ANTI-FORENSIC RESISTANT LIKELIHOOD RATIO COMPUTATION: A CASE STUDY USING FINGERPRINT BIOMETRICS

Norman Poh, Nik Suki, Aamo Iorliam and Anthony TS Ho

Department of Computing, University of Surrey
Guildford, GU2 7XH, Surrey, UK

ABSTRACT

One of the major utilities of biometrics in the context of crime scene investigation is to identify people. However, in the most sophisticated cases, criminals may introduce the biometric samples of innocent individuals in order to evade their own identities as well as to incriminate the innocent individuals. To date, even a minute suspect of an *anti-forensic* threat can potentially jeopardize any forensic investigation to the point that a potentially vital piece of evidence suddenly becomes powerless in the court of law. In order to remedy this situation, we propose an anti-forensic resistant likelihood ratio computation that renders the *strength of evidence* to a level that is proportional to the trustworthiness of the trace, such that a highly credible evidence will bear its full strength of evidence whilst a highly suspicious trace can have its strength of evidence reduced to naught. Using simulation as well as a spoof fingerprint database, we show that the existing likelihood ratio computation is extremely vulnerable to an anti-forensic threat whereas our proposed computation is robust to it, thereby striking the balance between the utility and threat of a trace.

***Index Terms*—** Anti-forensic, likelihood ratios, tampered images, trustworthiness, strength of evidence

1. INTRODUCTION

1.1. Motivation

The trustworthiness of biometric evidence is a major concern when presenting such evidence in court. As a forensic practitioner, the assessment of likelihood ratio (LR) is the most genuine means to evaluate biometric trace [1]. Anti-forensic techniques are recently so rampant to alter biometric trace that this becomes a great concern to the forensic and biometric communities. Digital tampering and sensor tampering are the two main ways used by anti-forensic practitioners. Even though the LR approach statistically estimates the value of a biometric sample, performing a LR on a digitally tampered or spoofed biometric sample will pose a threat on innocent people.

In order to reduce the threat posed by the aforementioned anti-forensic issues, we are motivated to develop a framework that determines the strength of the evidence and also show if such a biometric sample is worth presenting as evidence in court or not.

1.2. Strength of Evidence

Meuwly and Veldhuis [2] proposed that forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements and statistical models in order to overcome the inconclusive category of evidence outcome. The calculated results may later be supported by an expert's opinion if needed. This is to ensure that one can use and evaluate all possible evidence collected and present it in court. A large body of literature [3–9] asserts that likelihood ratio computation is a reliable, scientifically validated, and approved method for evaluating forensic biometric trace.

Validity and reliability are two virtues that cannot be overly emphasized in forensic science. Morrison *et al.* [10] addressed that one should be able to present the accuracy of their output results together with how precise their approach could be. It will be misleading to court if forensic practitioners only report the latter result without giving a statement concerning the reliability of their result. At the end, it is the judge who has the ultimate power to admit or reject the presented evidence based on investigator's testimony [11].

The main concern in this paper is that if the trace collected from a crime scene is not original or has been tampered with *in any way*, evaluating the calculated result from a trusted likelihood ratio framework could be jeopardised. By taking into consideration the importance of knowing the trustworthiness of the evidence, we propose an anti-forensic methodology that can take into account the trustworthiness of piece of biometric trace. The context of *trustworthiness* in this paper, we refers to the ability to place trust on the collected trace and being assured that the trust shall not be betrayed. Latter we implement it in the usual likelihood ratio computation framework, thereby striking the balance between the utility and threat of a trace proportionately to its level of trustworthiness.

1.3. Questionable Images as Evidence

In general, there are two types of tampered images: (i) digital tampering; and (ii) spoofed samples. Taking into consideration of a biometric sample, digital tampering refers to a process where a trace image is maliciously tampered with in order to be presented as evidence in the court or to tarnish someones

All authors are with Department of Computing, University of Surrey, Guildford, GU2 7XH, Surrey, UK. E-mails: {n.poh, n.niksuki, a.iorliam, a.ho}@surrey.ac.uk

prestige [12]. Fortunately, very often when a digital image is tampered with, the act of tampering leaves traces which can be analysed by a forensic expert [12]. Even though Stamm *et al.* [12] showed that such traces could be cancelled using a suitable dithering noise signal, Giuseppe *et al.* [13] recently argued and showed that the dither noise presents a visible distortion in the attacked image. On the other hand, spoofing refers to any attempt to masquerade as someone else by falsifying the victim's data e.g. wearing a 3D mask, using a printed image, and using artificial materials to replicate a fingerprint to gain illegitimate access or benefits. Indeed, some literature [14–16] suggests that current biometric recognition system are vulnerable to spoofing attacks, thus demonstrating the feasibility of spoofed samples being introduced into a crime scene. Later in this paper, we investigate such a scenario in an experimental setting.

