R Guest, O Miguel-Hurtado, SV Stevenage, GJ Neil & Sue Black (2014). Biometrics within the superidentity project: a new approach to spanning multiple identity domains. Paper presented at the International Carnahan Conference on Security Technology (ICCST).

Final version available at:

http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6986992&refinements%3D4277039428%26sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A6986962%29

DOI:

10.1109/CCST.2014.6986992

Biometrics within the SuperIdentity project: a new approach to spanning multiple identity domains

Richard Guest, Oscar Miguel-Hurtado
School of Electronic and Digital Art
University of Kent
Canterbury, United Kingdom
{R.M.Guest, O.Miguel-Hurtado-98}
@kent.ac.uk

Sarah V. Stevenage, Greg J. Neil
School of Psychology
University of Southampton
Hampshire, UK
{S.V.Stevenage, gjn1w07}@soton.ac.uk

Sue Black
Centre for Anatomy and Human
Identification, University of Dundee
Dundee, United Kingdom
s.m.black@dundee.ac.uk

***Abstract*— this paper presents a new approach to user identity currently being explored within the SuperIdentity project and outlines the specific role that biometric measurements can contribute towards a modelling process. The SuperIdentity project aims to define a novel holistic model wherein the proven strengths of relationships between facets of identity are mapped. These facets are drawn from both real and digital worlds and are categorised into four domains: biographic, psychological, biometrics and cyber-behaviour. Examples from a sample database explicitly collected to accomplish the aims of the project are also presented.**

***Keywords*— Biometrics; Identity management; Biometric database; Privacy; Security.**

I. Introduction

Identity is a complex concept which comprises multiple elements and can be expressed in different ways depending on the context. These different contexts in the physical-world may relate to work, friends, family, hobbies, etc. Similarly, there are different contexts in the digital world: through work-related environment such as LinkedIn, and predominately social environment like Facebook and Twitter or social media platforms attached to games and interests.

In [1] the authors suggest that identity is revealed through many different fragments which, when combined, approach a complete sense of who someone might be. This work also noted that a person may adopt different roles across different contexts, and in each role they may reveal different aspects of their personality, shaped both by our backgrounds and by the norms of each situation or context. Taken together, all of this information can be combined into a single “SuperIdentity”.

Taking this view of identity, the SuperIdentity (SID) project [2] was conceived as a novel and ambitious approach which aims to describe a holistic model of identity, creating a framework through which to gather the numerous identity cues and their relationships. These elements are drawn from across both physical and digital worlds and are categorised into four *identity domains*: biographic, psychological, biometrics and cyber-behaviour.

Within a domain, individual elements are often related. For instance in the biographic domain, gender and height are related to each other. Considering that the individual elements within each domains are often related, the novelty of the SID project lies in revealing these links, with a particular interest in revealing links that cross from the physical to the digital domains. Figure 1 represents the four domains and their potential relationships. Relationships are derived both from previously published work and from new empirical studies, with the new studies focusing on both algorithmic computer-based assessment and psychological analysis of human recognition performance. Whilst some relationships may be well understood (for example certain biometric features mapped to biographical traits such as gender and age) others are novel (e.g. dynamic signature feature to personality traits). Full understanding of how the inter-domain relationships will enable a representation of the SuperIdentity that can be applied to a variety of end-use scenarios.

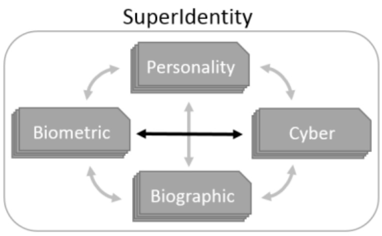

Fig. 1. SID Identity Domains

The SID project brings together a broad interdisciplinary team which includes anatomical experts from the University of Dundee, cognitive psychologists and IT legal experts from the University of Southampton, cyber-psychology experts from the University of Leicester, human computer interaction experts from the University of Bath, cybersecurity experts from Oxford University, visual analytic experts from the Pacific Northwest National Laboratory and biometric experts from the University of Kent.

