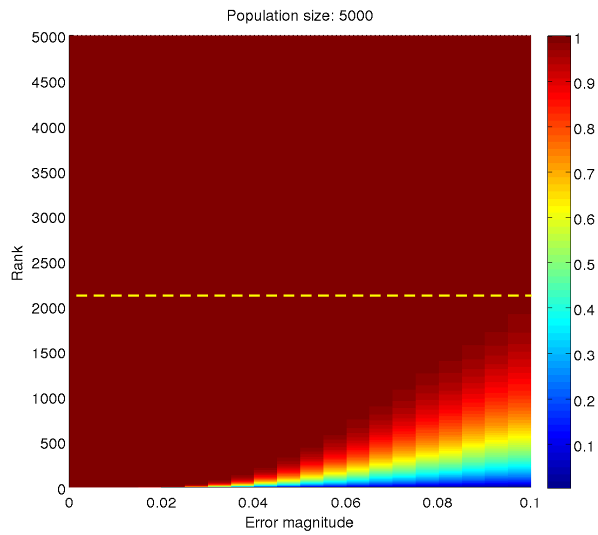
$$F_\epsilon = [f_1(1 + \alpha_1 w_1), \ldots, f_n(1 + \alpha_n w_n)] \quad (1)$$

where α_i is a binary random variable that takes values in the set $\{-1, +1\}$, and the error is computed as function of the original value (f_i) times the power assigned to the noise at a given iteration (w_i).

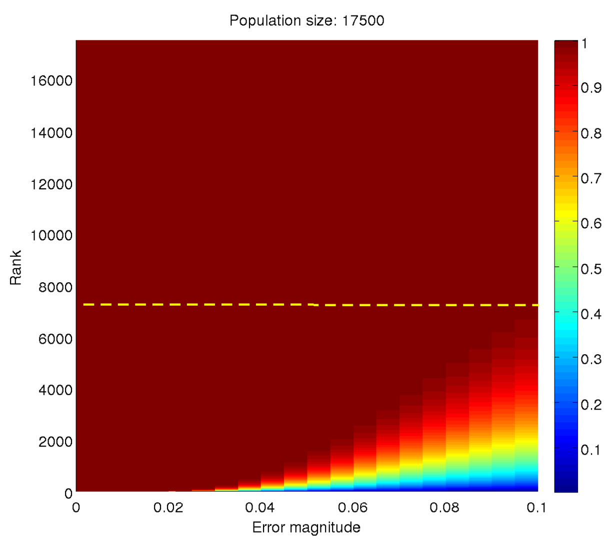
To compare our error to the one observed in [4, 9] we summarize the error statistics in Table 1. Considering the average error, one can see that we are close to what one could expect from a worst case scenario.

For each given error magnitude we iterate over the entire dataset, at each iteration every biased vector is compared against the original dataset and a distance matrix is built. From the distance matrix we are able to compute a cumulative matching characteristic curve (CMC) that summarizes the performance of the pruning. The curve indicates the probability of observing the client in the first N-best candidates.

At the end of our analysis we obtain a CMC curve for each error magnitude considered. We can hence plot the result as 3D plot (fig. 3) in order to show the trend of the graph; or in 2D (fig. 4) to better compare different experiments.

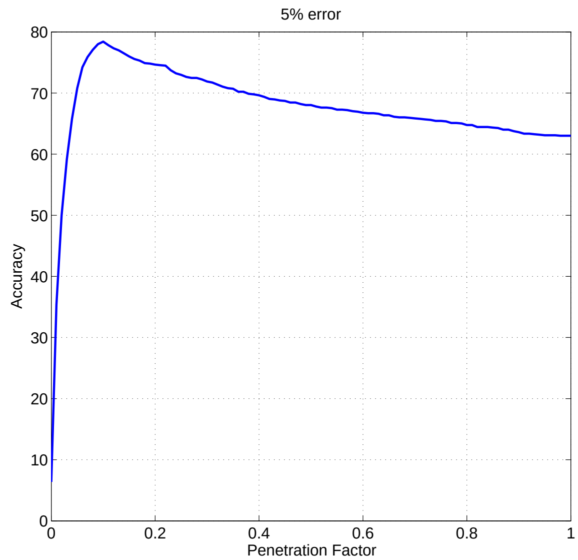


(a)

