Person Recognition using Soft Biometrics

Date of Submission: 27th April, 2017

Nipun Jain 2013B3A7608P

Pranav Misra 2014A7PS096P

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

Table of Contents

1. Abstract

The field of person recognition has seen extensive research work in the recent years. A wide array of person-recognition systems exists that perform automatic recognition based on different physiological and behavioural characteristics like face, fingerprint, iris, gait etc. Our work is based on implementation of a Soft Biometrics based recognition system. Soft Biometrics can be defined as traits that contain some information about the user, however, lack the necessary discrimination power to recognize a person with a good efficiency. Here, we analyse the effectiveness of a Soft Biometrics based recognition system.

2. Introduction

The field of person recognition has seen extensive research work in the recent years. A wide array of person-recognition systems exists that perform automatic recognition based on different physiological and behavioural characteristics like face, fingerprint, iris, gait etc. The person recognition systems that use just one trait for recognition are called Unimodal Biometric Recognition Systems. However, these systems suffer with problems like noise in data, lack of distinctiveness power of the trait etc. Some of these problems can be solved by using multiple traits for recognition. Such systems are called Multi-modal recognition systems. A typical multimodal recognition system combines various features like face, fingerprint, gait etc to improve the overall efficiency of the system. Most multimodal systems consist of a primary recognition system and a secondary system that uses features in addition to those used in primary system.

Most of the secondary recognition systems that rely on fingerprint, and iris have are intrusive in nature. They require assistance from the user end to collect the required data. A possible, less intrusive and autonomous secondary system can involve using features like height, weight, age, gender, skin color, eye color, ethnicity etc. Such features are called as Soft- Biometrics. Soft Biometrics can be defined as traits that contain some information about the user, however, lack the necessary discrimination power to recognize a person with a good efficiency.

We propose a proof-of-concept system that establishes the feasibility of Soft Biometrics as a suitable secondary biometric recognition system.

3. Literature Survey

A plethora of existing research exists in the field of person recognition. Most of the research is focused on Unimodal recognition systems based on some of the prominent features like face, fingerprint, and iris. Very limited research exists in Multimodal systems that use Soft Biometrics.

In a research, Heckathorn et al. [4] showed that soft biometrics attributes like age, race, sex, height, eye color and other visible identification marks could only identify people with limited accuracy. Thereby making it unfeasible to develop a person recognition system based solely on Soft Biometrics.

Wayman [5] used Soft Biometrics to reduce the search space for the primary biometric system. For example, if the person to be verified is a young female, the search space is reduced now to only match against people that satisfy the above mentioned criteria. Using Soft Biometrics improves the overall efficiency and search speed. Givens et al. [6] showed that factors like race, age and occupation negatively impacted the efficiency of Recognition systems. uch as age, gender, race, and occupation can affect the performance of a biometric system. For example, a young female Asian mine-worker is seen as the most difficult subject for a fingerprint system [7].

Tome et al. [1] studied the relation between the distance of the person from the input device and performance of Soft Biometric systems. They developed ways to integrate soft biometric traits to handle different distances. Their study showed an improvement in the overall performance of the recognition system when the primary recognition system used Soft biometrics as a secondary recognition system.

Jain et al. [7] developed a mathematical framework to integrate the information from soft biometric traits with the output of primary biometric systems using Bayesian decision theory. They demonstrated that using information from soft biometric traits like ethnicity, height, gender improved the performance of traditional biometric systems like fingerprint.

4. Methodology

The data was first analysed using some statistics like the discrimination power of the attributes , the stability of the human annotations for all the attributes, and correlation between different attributes. Further, a proof-of-concept verification system was developed to analyse the feasibility of Soft Biometrics as a reliable recognition system.

4.1. Data

The analysis has been performed on the Southampton University Tunnel DB Soft Biometrics Database (TunnelDBSoftBio) [1][3]. The database comprises of manually annotated data for 23 different physical traits, from 10 human annotators for 58 different users. The physical traits comprise of several nominal and ordinal traits like sex, age, height, skin colour etc. Table I. shows a complete list of all the physical traits included in the database. For further analysis the values for each attribute is converted to numerical values accordingly.