The SID framework has been developed containing multiple elements within each of the 4 domains, and often further containing numerous features for individual elements. For example, the voice as an element may be described by features such as fundamental frequency, loudness, etc. These individual elements are linked through the relationships obtained both theoretically and empirically. These relationships can be directly implemented in the framework, such as with direct translations and simple or conditional correlations, or be coded as abstract classes for more complex transformations. The framework also contains confidence propagators. The confidence propagators enable the SID framework not only to know which unknown elements can be inferred from known elements, but also provide the level of certainty of those inferences. This is achieved by propagating the relationship level of certainty through the sequence of links (relationships) which connect the known elements with the unknown element(s). Over this framework, advanced visualization techniques are used to create usable interfaces. These interfaces have been developed by a survey of use-cases provided by UK and US analysts who deal daily with identity decisions.

The SID model will provide two capabilities that are unique– "investigative mode" and "defensive mode". *Investigative mode* explores the associations that can be made between different identity elements. The value of these associations is that known pieces of information may then be used to predict other previously unknown pieces of information (e.g. if you know someone's gender and shoe size, you can predict their likely height). The SID model also offers the capacity to quantify the certainty associated with such predictions. This enables the end-user to have a level of confidence (or risk) in their decision, and to make a judgment as to whether additional information is required before sufficient confidence is reached. This approach has not been used previously in the realm of identity, and offers significant value to security and intelligence services.

Defensive mode has a different emphasis. This mode will help citizens, organizations and policy-makers to understand, and measure the privacy risks to which to the leak of information can lead. This capability can help policy-makers to perform privacy threat assessments. Privacy attacks can come from obtaining identity elements that have been inadvertently shared by citizens, disclosed by third-parties or malevolently intercepted. The SID framework will allow policy-makers to measure the extent to which privacy can be improved by safeguarding specific identity elements, making it possible to quantitatively compare the effects of particular policies. Also, it also can help to make citizens more aware of the need to protect personal information. By interacting with the SID framework, citizens will be able to observe personally the implications and risks of disclosing their personal identity information.

In contrast to the common use of biometrics as a means of authentication or identification, one of the challenges of the SID project is to use biometric data to establish links with the other three domains, with an emphasis on the links between the biometric and cyber-behaviour domains (Fig. 1). Considering the difficulty in identifying direct links which connect cyber-behaviour with biometrics, our hypothesis is that biographic or personality traits may act as intermediary measures.

In this paper we review our methodology for assessing features and relationships, and the cross-domain experimentation currently being undertaken. In order to illustrate our approach, we discuss the examples of signatures, hand geometry and keystrokes, and their links with other domains. For each element of identity, a statistical methodology has been applied to explore biometric links to biographic and personality traits. Furthermore a comparison has also been performed between machine and human judgements through the use of a number of the individual elements. In Section II of this paper we discuss examples from the database collected to achieve the SID project's aims. Section III will explain the methodology used to find links between biometrics features and personality/biographic traits, giving preliminary results for several biometrics modalities. An outline of the work undertaken on the comparison of human vs machine-based authentication systems and how they can be complimented will be detailed in Section IV. The summary in Section V will present the conclusion and future work planned.

II. SSD - The SuperIdentity Stimulus Database

One of the major challenges when trying to find links between domains is the availability of data. There are many high quality multi-modal datasets to draw on in the literature, such as BIOSECURE [3] and MyIDEA [4]. However, each of these datasets tends to focus purely on biometrics. Although some of the datasets provide biographic information, none of them provide personality information. As we noted above, links between the biometric domain and other domains are likely to run through biographic and personality data. Consequently, the lack of such information in existing datasets makes it difficult to examine most links between domains.

In order to solve this problem, and to explore the relationships across multiple domains, a novel sample dataset has been collected: the SuperIdentity Stimulus Database (SSD). Full details of the SSD will be released in the future, but here we discuss the parts of the dataset that are relevant to our examples of signatures, hand geometry and keystrokes. For a summary of all data collected, see Table 1.

