Multi-modal Open-Set Person Identification in HRI

Bahar Irfan
CRNS, Plymouth University, UK
bahar.irfan@plymouth.ac.uk

Natalia Lyubova
SoftBank Robotics Europe, France
nlyubova@softbankrobotics.com

Michael Garcia Ortiz
SoftBank Robotics Europe, France
mgarciaortiz@softbankrobotics.com

Tony Belpaeme
IDLab - imec, Ghent University, Belgium
CRNS, Plymouth University, UK
tony.belpaeme@plymouth.ac.uk

ABSTRACT

In this paper, we describe a multi-modal Bayesian network for person recognition in a HRI context, combining information about a person's face, gender, age, and height estimates, with the time of interaction. We conduct an initial study with 14 participants over a four-week period to validate the system and learn the optimal weights for each of the metrics. Several normalisation methods are compared for different settings, such as learning from data, face recognition threshold and quality of the estimation. The results show that the proposed network improves the overall recognition rate by at least 1.4% comparing to person recognition based on face only in an open-set identification problem, and at least 4.4% in a closed-set.

KEYWORDS

Person recognition; Bayesian network; multi-modal data fusion; soft biometrics; personalisation

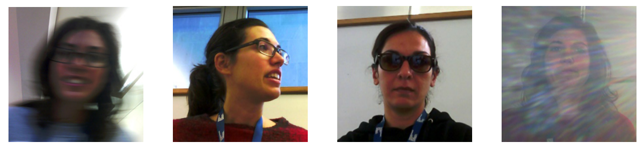
1 INTRODUCTION

Recognising a person is an essential step in establishing a personalised long-term human-robot interaction (HRI). In contrast to verification problems, where a user would state her identity and the system confirms or rejects it, in an HRI scenario, automatic recognition is desired for a natural interaction. In addition, the user might not be encountered before, in which case, the robot is expected to "meet" the user, i.e. enroll the user into the system. This problem is classified as an *open-set identification problem*, which is more difficult than closed-set identification or verification problems [5].

Biometric systems generally perform user recognition based on face recognition (FR). However, most FR challenges such as Face Recognition Vendor Tests[1] evaluate algorithms that perform verification. To this date, the only available open-set identification challenge is the Unconstrained Face Detection and Open Set Recognition Challenge[2], which shows that the algorithms achieve good identification accuracies at high false identification rates [6].

[1]https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt
[2]http://vast.uccs.edu/Opensetface

HRI '18 Workshop, March 5-8, 2018, Chicago, IL, USA



(a) (b) (c) (d)

Figure 1: Examples of unreliable face recognition from our study: (a) a blurry image; (b) an oblique viewing angle; (c) occlusions, e.g. glasses; (d) lighting condition, e.g. direct light.

Similarly, during a real-time interaction, FR could be unreliable due to a number of reasons including changing facial features, expressions, and lighting conditions [16] (see Fig. 1). Another example is the recent release of a smart phone with the built-in FR system for unlocking the phone that struggle to distinguish family members due to similarity of their facial features [4]. This issue increased awareness of the security and privacy problems that might arise from using a uni-modal biometric system that might not be as reliable as using a pass code.

Moreover, a biometric system may not be able to obtain meaningful data in some cases, resulting in a *failure-to-enroll* (FTE) error [13]. For instance, a face may not be detected in a blurry image where a person is moving. In addition, the upper bound on identification accuracy would limit the matching performance of a uni-modal biometric system. However, multi-modal biometric systems can improve the matching accuracy of a recognition by fusing information from multiple sources that could reduce the effects of noisy data, decrease FTE error, and eliminate the upper bound issue for a better determination of the identity. Robots, due to the rich sensor suite they carry, lend themselves well to multi-modal recognition.

In this paper, we explore a multi-modal Bayesian network (BN) for integrating soft biometric information, such as a user's gender, age, height, and time of interaction, together with the primary biometric information provided by face recognition. These biometric modalities are non-intrusive and can be obtained using the camera embedded on the robot. We designed a pilot study to validate our system in a real-time HRI scenario. We compare performances of several normalisation methods using optimised weights for each comparison. The proposed recognition system is intended to be used in a real-world application in Cardiac Rehabilitation therapy with a personalised robot [10].

