# Biometric Permanence: Definition and Robust Calculation

John Harvey
Dept. of Systems and Computer Engineering
Carleton University
Ottawa, ON, Canada
Email: jharvey@sce.carleton.ca

John Campbell
Bion Biometrics Ltd.
Ottawa, ON, Canada

Stephen Elliott
and Michael Brockly
Dept. of Technology Leadership and Innovation
Purdue University
West Lafayette, IN, USA

Andy Adler
Dept. of Systems and Computer Engineering
Carleton University
Ottawa, ON, Canada

***Abstract*—In this paper, we develop a novel metric, which we call biometric permanence, to characterize the stability of biometric features. First, we define permanence in terms of the change in false non-match ratio (FNMR) over a repeated sequence of enrolment and verification events for a given population. We consider how such a measure may be experimentally determined. Since changes in FNMR, for most biometric modalities, are small, any variability in the biometric capture over time will camouflage the changes of interest. To address this issue, a robust methodology is proposed which can isolate the visit-to-visit variability, and substantially improve the estimation. We develop a model for the visit biases, and provide extensive simulation results supporting the efficacy of the improved method.**

## I. Introduction

Biometric systems allow identification of people based on analysis of images of their biometric features [8]. When a biometric is used for verification, a biometric sample image is tested against a previously captured sample from the person to be verified [9]. In verification, the performance of the system is measured in terms of its Type-I and Type-II error rates. One key criterion for a biometric modality is the stability of the underlying features. For example, for fingerprint recognition, the structure of the friction ridges is considered to be a unique and stable characteristic of each individual [9].

However, it is widely known [4], [11], [16] that, for any biometric modality, some degree of variation in the biometric features occurs over time. An example is the damage that can occur to fingerprints, which is more common in certain population groups and occupations [2]. Variation in biometric features over time, known as *template aging*, results in a decrease in biometric recognition accuracy over time [7]. The importance of template aging varies across different applications of biometrics. It is especially significant for many government programs, such as border security, in which stored templates are intended to be used for comparison over years or decades. There have been several studies done in assessing or describing the impact of template aging [12], [15]. However, many of these studies have very small datasets (in terms of sample sizes and time periods). Several challenges associated with permanence were identified, including those associated with specific occupations and other environmental factors [3], [13]. Although biometric permanence has seen some investigation, to our knowledge no statistical measures have been defined to measure it or calculate it robustly. To address this deficiency, we define a new term, *biometric permanence*, $P_B$, and develop methods to calculate it. $P_B$ has an inverse significance to template aging: a biometric modality with high $P_B$ shows little change over time.

We first propose a definition for *biometric permanence,* $P_B$, and a reference method to calculate it based on a traditional detection error trade-off (DET) analysis. Next, we consider how to measure $P_B$ robustly for a given biometric modality. A sample population is recruited and biometric measures are performed at intervals over time ($\Delta t$), from which a complete set of cross comparisons is calculated [6]. When calculating $P_B$, the major difficulty in analyzing these data arises in separating the visit-dependent factors from the $\Delta t$ values, which are of course implicitly dependent on the absolute times of the visits. Since the effects of aging can be small, the traditional method is highly sensitive to estimation variance. To address this issue, we propose a strategy to improve the measure, which we call the *matched delta* method.

## II. Methodology

An overview of the matched delta method is presented in Fig. 1. We base the test protocol on that of [6]: in this paper however, we synthesize appropriate match scores as described in Section III. No actual human subject data is involved.