TABLE 1 SUPERIDENTITY STIMULUS DATABASE SUMMARY

Domain	**Trait**	
Biographic	Age	Weight
	Gender	Shoe Size
	Ethnic	Hobbies
	Handedness	Places of residence
	Height	Occupations
Biometric	Face	Palm vein
	Fingerprint	Gait
	Iris	Voice
	Signature	Hand geometry
Personality	Self-monitoring	Impulsivity
	Social desirability	Situational self-awareness
	Sex Role	Five Factor
Cyber	Keystroke	Swipe
	Facebook & LinkedIn profile page	
	Facebook & LinkedIn privacy settings	
	Facebook, Goggle+ & Flickr number of friends	
	Twitter number of followers and followed users	

The SSD comprises 116 participants (57 males and 59 females). To limit the variability on the metrics captured the data set population was restricted to Caucasians, within an age range between 18 and 35 years and with English as their first language. 11 of the participants were left handed. Each participant contributed a broad range of identity cues from the four different domains outlined in Section I.

A. Biographic Data

The biographic data was self-reported by participants using an online questionnaire. In this questionnaire, basic biographic data such as age, sex classification, ethnic origin, handedness, height, weight, and shoe size were collected. The questionnaire also includes more personal information such hobbies, previous places of residence and occupation.

B. Biometric Data Examples

A broad range of biometric modalities were captured in the SSD. Here, we discuss signatures and hands data:

- **Dynamic signature:** Each participant provided a total of nine samples collected with a LCD signature tablet. Dynamic constructional data of the signatures were recorded, so that analysis of the signatures could look at measures from the real-time construction of the signature and from the static finished product. Thus, pen x and y positions were recorded together with pressure and tilt.
- **Hand geometry:** Using bespoke equipment a camera was fixed in position facing a flat surface. Upon the surface was laid a guide as to where participants should place their hands, along with two mirrors set at 60° angles. Three photographs were taken of the back of each hand, with the mirrors enabling the photograph to include a view of either side of the hands.

C. Personality Data Examples

The SSD provides a wide variety of personal scales that have been measured through the use of the most common standardised personality inventories. Online questionnaires were self-reported by participants. Several scales were collected, discussed here are the scales that are relevant to our analysis examples in the following section:

- **Five-factor personality inventory** [5]**:** these five scales measure five broad dimensions of personality: neuroticism, extraversion, openness, agreeableness and conscientiousness.
- **UPPS impulsive behaviour scale** [6]**:** these four scales measure four dimensions of impulse behaviours: premeditation (lack of), urgency, sensation seeking and perseverance (lack of).
- **Situational self-awareness** [7]**:** these three scales measure how aware a person is in three different situations: private, public and situational.

D. Cyber Data

"Cyber data" in the SSD covers two broad aspects. The first concerns how we physically interact with devices that allow us to access the digital world. One example that is relevant to our data analysis here is keystrokes. Keystroke information was collected by asking participants to type a variety of phrases into a keyboard. Whilst they typed, the identity and timing of each key strike were recorded. The second aspect is how our attitudes and behaviours are expressed in the digital world. Although our analysis below does not involve the use of this second type of measure, examples include Facebook profile pages and privacy settings.

III. Biometrics and Personality

As previously mentioned, within the SID project, biometrics are not intended as a mean for authentication, but rather as link to the other identity attributes including the personality of the person providing the biometric sample. In this way, using the SID model, our hypothesis is that if it is possible to predict personality traits from biometric samples, it may be also possible to predict other identity cues such as cyber-behaviours from these personality traits. In order to reveal the potential links between biometric and personality traits, and as well some biographic traits such as height, weight, gender, the following methodology has been implemented.

A. Biometric feature extraction

For each of the different biometric modalities included in the SSD, a review of the state-of-the-art has been undertaken. The aim of this review is to find the most 'powerful' set of biometric features (in terms of inter-subject discrimination), related to each specific modality. Having extracted these features, related to each specific modality. These features are then extracted. Where multiple single feature samples exist for an individual, the median value of each feature is calculated, thereby reducing the risk of data being skewed by outlying values.

B. Correlation analysis

After calculating the features for each participant, an analysis is performed to find whether a particular personality or biographic trait is significantly correlated with any of the biometric features. For continuous features, the Pearson's correlation coefficient is calculated. For categorical values, an independent-samples t-test is performed. The p-values from these tests indicate significant relationships.

From this analysis, as a visualization tool, a correlation matrix is obtained depicting which personality/biographic traits correlate with which biometric modality features. This information can also be used to further analyse specific types of features. As an example, Figure 3 shows the dynamic and static signature features (represented in x axis by ID numbers 1 to 135) correlations with personality/biographic traits (represented in y axis by ID Numbers 136 to 156). It can be seen that a larger number of dynamic signature features significantly correlate to the personality/biographic traits. Black markers indicate significant correlations between biographic traits and signature features whilst white markers indicate significant correlations between personality traits and signature features.