2 RELATED WORK

Several post-classification fusion methods have been proposed for integrating multi-modal information in biometrics. They can be classified into three main categories: decision, rank, and confidence level fusion [8]. Decision level fusion (e.g. majority voting, AND/OR rule) is based on combining individual best matches from each biometric matcher. Rank level methods (e.g. highest rank, logistic regression) are used when the output of each biometric matcher consists of ranked matches. Confidence level fusion is the most common approach, as it allows a weighted decision from multiple biometric classifiers. There are two main approaches for combining the scores for confidence level fusion: classification methods and combination approaches. Classification methods treat the classifications of individual classifiers as input for a new classifier, such as ANNs and Support Vector Machines, which allows combining non-homogeneous data without preprocessing. The combination approach on the other hand consist of three steps: (1) normalisation of scores from different modalities into a common domain, (2) combination of scores through a method such as *sum* or *product rule*, and (3) thresholding to obtain the identification results. Performances of the combination approaches depend on the method and threshold chosen at each of the steps.

Although most biometric systems utilise primary biometrics for person recognition, such as fingerprint or face, other attributes of an individual such as age, gender, and clothing –referred to as soft biometrics– can provide additional information to improve the recognition performance [2]. In [9], the authors proposed combining a primary biometric trait (fingerprint) with soft biometric traits, such as a person's gender, ethnicity, and height, using a BN. In the weighting scheme, the traits with smaller variability and larger distinguishing capability were given more weight in the computation of the final matching probabilities. Furthermore, smaller weights were assigned to the soft biometric traits, so that if a soft biometric trait is measured incorrectly (e.g. a male user being identified as a female) the rejection probability is decreased. They achieved a 4% improvement in the genuine acceptance rate, however, the fusion weights were not optimised.

To the best of our knowledge, our approach is the first in combining soft biometrics with a primary biometric to identify a user in real-time, in the field of HRI. Moreover, it is the first time that the presented modalities (face, gender, age, height and time of interaction) are fused together, although they have shown improvement when fused separately or with other biometrics [2, 11, 15].

3 METHODOLOGY

We developed a BN based on [9] integrating multi-modal biometrics for reliable recognition in real-time human-robot interaction. We fuse face recognition (F) information (primary biometric) with gender (G), age (A), and height (H) estimations and the time of the interaction (T) (soft biometrics). Conditional independence is assumed between nodes, given the identity (I). The pyAgrum [3] library is used for implementing the structure.

3.1 Structure

The states of each node are determined by: the number of known users (for F and I), the available range of the modality (for A and H), and the pre-defined values (for G, "female" and "male", and for T, the day of the week combined with the time of the interaction). The data available about each user are converted to probabilities for each state within a node. These values are used as evidence in the network, and the maximum posterior for the I node determines the estimated identity of the user.

FR values are assumed to be similarity scores, such that, each score gives the percentage of similarity of the current user to the faces in the database. These scores are normalised to find the probabilities of the states. Age, height, and time are considered as discrete random variables (e.g. age is taken as 26, between 26 and 27). We estimate the probabilities of the remaining states by assuming a discretised and normalised normal distribution, $N(\mu, \sigma^2)$, defined by Eq. 1, where X is the estimated value, Z is the z-score, and C is the confidence of the biometric indicator for the estimated value.

$$\mu = X, \quad P(\frac{-0.5}{\sigma} < Z < \frac{0.5}{\sigma}) = C \tag{1}$$

Generally, in FR systems, if the highest similarity score or probability is below a given threshold, the user is declared as "unknown". However, in a BN, the posterior probabilities can be quite low due to the multiplication of probabilities during inference and the value can decrease with increasing number of states in a node. Thus, instead of using a fixed threshold in our system, we use the quality of the estimation (Q), in which we compare the highest probability (P_w) to the second highest probability (P_s), as shown in Eq. 2. The difference is multiplied by the number of people in the database (n_p), because the difference between the probabilities decreases as the number of people increases (the sum of probabilities is 1.0). Initially $Q = 0$, which eliminates the cases where the first and the second highest probabilities are the same.