Body

Trait	Range of Values
1. Arm Length	Very Short, Short, Average, Long and Very Long
2. Arm Thickness	Very Thin, Thin, Average, Thick and Very Thick
3. Chest	Very Slim, Slim, Average, Large and Very Large
4. Figure	Very Small, Small, Average, Large and Very Large
5. Height	Very Short, Short, Average, Tall and Very Tall
6. Hips	Very Narrow, Narrow, Average, Broad and Very Broad
7. Leg Length	Very Short, Short, Average, Long and Very Long
8. Leg Direction	Very Bowed, Bowed, Straight, Knock Kneed and Very Knock Kneed
9. Leg Thickness	Very Thin, Thin, Average, Thick and Very Thick
10. Muscle Build	Very Lean, Lean, Average, Muscly and Very Muscly
11. Proportions	Average and Unusual
12. Shoulder Shape	Very Rounded, Rounded, Average, Square and Very Square
13. Weight	Very Thin, Thin, Average, Big and Very Big

Global

Trait	Range of Values
14. Age	Infant, Pre Adolescence, Adolescence, Young Adult, Adult, Middle Aged, Senior
15. Ethnicity	European, Middle Eastern, Indian/Pakistan, Far Eastern, Black, Mixed, Other
16. Sex	Female, Male

Head

Trait	Range of Values
17. Skin Color	White, Tanned, Oriental and Black
18. Facial Hair Color	None, Black, Brown, Red, Blond and Grey
19. Facial Hair Length	None, Stubble, Moustache, Goatee and Full Beard
20. Hair Color	Black, Brown, Red, Blond, Grey and Dyed
21. Hair Length	None, Shaven, Short, Medium and Long
22. Neck Length	Very Short, Short, Medium and Long
23. Neck Thickness	Very Thin, Thin, Average, Thick and Very Thick

Table I. Physical traits and their values

The physical traits can be categorised into the following categories:-

1. Body - These attributes describe the user's perceived somatotype. These are correlated with the type of clothes a user wears.
2. Global - These attributes typically remain constant over the lifetime of a user.
3. Head - These are the attributes which are primarily used by us humans to identify people.

4.1.1. Correlation

For the set of 580 annotations, Pearson's correlation coefficient was calculated for all the pairs of attributes. The coefficient of correlation is calculated as:

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}} \quad (1)$$

σ_{XY} is the covariance between X and Y. σ_X and σ_Y are standard deviation of X and Y respectively. The coefficient of correlation varies from -1 to 1, with 1 meaning variables are perfectly positively correlated and -1 meaning the variables are perfectly negatively correlated.

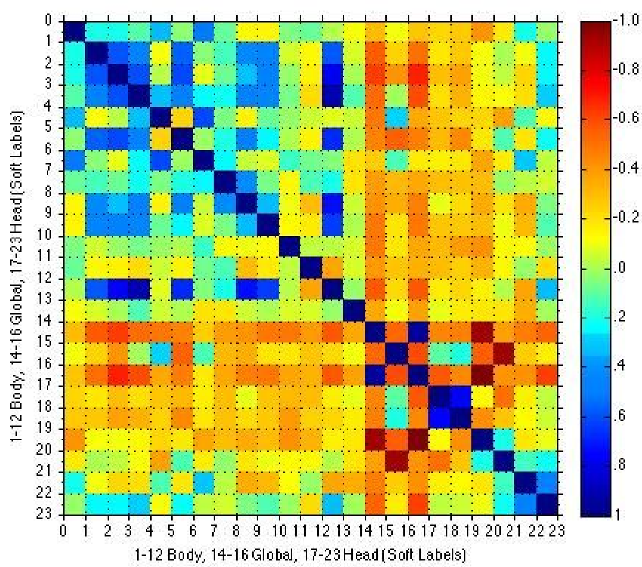

Fig 1. Coefficient of Correlation between attributes

Fig 1. shows the coefficient of correlation matrix for all the pairs formed between the 23 attributes. Colours near red represent a coefficient of correlation close to -1 while colours near blue represent coefficient of correlation close to 1.

4.1.2. Stability

Here we find the stability of the annotations made by the human annotators for all the attributes.

stability coefficient is defined as:

$$\text{Stability}_X = 1 - \frac{1}{SA}\sum_{i=1}^{S}\sum_{a=1}^{A}|X_{ia} - \text{mode}_a(X_{ia})| \quad (2)$$

In this equation, X_{ia} is the annotated value for subject i by annotator a, A = 10 is the total number of annotators, S = 58 is the total number of subjects, and $\text{mode}_a(X_{ia})$ is the statistical mode across annotators (the value most often annotated for subject i).