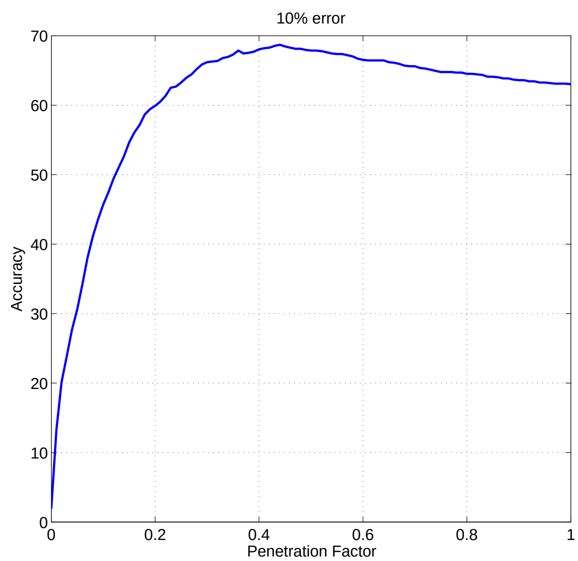


(b)

Figure 4: The plot shows the CMC curves for two different population sizes: (a) 5000, (b) 18000 subjects. The intensity of the color indicates the probability of the CMC curve.



(a)



(b)

Figure 5: Rank one accuracy variation as the Penetration rate increase. Error at 5% (a) and 10% (b) magnitude are considered.

In figure 3 we can clearly see that if the error magnitude increases, we have to consider bigger portions of the results to guarantee that the client is among the first results.

We performed our analysis over different population sizes drawn randomly from the original dataset. For the sake of brevity, in figure 4 we show the experiments conducted with 5000 and 17500 subjects considered, respectively 30% and 100% of our dataset. In the first case the experiment was conducted 100 times and the results were averaged.

If we consider the maximum values admitted for the error magnitude (10% of the original value) we clearly see that a penetration rate of 50% can be obtained with no effort. In the first case (4.a) the client is always (probability $= 1$) within the first 2000 results. We remark that this penetration rate corresponds to 100% accuracy of the system, thus it cannot reduce the performance of the recognition algorithm applied afterwards. The second case (4.b), that considers the full original dataset, confirm our results since 50% of the dataset (8000 subjects in this case) is still a valid choice to have 100% accuracy of the pruning system.

4.2 Recognition accuracy increase by pruning

As previously mentioned, the performance of a recognition algorithm could possibly be affected by the error introduced by the pre-classification algorithm, or contrarily the recognition algorithm could benefit from the pruning both in speed and accuracy. The analysis we conducted in the previous part, showed that a conservative choice could be the selection of just half database, which will preserve the performance of the recognition algorithm even in the worst case (that we consider being the 10% magnitude error).

To verify our assumption we analyze in this section the performance of the cascade composed by our pre-classification scheme based on anthropometric measures, and a face recognition baseline system. The baseline produces at rank 1 a result of 63% accuracy. To analyze how the pruning affects this accuracy, we decided to consider two error magnitudes: 5% and 10%. Although a conservative choice could be made by selecting the penetration rate that provides us with the certainty of finding the client in pruned results, for the sake of completeness we tested our full system for increasing values of penetration rate. This analysis shows us the best trade-off between pruning and accuracy and leads us to the best performing system. At each iteration we used the anthropometric feature vector associated with the identity to perform a fast search in the anthropometric dataset. After all the distances are computed and the identities are ranked accordingly, the face recognition is performed on the subset defined by the penetration rate.

Figure 5 summarizes the results obtained by such analysis. Our assumptions are confirmed: as we consider the lowest penetration rate, the final recognition accuracy suffers from the poor performances of the pruning algorithm, that is not able to provide the client with a given certainty (*e.g.*in the range $[0, 0.1]$ of fig 5.a). In both cases the curves behavior is very similar; the performance at rank 1 accuracy increases up to a global maximum, and then, asymptotically, falls back to the baseline accuracy result (63%) as the portion of pruned identities considered gets larger. Indeed, as we consider more subjects in the recognition process, we analyze more images that could include some face samples that increase the false acceptance rate. The maximum corresponds to the operating point that best put together the benefit of the pruning and recognition algorithm. By selecting the two maxima we can define the optimum operating points for the two systems (see fig. 5) that have to operate with 5% and 10% error magnitude. In the first system the maximum indicates the best penetration rate of 0.1, then 10% of the pruned results; while in the second case the range falls between 0.4 and 0.5.