1.4. Our Contributions

Our contribution in this paper is two-fold:

- Proposal of a likelihood ratio framework for anti-forensic resistant, and
- Empirical validation of the framework using a fingerprint database under spoofing attack.

1.5. Paper Organisation

This paper is organised as follows: Section 2 gives our methodology where we explain our proposed anti-forensic resistant computation. Section 3 shows our experiments, database used and results. Finally conclusions and future works are drawn in Section 4.

2. METHODOLOGY

The usual likelihood ratio framework as used in the court of law is:

$$LLR(E) = \log \frac{p(E|H_0)}{p(E|H_1)}$$

in the logarithmic domain, where E is a piece of evidence and H_0 is the null hypothesis and H_1 the alternative hypothesis. The null hypothesis states that the evidence belong to the suspect whereas the alternative hypothesis is that the evidence belongs to someone else. Put differently, H_0 represents the prosecution hypothesis whereas H_1 represents the defence hypothesis.

However, the log-likelihood term, $LLR(E)$ has no mechanism that considers the trustworthiness of the evidence. If the evidence has been tampered with, than, we would like the $LLR(E)$ term to reduce to zero so that it does not carry any strength any more. On the other hand, if the $LLR(E)$ is trustworthy, we shall keep $LLR(E)$ to assume its original value. In other words, we need a trustworthy log-likelihood ratio, just in case our evidence has been tampered with by an anti-forensic technique. Intuitively, the anti-forensic resistance LLR should be of the form of:

$$LLR^{tamper}_{resist}(E) \approx w \ \log \frac{p(E|H_0)}{p(E|H_1)} \tag{1}$$

where $w \in [0, 1]$.

We describe a procedure that can achieve this.

Let $P(T = 1|t)$ be the probability of the trustworthiness T of the sample given the measurement of potential tampering, t which is usually a feature vector. In the case of a digital image, t is deviation from the Benford's law, which is a probability distribution that has been widely used to detect tampering in financial data [17]; and has recently been used to detect digital tampering of fingerprint images [18]. The feature vector t can also characterise the liveness of a biometric sample, e.g., local binary patterns (LBP) [19], so that $P(T = 1|t)$ can be interpreted as the probability of a sample being taken from a live finger rather than spoof materials. In any case, we recognise that the probability of trustworthiness, $P(T = 1|t)$, is a function of the tampering measure, t; hence, this can be written as $w(t) \in [0, 1]$.

A tamper-resistant likelihood can then be defined as:

$$p(E|H_k, T = 1, t) = \frac{1}{Z_k(w(t))} p(E|H_k)^{w(t)} \tag{2}$$

for both prosecution and defence hypotheses, $k \in \{0, 1\}$, where $Z_k(w(t))$ ensures that the left hand side term is properly normalized, i.e., $Z_k(w(t)) = \int_E p(E|H_k)^{w(t)}$, noting that this term is a function of the probability of trustworthiness term, $w(t)$, but not of the evidence, E, itself since it is integrated out of the equation.

Using the likelihood ratio framework, the tamper-resistant LLR can consequently be written as:

$$\begin{aligned}
LLR^{tamper}_{resist}(E) &= \log \frac{p(E|H_0, T = 1, t)}{p(E|H_1, T = 1, t)} \\
&= \log \left(\frac{(Z_0(w(t)))^{-1} p(E|H_0)^{w(t)}}{(Z_1(w(t)))^{-1} p(E|H_1)^{w(t)}} \right) \qquad (3) \\
&= \underbrace{w(t) \log \left(\frac{p(E|H_0)}{p(E|H_1)} \right)}_{\text{evidence-dependent term}} - \underbrace{\log \frac{Z_0(w(t))}{Z_1(w(t))}}_{\text{normalizing term}} \\
&= w(t) LLR(E) + \epsilon(w(t)) \\
&\approx w(t) LLR(E) \qquad (4)
\end{aligned}$$