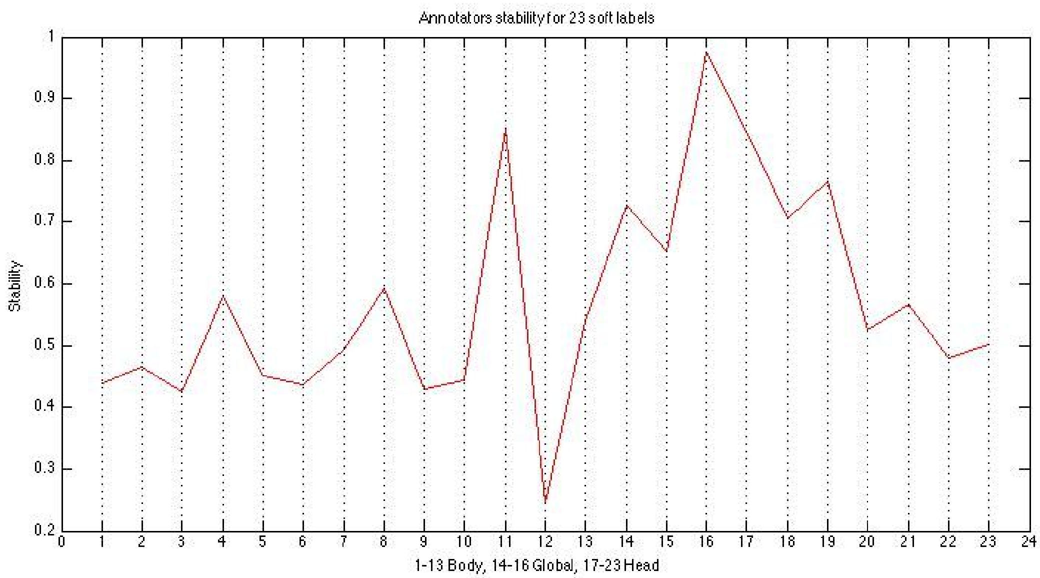

Fig 2. Annotators' Stability for the 23 soft labels

We can see that gender is the most stable attribute amongst all due to low variability. Attributes like arm length, leg length etc. have low stability due to high variability in the values.

4.1.3. Discrimination Power

Discrimination power of variable X is calculated as:

$$\text{Discrimination}_X = \frac{\frac{1}{S(S-1)}\sum_{i=1,i\neq j}^{S}\sum_{j=1}^{S}|\mu_i - \mu_j|}{\sigma} \quad (3)$$

$$\mu_i = \underset{a}{\text{mean}}(X_{ia}), \quad \mu_j = \underset{a}{\text{mean}}(X_{ja}), \quad \sigma = \frac{1}{S}\sum_{i=1}^{S}\sigma_i \quad (4)$$

where $\sigma_i = \text{std}_a(X_{ia})$, i and j index subjects, and a indexes annotators.

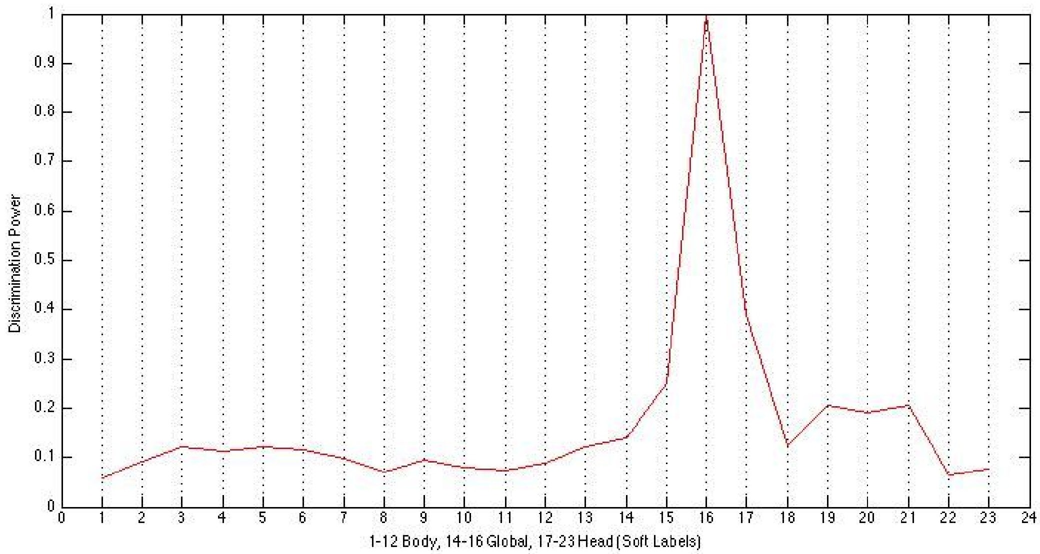

Fig. 3 Discrimination Power of Attributes

As can be seen in the figure above, the global and head traits are more discriminant than body attributes. Traits like skin colour, and ethnicity have high discrimination power. Sex has the highest discrimination power due to the easily identifiable nature of the attribute by the annotators.