$$Q = [P_w(I|F, G, A, H, T) - P_s(I|F, G, A, H, T)] * n_p \tag{2}$$

The FR threshold (θ_{FR}) is maintained in the system through the introduction of the "unknown" (U) state in face and identity nodes. The similarity score of U for FR is set to the θ_{FR}, hence, when normalised, the similarity scores below the threshold have lower probabilities than U. Similarly, those that have higher similarity scores than the threshold will have higher probabilities than U.

3.2 Learning

Our hypothesis is that the recognition could be improved by learning the likelihoods of the system through evidence. Hence, as our contribution, we propose a BN where the likelihoods of the system are learned from data.

A possible solution could be to create a model that depends on time-series data, like a dynamic BN. However, in a dynamic BN, only the immediate prior value at the previous time step is used, which differs from an open-set identification problem, where the previous state can contain values that applied to another user. For example, user "1" might be encountered right before user "2"; in which case, the evidence for user "1" might not have an effect on the identity estimation for user "2".

Therefore, we designed our own approach in learning from data. We initially use the prior knowledge in setting the likelihoods for each variable (e.g. $P(F|I)$, $P(G|I)$). When the robot identifies a user, and the user confirms her identity, the recognition information and the identity of the user are fed as evidence to the network, and the current posteriors are summed and normalised with the previous posteriors, to update the posteriors for this user. In our example (see Fig. 2), initially, $P(F = "1" | I = "1")$ is set to be much greater than the rest of the likelihoods. However, the FR evidence gives a higher probability score for "3" than "1", which might be due to the similarity in their appearance. After the identity confirmation of the user, using the face evidence, and the evidence for the other modalities, the likelihood is updated by summing with the previous posterior and then normalising it. Updating the posteriors would allow the network to learn their similarity, hence, at the next encounter, the probability for mistaking "1" with "3" would be decreased.

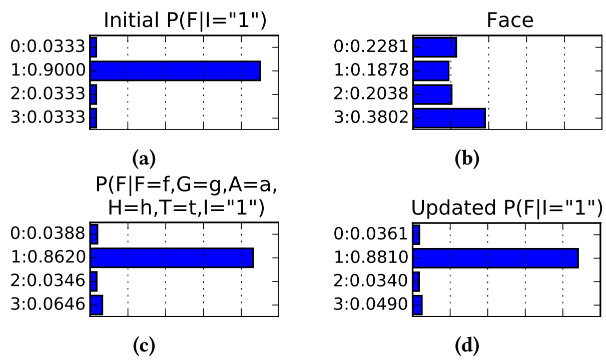


Figure 2: Learning: (a) Initial likelihood of F given I = "1", (b) F evidence, (c) posterior using the evidence, (d) updated posterior.

Likewise, if the recognised user was not previously enrolled into the system, then posterior of $P(F|I = "0")$ is updated. However, gender, age, height, and time posteriors for U are not changed, as they should be uniformly distributed. In order to allow the network to learn from enough data to make meaningful estimations, the output of the system is returned as U if the number of recognitions is less than a predetermined threshold (here, we chose 5).

3.3 Weights

We smooth the recognition results of each modality by using the weights as an exponent to the results for the evidence, due to using product rule as the combination method, as opposed to the sum of logarithms method in [9]. Also, we do not restrict the sum of weights to 1.0, as this could deteriorate results of the primary biometric trait (face), and instead set the weights to the range from 0.0 to 1.0.

We designed a pilot experiment, described in Section 4, for collecting data to optimise the weights, that would minimise the overall recognition error. The weights were optimised for each parameter separately, except for the weight of F, which is always 1.0. The weights that corresponded to the minimum number of incorrect recognitions were combined to get the optimum weights, based on the assumption that each node is conditionally independent.