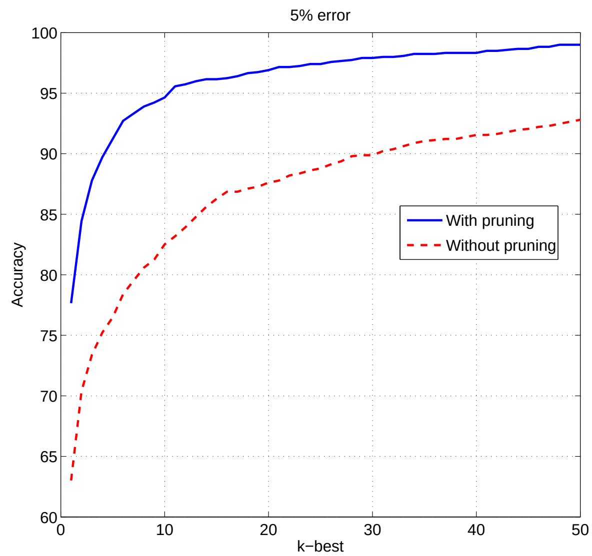
Furthermore, to complete our analysis we show in figure 6 the full rank curves of the two systems. In the first case (fig. 6.a) the chosen penetration rate consider the first 100-best while in the second case (fig. 6.b) the top 600 results obtained by the pruning algorithm are further analyzed. The gain is both high in the sense of accuracy and speed performance (since the penetration rate is smaller) in the former, while in the latter both the gains are reduced, but still our system performs better than the baseline algorithm and we are able to prune half database out from our recognition, speeding up the recognition phase by a factor of $2\times$.

In case of a recognition algorithm more expensive in terms of resources, one could be interested in reducing at the minimum the search space size. Therefore, a measure of the loss of performance caused by selecting a smaller penetration rate is needed.

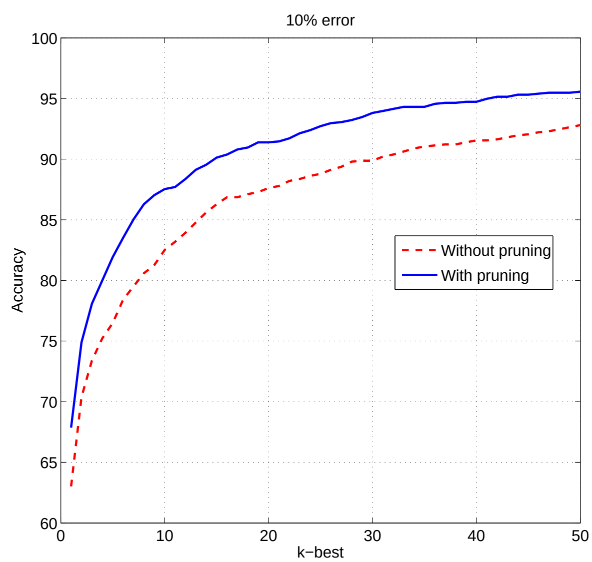
Then, in order to analyze the full response of the system to both the penetration rate and the error in the anthropometry mensuration system, we present the result of figure 7. Here we show the performance of the global system in terms of rank 1 results as both the penetration factor and the error magnitude vary. One could exploit such graph to understand whether to leverage on the first parameter to obtain a faster system, or to invest into a better mensuration system to lower down the second factor, thus approaching the best possible results. In our case we can clearly see that choosing small penetration rate values we can guarantee good performance only if we are able to lower down the error of the mensuration system. If we are not aware of the amount of error generated by our mensuration system, a good choice could be a penetration rate of 20% (*i.e.*the face recognition algorithm would check only 1 image out of 5). That choice guarantees us very good performance in case of 5% error magnitude, and close to the baseline rank 1 accuracy in case of 10% error.

5 Conclusion

We presented a work that exploits body soft biometric traits to pre-classify people of a database to improve both speed and accuracy of a hard biometric identification system. The anthropometric measures used as feature vector were extracted from a large scale medical dataset, thus guaranteeing statistical meaning



(a)



(b)

Figure 6: The accuracy increase achievable after setting the penetration factor at its best in both (a) 5% and (b) 10% noise magnitude.