4.2 Verification/Identification System

Person recognition using soft biometrics was verified using similarity based matching and also using a neural network. Out of the 580 user records (58 users x 10 annotators), 58 records (1 per user) were separated and used as the test data for the verification system. The rest of 522 records were used as the database to which the test users were matched. The data was tested in the verification system using normalised as well as the data the values of which were not normalised. The data values were normalised in the range 0 to 1 using a tanh estimator [1]:

$$X'^k = \frac{1}{2}\left\{\tanh\left(0.01\left(\frac{X^k - \mu_{X^k}}{\sigma_{X^k}}\right)\right) + 1\right\} \qquad (5)$$

In equation 5 X_k is the k = {1,...,K} soft label (K = 23), X'_k denotes the normalized label, and μ_X^k and σ_X^k are respectively the estimated mean and standard deviation of the label under consideration (see Table I for the list of the labels).

For similarity based matching three different types of cost functions were used when measuring the similarity between attributes. For each cost function the identification of person is based on minimizing the cost when the input attribute vector of 23 values is compared against 522 target vectors, with 9 target vector per unique person id i.e 58 x 9 gives 522. The first cost function is a trivial cost function summing the absolute difference in vector attributes, for all 23 attributes.

$$Cost = \Sigma_i |u_i-v_i| \quad i \in \{1...23\} \qquad (6)$$

The second cost function scales the absolute difference depending on the discrimination power of the attribute.

$$Cost = \Sigma_i |u_i-v_i|w_i \quad i \in \{1...23\} \qquad (7)$$

In the above equation, w_i stands for the weight of i^{th} attribute given by Eq. 3 and can be inferred from Fig. 3. The third cost function is also based on Eq. 3 and Fig 3 but the value added to the cost is not proportional to the w_i for all $|u_i-v_i|$. It assigns a penalty of $2^* |u_i-v_i|$ when wi is greater than 0.1 but less than 0.3 and $|u_i-v_i| <=1$. It assigns a penalty of $10^* |u_i-v_i|$ when wi is greater than 0.3 and $|u_i-v_i| >=1$. It assigns a penalty of $|u_i-v_i|^2$ when wi is greater than 0.1 but less than 0.3 and $|u_i-v_i| >1$. It assigns a penalty of -3.9 when wi >0.3 and $|u_i-v_i| ==0$. It assigns a penalty of $|u_i-v_i|$ for wi < 0.1. These penalties are added to the cost function.

For verification using neural network, a network using scaled conjugate gradient descent and 45 hidden neurons was used. Training was stopped after 6 validations. The target matrix was an array of 58x522 with 9 entries corresponding to the row number id as ones and all other entries of that row as zeros. The training accuracy was 23.68 %. Normalised values of attributes were used for training. For testing the normalised testing data was used, and the targets were 58 unique values corresponding to 58 people used in the experiment.

5. Results

The concept of soft biometrics in the study conducted uses discrete values for all the attributes. This results in a finite number of permutations of word based soft attributes that can be assigned to the subjects. Therefore, for a large enough data soft biometrics can only be used to reduce the search space before using more computationally expensive techniques to identify a person. Keeping this in mind, the number of correct outputs as seen in the top three results predicted by the verification/identification system for each type of cost function in the similarity based approach and also the neural network were plotted. The second and the third type of cost functions that used the discriminative power of the attributes to assign the cost did relatively better than the trivial cost function. The third cost function performed the best with 77.6% accuracy for the test data and was followed by the second type of cost function with 70.7% accuracy. The accuracy of the neural network was 69.0%. Figure 4 shows the comparison of the accuracy various approaches for their top 3 predictions as seen on the test data.