3.4 Normalisation

A good normalisation method should be insensitive to the outliers and provide a good estimate of the real distribution [8]. For analysing the effects on the performance, we compared such normalisation methods that scale the values to $[0, 1]$ range to be used as probabilities within the BN: min-max, tanh [7], softmax [1], and norm-sum (dividing by the sum of values).

BNs use the *product rule* for combining the results of each node, hence, if a probability of a classifier is zero, it results in an overall zero probability for a class irrespective of the results from other classifiers. In order to overcome this problem we used a small cut-off probability threshold as $p_t = 10^{-6}$.

3.5 Extendability

Our approach relies on FR primarily, but the described system can be extended with other primary biometric traits such as voice and fingerprint, and soft biometric traits, such as the location of the interaction, and ethnicity. It is intended to increase the recognition rate from a single image, and tracking is not applied between images. In order to increase the reliability of the system, multiple images (3 images here) are taken in succession during the pilot study and the results are normalised to estimate the identity of the user, which allows discarding the images without a face detected.

The system does not require heavy-computing, hence, it is suitable for use on commercially available robots. We use a Pepper robot[3] in our study with Naoqi[4] software modules (providing a user’s face ID, gender, age, and height that we used as input modalities), however, the network is applicable to any recognition software.

4 STUDY IN USER IDENTIFICATION

The objective of this study is to gather data for finding the optimal weights for the proposed network for user identification.

4.1 Protocol

The user initially enrolls to the system by entering his/her name, gender, age, and height, and then the robot takes photos of the user. During next encounters, the robot predicts the identity of the user and asks for confirmation.

We ran the study with 14 participants (4 female, 10 male, of age range 24-40) and collected a total of 66 images per user over the four weeks period. The recognition process took approximately 5 seconds: ∼2-3s for user detection, ∼1.5s (0.5s each) for image capture, ∼1s to load the network parameters, ∼0.6s (0.2s each) for recognition from modalities, ∼0.9s (0.3s each) for estimation of the identity using the network. The robot stayed in a fixed position before the interaction, and only when a user was identified, it would become animate to ensure a natural interaction. We aimed to achieve better quality of images by keeping the robot fixed, however, since the robot did not notify participants when taking images, some of the captured images include people looking sideways, smiling, partially covering their faces or moving (see Fig. 1).

We use single-user recognition within the images, that is, only one user is assumed to be present in front of the camera. Hence, the image database was cleaned of images with multiple people or any other user rather than the claimed identity for cross-validation. However, in the future, the position of each user can be considered for multiple people recognition and interaction.

[3] https://www.ald.softbankrobotics.com/en/robots/pepper
[4] http://doc.aldebaran.com/2-5

4.2 Results

In order to validate our system, 5-fold cross-validation is applied with 13 images per user in each bin with a different randomised initial ordering of the users, and the results are averaged. Detection and identification rates (DIR) and false alarm rates (FAR) are reported for the pilot study along with receiver operating characteristics (ROC) curves (see Appendix and Fig. 3), which are the performance measures for the open-set identification problem [12].

The average failure to enroll error (FTE) is 0.214 (0.008), which corresponds to the fraction of images where a face cannot be detected. The identity was not estimated by the network in those cases because the only primary biometric in our system is FR and soft biometrics do not have the deterministic characteristic to estimate the identity on their own.

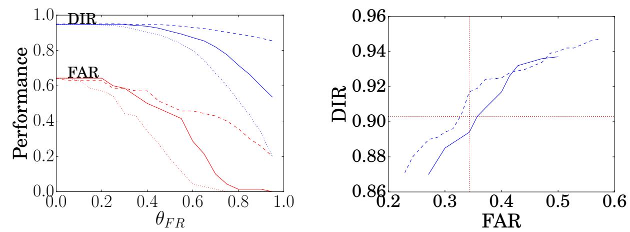
The optimised weights (see Appendix) show that in our study the age is the least effective soft biometric in determining the identity, whereas height is the most effective one. However, this might be due to the characteristics of the population in our pilot study, as the participants' ages are close to each other. Another important factor is the reliability of the age recognition software. The standard deviation of the estimated age of a user on average was 9.3. Hence, we cannot conclude that age should not be used to supplement the FR in general, but if used, the accuracy of the software used should be high, especially in a population with a narrow age range. On the other hand, the effectiveness of the height can also be explained by the nature of the population (3 relatively tall (> 180 cm) and 2 relatively short (< 160 cm) users), even though the average standard deviation of estimated height was 6.3 cm. A more balanced dataset would allow observing the true effects of these parameters.