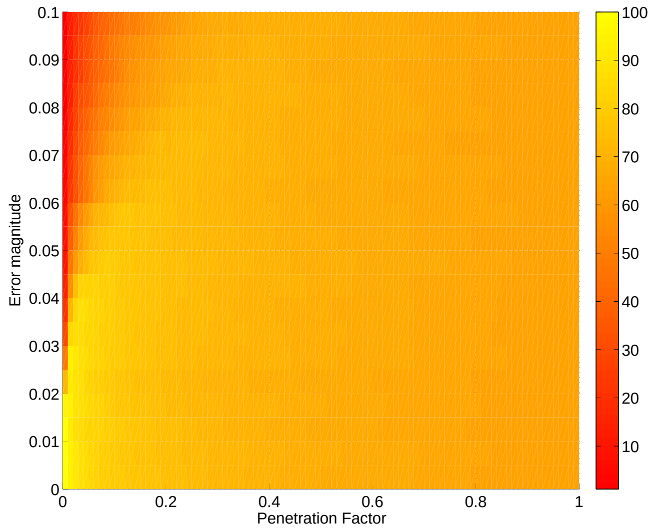


Figure 7: The graph shows the accuracy achievable by the cascade of our pruning module followed by the face recognition one. The plot helps identifying the best operational point. One can see that lowest penetration factors are effective only if mensuration error is reduced accordingly. By knowing the possible maximum error of our mensuration scheme, it is easy to understand which penetration factor to use in order to maximize the performance of our system.

of the results. Our work clearly identifies the trade off between penetration rate (*i.e.*the speed up achievable by pruning the dataset) and the accuracy of the identification performed by the system after pruning, and the response to the noise of the mensuration system. The behavior of a complete system is analyzed by using a well known face recognition technique and database in cascade to our preclassification module. Our results distinctly show that even in the worst case of $\pm 10\%$ error magnitude in the anthropometric measures, the pruning system is able to speed up the search of $2\times$ factor while guaranteeing an increase of accuracy performance. A normal continuation of this work could focus on the identification of a confidence interval for each of the measures, so that the error considered will not depend exclusively on the magnitude of the measure itself, but also on the mensuration system. Moreover, increasing the number of anthropometric measurements considered in this study will undoubtedly ameliorate the performance of the classification scheme.

References

[1] A.K. Jain, S.C. Dass, and K. Nandakumar. Soft biometric traits for personal recognition systems. *Biometric Authentication*, pages 1–40, 2004.

[2] A.K. Jain, S.C. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? *age*, 20:39.

[3] A. Dantcheva, C. Velardo, A. D'angelo, and J.-L. Dugelay. Bag of soft biometrics for person identification : New trends and challenges. *Mutimedia Tools and Applications, Springer, October 2010*, 2010.

[4] D.B. Ober, S.P. Neugebauer, and P.A. Sallee. Training and feature-reduction techniques for human identification using anthropometry. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–8. IEEE.

[5] C. Velardo and J.-L. Dugelay. Weight estimation from visual body appearance. In *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, September 27-29, 2010, Washington DC, USA*, 09 2010.

[6] V. Tillmann and PE Clayton. Diurnal variation in height and the reliability of height measurements using stretched and unstretched techniques in the evaluation of short-term growth. *Annals of Human Biology*, 28(2):195–206, 2001.

[7] G.S. Daniels. The "average man" ? Technical report, Air Force Aerospace Medical Research Lab Wright-Patterson, 1952.

[8] R.D. Green and L. Guan. Quantifying and recognizing human movement patterns from monocular video images-part ii: applications to biometrics. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(2):191–198, 2004.

[9] A. Godil, P. Grother, and S. Ressler. Human identification from body shape. In *3-D Digital Imaging and Modeling, 2003*, pages 386–392, 2003.

[10] J.L. Wayman. Large-scale civilian biometric systems - issues and feasibility. Card Tech / Secur Tech ID, 1997.

[11] S. Samangooei and M. Nixon. Performing content-based retrieval of humans using gait biometrics. *Semantic Multimedia*, pages 105–120, 2008.

[12] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of pca, ica, and lda on the feret data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005.

[13] R. Dass and S. Pal. Challenges of identity management a context in rural india. In *Information Security and Digital Forensics*, volume 41 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pages 172–183. Springer Berlin Heidelberg, 2010.