where in (3), we have plugged in the tamper-resistant likelihood term as defined in (2). By rewriting the equation as a function of the conventional $LLR(E)$, we observe that the tamper-resistant LLR is made up of two terms, i.e., an evidence-based log-ratio term and a normalizing log-ratio term that is independent of the evidence. While the first term rescale the conventional $LLR(E)$ by $w(t)$, the second term, $\epsilon(w(t)) \equiv -\log \frac{Z_0(w(t))}{Z_1(w(t))}$ introduces the bias required in order to achieve equality. In practice, the second term tends to cancel out each other, causing it to assume a significantly small value that is close to zero, i.e., $\epsilon(w(t)) \approx 0$. Put differently, the absolute value evidence-dependent term is many orders larger than that of the normalizing term:

$$\left| \log \left(\frac{p(E|H_0)}{p(E|H_1)} \right)^{w(t)} \right| \gg \left| \log \frac{Z_0(w(t))}{Z_1(w(t))} \right| .$$

Following this rationale, by dropping $\epsilon(w(t))$, we obtain the intuition as specified by (1). Despite being a less important term, it should not be neglected because the absolute value of LLR is often used to interpret the *strength* of a piece of evidence. In summary, although we started with an intuition, by using probability axioms, we have derived a tamper-resistant LLR that is a shifted and scaled version of the conventional LLR, thus introducing minimal modification to a widely accepted practice. Since the modification is easy to understand, it is more likely to be accepted and adopted.

2.1. Assessing the Probability of Trustworthiness using the Bayes Theorem

The key to obtaining a resistant-tampering likelihood ratio is to evaluate $w(t) = P(T = 1|t)$. This term can be calculated using the Bayes theorem [20] or through a discriminant function such as logistic regression that gives probabilistic output. Here, we shall present the first method; and the readers are referred to [20] for the second method.

From the Bayes theorem, the probability of Trustworthiness is given by:

$$P(T = 1|t) = \frac{P(T)p(t|T = 1)}{\sum_{T' \in \{0,1\}} P(T')p(t|T')}$$

where $p(t|T)$ is the likelihood of evidence of tampering given that its state of tampering, T, which can be either true or false. This is because a sample has either been tampered with or it has not. $P(T)$ is the *prior probability* of tampering. Importantly, $p(t|T)$ is obtained from a training database of tampered and untampered samples whereas the prior $P(T)$ is manually set for each and every case work.

The value of $P(T)$ depends on a number of factors. When solving a case work, if there is a reason to be believe that anti-forensic could have taken place, then, it is sensible to set the prior probability of tampering $P(T)$ appropriately. On the other hand, if an evidence is considered 100% trustworthy, one can simply set $P(T) = 1$ so that the tamper-resistant LLR is *exactly* the same as the conventional $LLR(E)$. In summary, $P(T)$ offers a flexible way of specifying trustworthiness that reflects an investigator's belief. This mechanism effectively renders the conventional LLR resistant to antiforensic attempts.

2.2. Analysis of the weighted LLR

Let us now analyse the range of values assumed by a tamper-resistant LLR as in (4). For this purpose, we shall focus on the evidence-dependent log-ratio term and drop the log-ratio normalizing term (which a very small value).

$$LLR^{tamper}_{resist}(E) \approx w(t) \underbrace{\log \frac{p(E|H_0)}{p(E|H_1)}}_{LLR(E)}$$

If the normal $LLR(E)$ is bounded in $[-b, b]$, the tamper-resistant LLR will be bounded in $[-bw(t), bw(t)]$. Since $w(t)$ takes a value between zero and one, the tamper-resistant LLR will be at most as large as $[-b, b]$ but as small as 0. Therefore, $w(t)$ directly controls the strength of evidence.

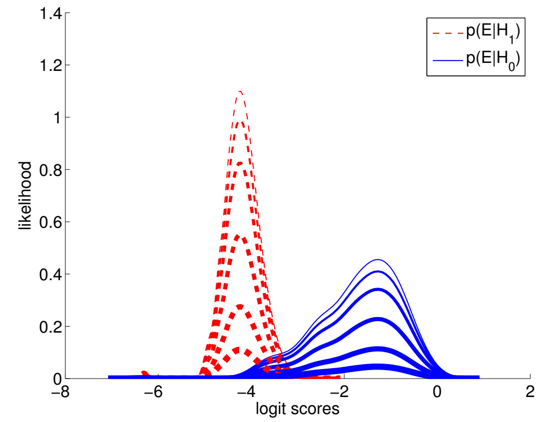


Fig. 1: A demonstration of likelihoods $p(E|H_k, T = 1)^w$ for $k \in \{0, 1\}$ with different weightings of $w \in \{1, 0.9, 0.75, 0.5, 0.25, 0.1\}$ from the thinnest to the thickest lines.