The cross-validation results for the optimised weights (see Appendix) show that combining soft biometrics using our proposed BN can increase the DIR, depending on the inner settings. It can be observed that although norm-sum and min-max methods provide good results without learning, the recognition rate drops below FR with learning, whereas softmax and tanh methods are not affected.

However, the FAR of the network for any normalisation method is greater than the FAR of FR. This is caused by the combination of multimodal data. For example, if the highest face similarity score is below the threshold, the FR reports the user as "unknown". The network, on the other hand, will still try to identify the user based on other sensor input, where errors might increase FAR.

In order to compare the effects of learning, we chose the min-max method without learning and with a cut-off threshold (N_{minmax}) and the softmax method with learning and no cut-off threshold ($NL_{softmax}$), because the former provides the second highest DIR but with lower FAR than that of the best methods (highlighted in blue), and the latter provides the best DIR in learning in both training and test sets. The results are presented in Fig. 3.

The trade-off between DIR and FAR can be observed in Fig. 3a. The ideal FR threshold (θ_{FR}) should maintain a low FAR with a good DIR. For example, at $\theta_{FR} = 0.7$, the FAR is very low for both FR and $NL_{softmax}$, however, the DIR has also decreased substantially. If we compare the results in the range where $FAR_{FR} \leq 0.5$ ($\theta_{FR} = 0.3$) and $DIR_{FR} \geq 0.8$ ($\theta_{FR} = 0.6$): N_{minmax} is better in identification ($0.93 \leq DIR_{minmax} \leq 0.949$) than $NL_{softmax}$ ($0.873 \leq DIR_{softmax} \leq 0.946$) and FR ($0.801 \leq DIR_{FR} \leq 0.933$). However, $NL_{softmax}$ misidentifies



(a) (b)

Figure 3: ROC curves (Dotted lines represent FR results, dashed line is N_{minmax}, solid line is $NL_{softmax}$): (a) Performance measures, DIR (in blue) and FAR (in red), for varying θ_{FR}; (b) ROC curve for varying Q values for $\theta_{FR} = 0.4$.

the "unknown" users much less ($0.286 \leq FAR_{softmax} \leq 0.543$) than N_{minmax} ($0.457 \leq FAR_{minmax} \leq 0.571$).

$\theta_{FR} = 0.4$ gives the highest detection rate for both N_{minmax} and $NL_{softmax}$ with lower FAR, hence, we compared the effects of quality of estimation (Q) at this rate (see Fig. 3b). The area of improvement for the open-set identification problem is where $FAR \leq FAR_{FR}$ and $DIR \geq DIR_{FR}$. $NL_{softmax}$ does not provide a value in this range, hence, we can conclude that the proposed learning method performs worse than the method without learning. DIR_{minmax} is 1.4% higher than DIR_{FR} where the FAR is equal ($Q = 0.31$), and FAR_{minmax} is 1.4% lower than FAR_{FR} where DIR is equal ($Q = 0.41$). On the other hand, if the problem was treated as a closed-set problem (where all the users are enrolled into the system), $Q = 0$ would be sufficient and the increase in DIR would be 4.4%.

5 CONCLUSION

Our results suggest that the use of soft biometrics increases the recognition rate, however, it can also increase the misidentification rate of unknown users. Increasing θ_{FR} and Q can indeed decrease the FAR, but it can decrease the DIR as well. On the other hand, our proposed learning method mostly performs worse than the traditional BN on this dataset.

Furthermore, the results indicated that our dataset might be biased due to the small population size and the characteristics of the population. To our knowledge, the only publicly available database that contains the soft biometrics used in our system (except the time of interaction) with a face database is the recently released BioSoft [14]. However, the number of subjects is limited to 75, and the height is defined in labels instead of numeric values. Therefore, as a future extension, we will generate an artificial database with a higher amount of subjects with differing soft biometrics, which would also allow setting the noise level in the modalities. We aim to compare the performance of our system with other classification methods such as Support Vector Machines on the artificial dataset.