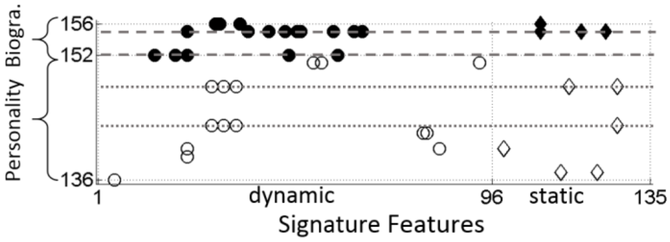

Fig. 2. Example of correlation matrix where it can be seen the significant correlations between signature features and personality/biographic traits

C. *Multi-linear regression models*

The biometric features which have shown a significant correlation with a specific personality/biographic trait are used to create a multi-linear regression model. In that model, the personality/biographic trait is used as a response whilst the biometric features are used as predictors. To ensure that a resultant model only contains the most relevant biometric features a stepwise feature selection approach is used.

This methodology can also use biometric features as the response of the multi-linear model and the personality/biographic traits as predictors. In this way a model can be used to analyse whether personality traits can predict biometric characteristics. Furthermore this methodology could also be applied to find links between biometric modalities, in order to find whether biometric features from one modality can predict biometric features from other modalities.

D. *Preliminary results*

The methodology described in Sections A to C has been applied to several biometrics modalities such as dynamic signature, hand geometry and keystrokes. In this Section we outline the preliminary results obtained from these modalities.

Dynamic Signature vs Personality & Biographics

For dynamic signature, 135 features were selected from the state-of-the-art review, emphasising the dynamic features (encapsulating the temporal process by which a signature was created). Dynamic features may be more advantageous because they are related to the way an individual moves, and the way an individual moves can be influenced by their personality. Thus, by focusing on both dynamic and static (pictorially related) features, we have the best chance to find any links to personality and thus to other domains.

Our analysis shows interesting relationships between both dynamic and static signature features and both personality and biographic traits. Yet, the greater number of significant correlations found were for biographic traits: sex classification and weight (see Figure 2).

The significant correlations found in the previous step were entered into multi-linear models using stepwise feature selection approach, which found that sex classification and weight could be predicted by signature features. Regarding personality traits, multi-linear model were created for contentiousness and UPPS premeditation, which enable to predict them by combinations of certain dynamic and static signature features. These results suggest there is renewed value in using a signature to reveal information about the signer.

Hand Geometry vs Personality & Biographics

From the hand images collected at the SSD, 21 hand lengths were extracted as depicted in Figure 3:

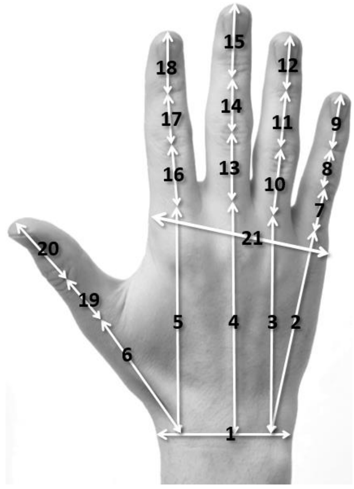

Fig. 3. Hand geometry features

From these lengths, several ratios including finger length, hand dimension and finger segment were calculated. Some of these ratios have been used in the literature as predictors for different personality [8]–[10] and biographic traits [11], [12].

The correlation analysis performed shows, as expected, a large number of correlations with biographic traits as well as correlations with personality traits. Using these correlations, multi-linear regression models have been derived for sex classification, foot size and height, all of which may be of interest within the forensic field in enabling a fuller physical description of a suspect from only a limit set of known information.

Regarding personality traits, as in previous studies [8]–[10], multilinear regressions model showed low prediction capability. However, some of the models found showed higher prediction capability than in previous studies, as for UPPS premeditation, perseverance and urgency, SSA private and five factors agreeableness and neuroticism. These new findings suggest that the use of greater variety of hand lengths rather than the second to four length ratio commonly used in the psychologist literature, could lead to a better prediction of personality traits.