We now illustrate this with an example. We first plot the likelihood of evidence given H_k, $p(E|H_k)$, where the evidence is a matching score, for the prosecution hypothesis H_0 ($k = 0$); as well as the defence hypothesis H_1 ($k = 1$). Because the likelihoods have been modelled on original, non-tampered data, we are really estimating $p(E|H_k, T = 1)$ to write more explicitly. Figure 1 plots the pair of likelihood controlled by w which is set to different values between zero and one. As can be observed, with decreasing w, the likelihoods $\frac{1}{Z_k(w)} p(E|H_k|T = 1)^w$ for both classes, $k \in \{0, 1\}$ also become smaller and smaller. This should lead to smaller strength of evidence in terms of log-likelihood ratio. In the experimental section, we will let w to be controlled by the probability of a live, non-tampered sample.

3. EXPERIMENTS

3.1. Database

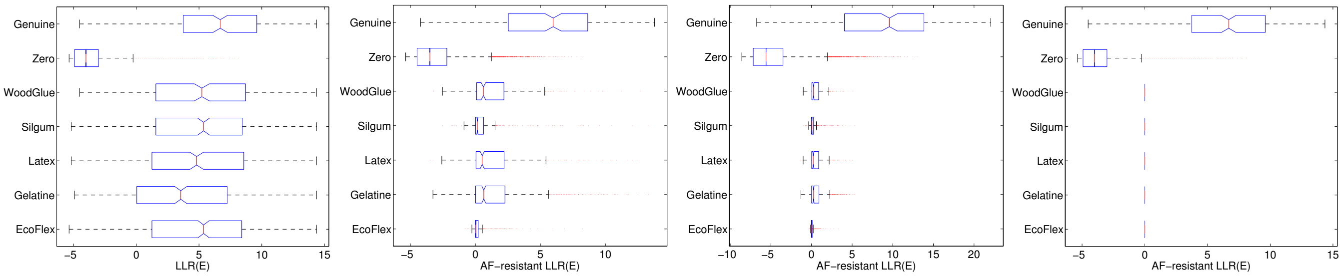
In order to study the potential effect of tampering, we have chosen to use a biometric database containing spoof samples. This enables us quantify objective how strong the proposed anti-forensic resistant framework can withstand tampering. Specifically, we wish to examine whether or not the strength of evidence for tampered biometric samples can be reduced by our proposed method with respect to the conventional method of likelihood ratio computation.

To this end, we have chosen to use the LivDet 2011 [15] database. An interesting aspect of this database is that it contains live fingerprint images of different quality levels as well as fake fingerprint images due to the use of different fabrication materials.

In addition, with the database, we can also measure the strengh of evidence under zero-effort non-match comparison which is essential to estimate $p(E|H_1, T = 0)$.

The LivDET 2011 database contains 8000 samples. The most important key statistics relevant for our experiments are:

- 144 unique fingers containing both live and spoof samples



(a) Conventional LLR (b) Anti-forensic LLR (c) Expert opinion-weighted w (d) Oracle

Fig. 2: A boxplot of the (a) conventional $LLR(E)$, (b) the proposed anti-forensic resistant $LLR(E)$, (c) the expert opinion combined with the inferred posterior probability of untampered images, and with (d) the ground-truth about the nature of tampering.

- 256 unique fingers containing only live samples
- 4000 fingerprints acquired using the Biometrika sensor, and another 4000 acquired using the Italdata sensor.
- 800 fake fingerprint samples for each of the five fabrication materials

We prepared the data by making an exhaustive pair-wise comparison of all the available 8,000 samples. For each of the 8,000 samples, we also estimated their liveness measure based on the Local Binary Patterns (LBP) features as described in [19]. The LBP features have been shown to outperform other competing liveness measures based on pores detection, Curvelet, Power spectrum, Wavelet energy signature evaluated on the LivDet 2011 fingerprint database. We modelled the probability of liveliness given the LBP features using logistic regression [20].