In the near future, we plan to use the proposed user recognition system in Cardiac Rehabilitation (CR) therapy, during which the robot will recognise the patients and personalise the interaction based on the information about the patients' previous sessions and their progress during the therapy [10]. The study will allow us to evaluate our system in a real-world application, and to observe the effects of personalisation in a long-term HRI.

Appendix: 5-fold cross validation mean (with standard deviation) of false alarm rates (FAR) on the training set, detection and identification rates (DIR) for rank 1 for training and test sets, and optimised weights for each normalisation method with varying learning method and cut-off threshold (p_t) settings with $\theta_{FR} = 0.3$. Highlights in blue show the best values obtained in learning and without learning conditions (minimum FAR, maximum DIR for training and test sets). Highlights in red show the chosen methods for the comparison of learning from data.

Learning	p_t	*Normalisation*	*FAR*	DIR_1 *(Training)*	DIR_1 *(Test)*	w_G	w_A	w_H	w_T
none	none	FR	0.443 (0.078)	0.933 (0.004)	0.945 (0.015)	0	0	0	0
none	none	norm-sum	0.629 (0.032)	0.951 (0.004)	0.967 (0.013)	0	0	0.1	0
none	none	min-max	0.629 (0.032)	0.951 (0.005)	0.965 (0.015)	0.2	0	0.1	0
none	none	softmax	0.571 (0.072)	0.947 (0.004)	0.965 (0.014)	0.1	0	0.6	0
none	none	tanh	0.571 (0.051)	0.942 (0.005)	0.955 (0.012)	0	0	0.1	0
none	1e-6	norm-sum	0.529 (0.081)	0.943 (0.003)	0.956 (0.015)	0	0	0.1	0.1
none	1e-6	min-max	0.586 (0.060)	0.949 (0.005)	0.965 (0.014)	0.2	0	0.1	0
none	1e-6	softmax	0.571 (0.072)	0.946 (0.003)	0.959 (0.015)	0.1	0.1	0.1	0.1
none	1e-6	tanh	0.543 (0.039)	0.942 (0.003)	0.957 (0.013)	0	0	0.3	0.1
evidence	none	norm-sum	0.629 (0.032)	0.782 (0.063)	0.694 (0.093)	0.1	0	0.1	0
evidence	none	min-max	0.629 (0.032)	0.776 (0.064)	0.692 (0.090)	0	0	0.1	0
evidence	none	softmax	0.586 (0.060)	0.946 (0.005)	0.961 (0.017)	0.1	0	0.6	0
evidence	none	tanh	0.571 (0.051)	0.943 (0.007)	0.955 (0.012)	0	0	0.1	0
evidence	1e-6	norm-sum	0.571 (0.072)	0.75 (0.082)	0.632 (0.127)	0.1	0	0.1	0
evidence	1e-6	min-max	0.643 (0.0)	0.776 (0.061)	0.697 (0.089)	0	0	0.1	0
evidence	1e-6	softmax	0.586 (0.060)	0.946 (0.005)	0.954 (0.025)	0.1	0	0.6	0
evidence	1e-6	tanh	0.543 (0.039)	0.943 (0.006)	0.961 (0.019)	0	0	0.3	0

ACKNOWLEDGMENTS

This work has been supported by the EU H2020 Marie Sklodowska-Curie Actions ITN project APRIL (grant 674868). The authors would like to thank Valerio Biscione for his constructive suggestions in the design of the Bayesian network, and the participants of the user study for their time and efforts.

REFERENCES

[1] Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag New York, Inc.