Keystroke vs Personality & Biographics

As with signature production the features extracted from the keystroke data once more emphasise the dynamic aspects of typing, and how they could link with the participant's personality.

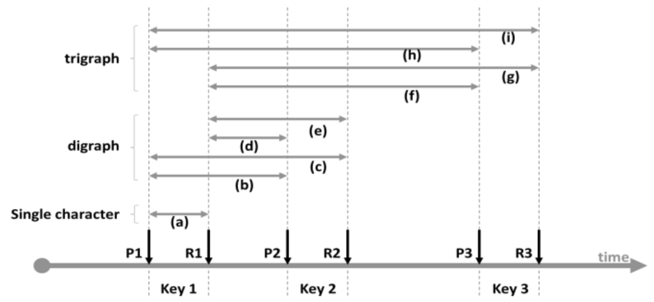

Fig. 4. Keystrokes pres and release times and features extracted from a sequence of three keystrokes

Figure 4 shows the key press (P) and key release (R) events over time for three consecutive keystrokes. Using this temporal information several features can be extracted for one single keystroke (i.e. a = R1 - P1) and the combinations of two (digraphs) or three (trigaphs) consecutive keystrokes. The median of the features has been calculated for each participant, as well as the mean and standard deviation.

Very few features were correlated with personality and demographic traits, leading to a small number of variables being entered into the multi-linear regressions models. What predictors have been found were rather weak in strength. Further analysis are being undertaken in order to extend these results.

Hand geometry vs Signature

As a final example of the biometric work that is being conducted within the SID project, the analysis of links between hand geometry and signature are being examined. The main aim of this work is to determine whether the shape and proportions of the hand might correspond with signature characteristics.

Using the aforementioned features for hand geometry and signature, the correlation analysis have shown some significance values. Different regression models have been found that hand geometry ratios, such as wrist width to hand breath and little to ring finger length, predict both dynamic and static signature features.

IV. Biometrics: Human Vs. Machine

The SID project has also been investigating the capabilities of human vs machine in matching biometric samples from the same individual with each other. Such sample matching underlies the process of authentication, which can be used in automatic biometric systems such as iris controlled access. By understanding the processes of how a human performs authentication, including which visual cues human take into account and the types of errors they make, automatic biometric systems can be improved [13]. Improved authentication performance can be achieved by either incorporating the human expertise to the automatic biometric algorithm or by combining both human perceiver and machine algorithm elements.

As an example of this work, the University of Kent along with the University of Southampton conducted an analysis of an iris recognition system. Previous work established that human perceivers can detect similarities between the left and right irises of an individual [14] and between the irises of identical twins [15] based on the similarity of the iris texture. These similarities are not found when an automated biometric algorithm is used, as left and right irises from the same person are often considered as belonging to different people. These results indicate that the biometric algorithms are missing information that the human brain can process. Our work presented in [13], explored the performance level achieved by both human-based and computer-based authentication system both assessing the same iris samples. By analysing the type of errors obtained, the aim of this work was to fuse both human and computer-based systems through a ‘committee-voting’ model, in which disagreements between human and machine are weighted by the human input. Our fusion system refines the decision of the computer-based system with the inclusion of this second, human decision-maker. Using this system, the results obtained showed that significant improvement is possible through the combination of human and computer-based system.

V. Conclusion and Future Work

In this paper the aims and objectives of the SID project have been presented alongside the important role that biometric data plays in cross-domain identity assessment. The SID project has set out to create a novel holistic model of identity, taking into account identity cues from four different domains: biographic, biometric, personality and cyber-behaviour. Multiple measures have been analysed and their relationships have been investigated, creating a detailed model of identity, where known information may be used to predict unknown information. Importantly, the SID model will be able to predict the confidence of these inferences. To serve the creation of this model we have had to collect our own data, in the form of the SSD, a unique multimodal database which combines multiple biometric modalities with biographic, personality and cyber-behaviour data.