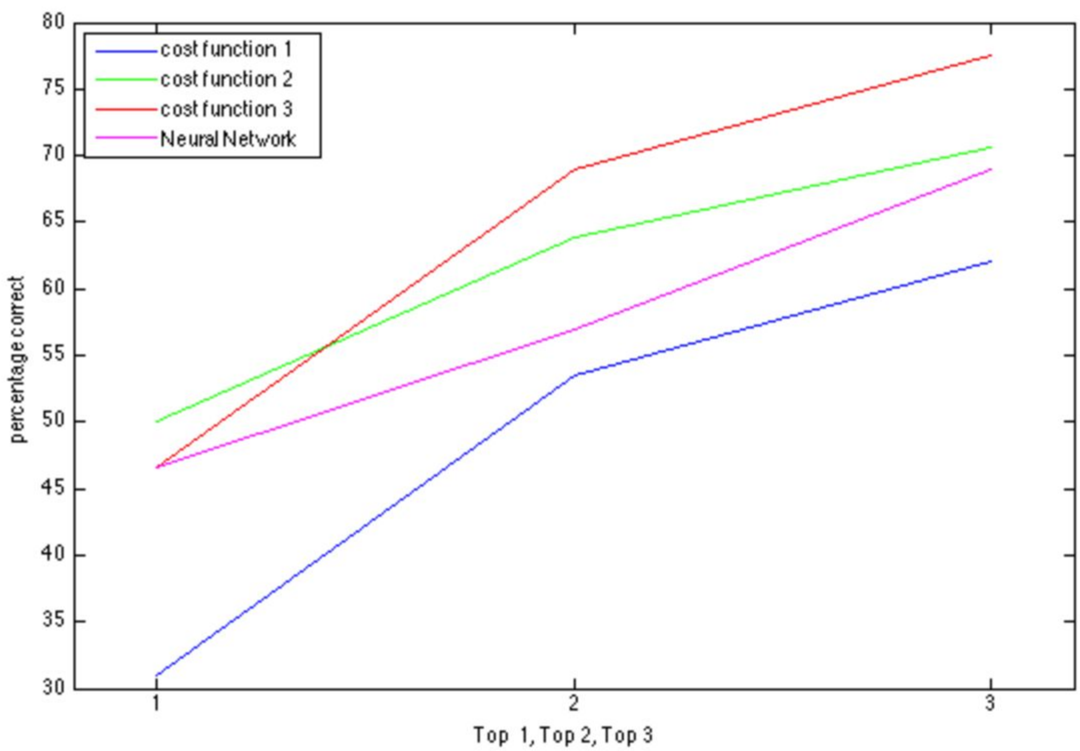

Figure 4. Showing percentage correct in top three values predicted. Results are cumulative.

6. Conclusions

The concept of word based soft biometrics assigns discrete values to all the physical traits. This results in a finite number of permutations of word based soft attributes that can be assigned to the subjects. This is an inherent limitation of this technique. Hence, for a large enough dataset soft biometrics can only be used to reduce the search space before using more computationally expensive techniques to identify a person.

Our findings suggest that Soft Biometrics based recognition system give a satisfactory best accuracy of 77.6%. This is coherent with our aim that was to give a proof of concept that soft biometrics and word based computing can be used effectively to reduce the search space when trying to identify or verify a person. Hence, Soft Biometrics based recognition systems are an excellent choice as a secondary recognition system of a larger Multimodal biometric systems.

Future extension of our work involves automatic extraction of Soft Biometrics attributes. Existing research in this field has not yielded satisfactory results. However, certain features like height, weight and sex can be automatically extracted with satisfactory efficiency. These are some really challenging problems themselves and considerable research should go into solving these.

7. References

[1] P. Tome, J. Fierrez, R. Vera-Rodriguez, M. S. Nixon, "Soft Biometrics and their Application in Person Recognition at a Distance", in IEEE Transactions on Information, Forensic and Security (TIFS), 2014. (DOI: 10.1109/TIFS.2014.2299975)

[2] D. Reid, M. S. Nixon and S. V. Stevenage, "Soft Biometrics; Human Identification using Comparative Descriptions", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013. (DOI: 10.1109/TPAMI.2013.219)

[3] https://atvs.ii.uam.es/atvs/tsb_db.html

[4] Heckathorn, D.D., Broadhead, R.S., Sergeyev, B.: A Methodology for Reducing Respondent Duplication and Impersonation in Samples of Hidden Populations. In: Annual Meeting of the American Sociological Association, Toronto, Canada (1997)

[5] Wayman, J.L.: Large-scale Civilian Biometric Systems - Issues and Feasibility. In: Proceedings of Card Tech / Secure Tech ID. (1997)

[6] Givens, G., Beveridge, J.R., Draper, B.A., Bolme, D.: A Statistical Assessment of Subject Factors in the PCA Recognition of Human Subjects. In: Proceedings of CVPR Workshop: Statistical Analysis in Computer Vision. (2003)

[7] Jain, A. K., Dass, S. C., & Nandakumar, K. (2004). Soft Biometric Traits for Personal Recognition Systems. *Proc. of International Conference on Biometric Authentication (ICBA)*, (July), 731–738. https://doi.org/10.1007/978-3-540-25948-0_99