[2] Antitza Dantcheva, Petros Elia, and Arun Ross. 2016. What else does your biometric data reveal? A survey on soft biometrics. *IEEE Transactions on Information Forensics and Security* 11, 3 (2016), 441–467. https://doi.org/10.1109/TIFS.2015.2480381

[3] Christophe Gonzales, Lionel Torti, and Pierre-Henri Wuillemin. 2017. aGrUM: a Graphical Universal Model framework. In *International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems (Proceedings of the 30th International Conference on Industrial Engineering, Other Applications of Applied Intelligent Systems).* https://doi.org/10.1007/978-3-319-60045-1_20

[4] Andy Greenberg. 2017. Watch a 10-year-old's face unlock his mom's iPhone X. (November 2017). Retrieved January 5, 2018 from https://www.wired.com/story/10-year-old-face-id-unlocks-mothers-iphone-x/.

[5] Manuel Günther, Steve Cruz, Ethan M. Rudd, and Terrance E. Boult. 2017. Toward open-set face recognition. (2017). arXiv:1705.01567 Retrieved from http://arxiv.org/abs/1705.01567.

[6] M. Günther, P. Hu, C. Herrmann, C. H. Chan, M. Jiang, S. Yang, A. R. Dhamija, D. Ramanan, J. Beyerer, J. Kittler, M. A. Jazaery, M. I. Nouyed, G. Guo, C. Stankiewicz, and T. E. Boult. 2017. Unconstrained face detection and open-set face recognition Challenge. (2017). arXiv:1708.02337 Retrieved from http://arxiv.org/abs/1708.02337.

[7] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions.* Wiley.

[8] Anil Jain, Karthik Nandakumar, and Arun Ross. 2005. Score normalization in multimodal biometric systems. *Pattern Recognition* 38, 12 (2005), 2270–2285. https://doi.org/10.1016/j.patcog.2005.01.012

[9] Anil K. Jain, Sarat C. Dass, and Karthik Nandakumar. 2004. Soft biometric traits for personal recognition systems. In *International Conference on Biometric Authentication (LNCS).* 731–738. https://doi.org/10.1007/978-3-540-25948-0_99

[10] Juan S. Lara, Jonathan Casas, Andres Aguirre, Marcela Munera, Monica Rincon-Roncancio, Bahar Irfan, Emmanuel Senft, Tony Belpaeme, and Carlos A. Cifuentes. 2017. Human-robot sensor interface for cardiac rehabilitation. In *2017 International Conference on Rehabilitation Robotics (ICORR).* 1013–1018. https://doi.org/10.1109/ICORR.2017.8009382

[11] Eric Martinson, Wallace Lawson, and Greg Trafton. 2013. Identifying people with soft-biometrics at fleet week. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction (HRI '13).* IEEE Press, 49–56.

[12] P. Jonathon Phillips, Patrick Grother, and Ross Micheals. 2011. Evaluation methods in face recognition. In *Handbook of Face Recognition* (2nd ed.), Stan Z. Li and Anil K. Jain (Eds.). Springer Publishing Company, Incorporated, 553–556. https://doi.org/10.1007/978-0-85729-932-1_21

[13] Arun Ross and Anil K. Jain. 2007. Human recognition using biometrics: an overview. *Annales Des Télécommunications* 62, 1 (2007), 11–35. https://doi.org/10.1007/BF03253248

[14] D. Sadhya, P. Pahariya, R. Yadav, A. Rastogi, A. Kumar, L. Sharma, and S. K. Singh. 2017. BioSoft - a multimodal biometric database incorporating soft traits. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA).* 1–6. https://doi.org/10.1109/ISBA.2017.7947693

[15] Walter J. Scheirer, Neeraj Kumar, Karl Ricanek, Peter N. Belhumeur, and Terrance E. Boult. 2011. Fusing with context: A Bayesian approach to combining descriptive attributes. In *2011 International Joint Conference on Biometrics (IJCB).* 1–8. https://doi.org/10.1109/IJCB.2011.6117490

[16] Waldemar Wójcik, Konrad Gromaszek, and Muhtar Junisbekov. 2016. Face recognition: Issues, methods and alternative applications. In *Face Recognition - Semisupervised Classification, Subspace Projection and Evaluation Methods*, S. Ramakrishnan (Ed.). InTech, Chapter 02. https://doi.org/10.5772/61471