The assessment of biometric features within the SID project has been described with specific reference to their predictive value when considering other aspects of identity. Of most interest and value may be the links that are possible across domains, describing how biometric measures may be used to predict psychological traits or cyber-behaviours. To discover new links between the physical and the digital domains, the most promising route uses strong links to personality traits as a bridge. The methodology chosen to find these associations has been detailed, and examples of preliminary results have been described. Perhaps most promising, though is the demonstration of clear benefit when a fusion model is adopted, as human and machine performance in combination can significantly improve the outcomes of the system.

The SID project is a highly complex and multidisciplinary project, representing a unique approach to describe identity. SID is on-going with further experimentation planned on other biometric modalities along with the exploration of new biometric to biometric relationships. Furthermore, is being considered the employ of more advanced machine-learning techniques, to augment our understanding of how one cue to identity relates to, and predicts, another.

ACKNOWLEDGMENT

This work is supported by UK EPSRC Grant EP/J004995/1 (SID: An Exploration of SuperIdentity). Colleagues on this grant are thanked for helpful comments on the work reported here.

REFERENCES

[1] K. Rannenberg, D. Royer, and A. Deuker, Eds., *"The Future of Identity in the Information Society"*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[2] University of Southampton, University of Kent, University of Leiceester, University of Dundee, University of Bath, University of Oxford, Pacific Northwest National Laboratories, "the SuperIdentity Project," 2013. [Online]. Available:

http://www.southampton.ac.uk/superidentity/.

[3] J. Ortega-Garcia, J. Fierrez, F. Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. W. R. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, and A. Savran, *"The multiscenario multienvironment BioSecure Multimodal Database (BMDB)."*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 6, pp. 1097–111, Jun. 2010.

[4] B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacrétaz, A. Humm, F. Evéquoz, R. Ingold, and D. Von Rotz, *"MyIDea-multimodal biometrics database, description of acquisition protocols"*, Proc. Third COST, vol. 275, pp. 59–62, 2005.

[5] T. Buchanan, J. A. Johnson, and L. R. Goldberg, *"Implementing a five-factor personality inventory for use on the internet"*, Eur. J. Psychol. Assess., vol. 21, no. 2, pp. 115–127, 2005.

[6] S. P. Whiteside and D. R. Lynam, "*The five factor model and impulsivity: Using a structural model of personality to understand impulsivity"*, Pers. Individ. Dif., vol. 30, no. 4, pp. 669–689, 2001.

[7] J. M. Govern and L. A. Marsch, *"Development and validation of the situational self-awareness scale"*, Conscious. Cogn., vol. 10, no. 3, pp. 366–378, 2001.

[8] R. A. Lippa, *"Finger lengths, 2D:4D ratios, and their relation to gender-related personality traits and the Big Five"*, Biol. Psychol., vol. 71, no. 1, pp. 116–121, 2006.

[9] J. Loehlin, S. Medland, and N. Martin, *"Relative Finger Lengths, Sex Differences, and Psychological Traits"*, Arch. Sex. Behav., vol. 38, no. 2, pp. 298–305, 2009.

[10] B. Fink, J. T. Manning, and N. Neave, *"Second to fourth digit ratio and the 'big five' personality factors"*, Pers. Individ. Dif., vol. 37, no. 3, pp. 495–503, Aug. 2004.

[11] J. H. Musgrave and N. K. Harneja, *"The estimation of adult stature from metacarpal bone length"*, Am. J. Phys. Anthropol., vol. 48, no. 1, pp. 113–9, Jan. 1978.

[12] T. Kanchan and P. Rastogi, *"Sex determination from hand dimensions of North and South Indians"*, J. Forensic Sci., vol. 54, no. 3, pp. 546–50, May 2009.

[13] R. M. Guest, H. He, S. V Stevenage, and G. J. Neil, *"An assessment of the human performance of iris identification"*, 2013 IEEE Int. Conf. Technol. Homel. Secur., pp. 623–626, Nov. 2013.

[14] A. Sgroi, K. W. Bowyer, and P. J. Flynn, *"The prediction of old and young subjects from iris texture"*, 2013 Int. Conf. Biometrics, pp. 1–5, Jun. 2013.

[15] K. Hollingsworth, K. W. Bowyer, S. Lagree, S. P. Fenker, and P. J. Flynn, *"Genetically identical irises have texture similarity that is not detected by iris biometrics"*, Comput. Vis. Image Underst., vol. 115, no. 11, pp. 1493–1502, Nov. 2011.