The data set is divided into two equal partitions, namely a training set and a test set. The training set is used for estimating any model parameters. In this case, the models are $p(E|H_k)$ for both $k = \{0, 1\}$, and also $w(t) \equiv P(T = 1|t)$. The test set is used uniquely to assess the properties of $LLR(E)$ and its tamper-resistant $LLR(E)$.

Each of the two sets is generated by a gallery of 72 distinct fingers. This means that the training and test sets are disjoint; hence, completely independent of each other. In each set, there are 720 genuine (match) attempts and 460800 zero-effort impostor attempts; whereas the spoof attempts are further divided into five sets depending on the fabrication materials, namely, Ecoflex, Wood glue, Silgum, Latex, and Gelatine. The number of these spoof attempts are $\{1000, 300, 350, 300, 300\}$, respectively. Each attempt consists of two observations: the biometric matcher output (E) and the liveness measure t of the query sample.

3.2. Results

The first experiment compares the conventional likelihood computation with its anti-forensic resistant version in terms of LLR. The comparative result is visualised using boxplots as shown in Figure 2, which shows the median, the first and third quartiles, as well as the fifth and 95-th percentiles of the data. Figures 2(a) and (b) compare the range of values according to the usual LLR with that of the proposed anti-forensic resistant LLR. We observe that the interquartile range of the conventional LLR, that is the range between the first and third quartile, is very similar to the genuine (match) scores. This implies that the spoof samples due to the five fabrication materials, that are, Ecoflex, Wood glue, Silgum, Latex, and Gelatine would generate LLR that is very similar to that generated by genuine comparison. With antiforensic LLR, Figure 2(b), the dispersion of innocent evidence in LLR still overlaps into the dispersion of a guilty evidence. Although this could still lead to misinterpretation of evidence, the risk of this is significantly reduced, as shown by the boxplot of the genuine sample and those of the spoofed materials, the dispersion of which is comparatively narrower.

In the second set of experiments as shown in Figure 2(c), we created an additional simulation where we allow the investigator to weigh in his/her opinion in order to question the validity of the evidence's trustworthiness. Therefore, rather than relying on a fully automatic liveness detector, the investigator is allowed to introduce a prior of $P(T = 1)$ to some values. In this case, we set $P(T = 1) = 0.8$ for all the live samples and $P(T = 1) = 0.2$ for the spoof samples made with any of the five fabrication materials. We also conducted a third set of experiments which is called the "oracle" where we set $P(T = 1) = 1$ for all the live samples and $P(T = 1) = 0$ for the spoof samples. This enables us to see the best possible that can be possibly achieved. The results are shown in Figure 2(d).

As can be observed, the interquartile range of the spoofed material with the expert's opinion become much more narrower in Figure 2(c). This shows that when the investigator exerts his/her opinion of trustworthiness to the proposed anti-forensic LLR, a much more accurate decision can be attained. Of course, if the opinion of the investigator turns out to be wrong, such an exercise will be counter productive, but can still do no worse than the conventional LLR.

The above experiments demonstrates the ability of the proposed anti-forensic computation to use the expert's opinion in weighting the uncertainty of the evidence's origin as discussed in Section 2.1. Furthermore, with the availability of additional prior knowledge about the integrity of the evidence (as represented by $P(T)$), the resultant strength of evidence is significantly reduced to a non-threatening level.

4. CONCLUSION

In this paper, we proposed an anti-forensic-resistant likelihood ratio computation that explicitly considers the trustworthiness of the evidence. We showed that the strength of evidence can be reduced to a non-threatening level when the evidence has been tampered with. Our empirical investigation shows that the strength of evidence due to spoofing can be significantly reduced in terms of the interquartile range of LLR, leaving the LLR of untampered samples, both for the match and non-match comparisons, to be roughly the same. The method can thus seamlessly be integrated with the widely accepted likelihood ratio computation without significantly modification. The proposed anti-forensic resistant computation also allows the expert's opinion to weigh in his prior belief about the trustworthiness of a piece of evidence by simply setting the prior of $P(T = 1)$. Possible future research directions include: (1) applying the proposed anti-forensic resistant computation to different biometric modalities; (2) investigating the effect of the proposed computation for digital tampered images; and (3) investigating the impact of various settings for $P(T = 1)$.

5. ACKNOWLEDGEMENTS

Dr Poh received funding from the BEAT project (grant no. 284989) to carry out this work. Nik Suki would like to acknowledge Universiti Malaysia Pahang (UMP), Malaysia for the sponsorship on her PhD study. Aamo Iorliam also would like to acknowledge the financial support of Benue State University, Makurdi, Nigeria for the sponsorship of his PhD scholarship. The authors thank Prof Raymond Veldhuis for a fruitful discussion.

REFERENCES

[1] D. Meuwly and R. Veldhuis, "Forensic biometrics: From two communities to one discipline," *Proceedings of the BIOSIG 2012, International Conference of the Biometrics Special Interest Group- (BIOSIG) on*, 2012.

[2] Geoffrey Stewart Morrison and Reinoud D. Stoel, "Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models a response to lennard (2013) fingerprint identification: how far have we come?," *Australian Journal of Forensic Sciences*, vol. 0, no. 0, pp. 1–11, 2013.

[3] A. Alexander, "Forensic automatic speaker recognition using bayesian interpretation and statistical compensation for mismatched condition," 2005.

[4] C.G.G. Aitken and F. Taroni, "Statistics and the evaluation of forensic evidence for forensic scientist," *2nd ed, Wiley, Chichester, UK*, 2004.

[5] Association of Forensic Science Providers, "Standards for the formulation of evaluative forensic science expert opinion," *Sci. Justice on*, vol. 49, no. 0, pp. 161–164, 2009.

[6] G.S. Morrison, "Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs," *J. Acoust. Soc. Americ. on*, vol. 125, pp. 2387 – 2397, 2009.

[7] Joaquin Gonzalez-Rodriguez, Andrzej Drygajlo, Daniel Ramos-Castro, Marta Garcia-Gomar, and Javier Ortega-Garcia, "Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition," *Computer Speech and Language*, vol. 20, no. 23, pp. 331 – 355, 2006.

[8] J. Buckleton, "A framework for interpreting evidence," *in J. Buckleton, C.M. Triggs, S.J. Walsh (Eds.), Forensic DNA Evidence Interpretation,CRC, Boca Raton, FL*, pp. 27 – 63.

[9] D.J. Balding, "Weight-of-evidence for forensic dna profiles," *Wiley, Chichester, UK*, 2005.

[10] J. Epps G.S. Morrison, T. Thiruvaran, "Estimating the precision of the likelihood-ratio output of a forensic-voice-comparison system on," June.

[11] "Daubert v merrell dow pharmaceuticals (92102) 509 us 579," 1993.

[12] M.C. Stamm, S.K. Tjoa, W.S. Lin, and K.J.R. Liu, "Anti-forensics of jpeg compression," *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 1694–1697, March 2010.

[13] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Revealing the traces of jpeg compression anti-forensics," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 2, pp. 335–349, Feb 2013.

[14] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Biometrics (IJCB), 2011 International Joint Conference on*, Oct 2011, pp. 1–7.

[15] Gian Marcialis, Aaron Lewicke, Bozhao Tan, Pietro Coli, Dominic Grimberg, Alberto Congiu, Alessandra Tidu, Fabio Roli, and Stephanie Schuckers, "First international fingerprint liveness detection competition livdet 2009," in *Image Analysis and Processing ICIAP 2009*. Springer Berlin / Heidelberg, 2009.

[16] Jukka Komulainen, Abdenour Hadid, and Matti Pietikinen, "Face spoofing detection using dynamic texture," in *Computer Vision - ACCV 2012 Workshops*, Jong-Il Park and Junmo Kim, Eds., vol. 7728 of *Lecture Notes in Computer Science*, pp. 146–157. Springer Berlin Heidelberg, 2013.

[17] M. Nigrini, "A taxpayer compliance application of benfords law," *Journal of the American Taxation Association on*, vol. 1, pp. 72–91, 1996.

[18] A. Iorliam, A. T. S. Ho, N. Poh, and Y. Q. Shi, "Do biometric images follow the benford's law?," in *the 2nd International Workshop on Biometrics and Forensics (IWBF 2014)*, 2014.

[19] S Nikam and S Agarwal, "Local binary pattern and wavelet-based spoof fingerprint detection," *Int. J. Biometrics*, vol. 1, no. 2, pp. 141–159, Aug. 2008.

[20] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1999.