To calculate $P_B$ under this protocol, data are required from a test crew of subjects who are biometrically tested over time at a series of visits. At each visit, $i$, enrolment $E_i$ and verification $V_i$ biometric samples are acquired. When biometric comparisons are made, match scores are calculated and assigned to a bin corresponding to the time difference ($\Delta t$) between visits. Thus, a comparison between $E_i$ and $V_j$ would be in bin $\Delta t_{ij}$. The highest match scores should be those from the same visit – during which no changes due to

template aging occur. Thus, comparisons of $E_i$ to $V_i$ have $\Delta t_{ii} = 0$. Given this set of biometric data, Fig. 1 shows how DET curves from the match scores in each bin are calculated. At a selected value of false match ratio (FMR), the false non-match ratio (FNMR) is calculated for $\Delta t = 0$ and compared to that for a chosen value of $\Delta t$, from which $P_B$ is calculated. We do not impose criteria on the selection of FMR; however, it should be chosen at some operationally meaningful level.

### A. Definition

Given these data, we define *biometric permanence*, $P_B(\Delta t)$, for a given elapsed time $\Delta t$, as follows:

$$P_B(\Delta t, \mathrm{FMR}) = \frac{1 - \mathrm{FNMR}_{\Delta t}}{1 - \mathrm{FNMR}_0} \tag{1}$$

where $\mathrm{FNMR}_{\Delta t}$ is calculated from match scores in the $\Delta t$ bin, and the "base" level, $\mathrm{FNMR}_0$ is calculated from scores based on data captured during the same visit (i.e. $\Delta t = 0$). Some features of this formulation are:

- $P_B \to 1$ as $\mathrm{FNMR}_{\Delta t} \to \mathrm{FNMR}_0$ i.e. if there is no increase in FNMR over time, then the permanence is high;
- $P_B$ decreases as $\mathrm{FNMR}_{\Delta t} \to 1$ i.e. as FNMR increases over time, permanence decreases.

In the pathological case where $\mathrm{FNMR}_{\Delta t} < \mathrm{FNMR}_0$, $P_B$ will be greater than 1.

### B. Robust calculation

Given the above definition, it would appear relatively straightforward to calculate biometric permanence from a set of repeated biometric captures. Unfortunately, robust calculation of $P_B$ is complicated. Primarily, the issue is that the effects of interest (small changes in the biometric features) occur in the context of many other changes which are difficult to control experimentally.

For example, in a longitudinal study over several years, there are changes in:

- *weather*: tests at different times of the year expose subjects to – not yet well understood – physiological changes which affect biometric performance (e.g. levels of skin dryness) [14].
- *test adminstrators*: over a period of several years there is inevitably some turn-over in test staff. Not all staff are equally well trained. Some will be more attentive in ensuring proper positioning and placement during biometric tests than others [1].
- *test adminstrator training*: Even it it were possible to eliminate turn-over of staff, the training level of test staff will adapt over time as they become more familiar with the procedure.
- *aging of the biometric sensors*: Biometric sensors are typically built of consumer grade electronics and not intended for many years of useful life. Degradation of some components in the sensors (e.g. lighting) can occur.

We address these issues with the matched delta methodology proposed in this section. In overview, match score data are used to estimate the visit-specific factors (which incorporate the variability above) and to separate them from the changes in match scores caused by template aging effects. These visit-specific factors may then be removed, leaving only the effect of template aging.

Since we collect both enrolment templates and verification images during each visit, we can match all of the enrolment templates against all of the verification images and then visualize the available single-finger match transactions as an $N \times N$ matrix ($N$ the number of visits in the study) in which the upper triangle elements are the 'forward time' matches and the lower triangle are the 'reverse time' matches. Along the diagonal are the *baseline* ($\Delta t = 0$) scores in which each finger image is matched against a template taken only a few minutes before, i.e. at $\Delta t = 0$. The essence of our proposed methodology is that we can substantially remove the per-visit score biases by looking at the difference in scores between a suitably chosen combination of visits and the corresponding baseline visits, and applying these to a composite distribution of the averaged baseline scores.

We use the terminology of "short-delay" to refer to visits spaced less than a few weeks apart, and long-delay for visits years apart. We proceed as follows:

- Calculate a visit pair match score, $G_{i,j,k}$, which represents the match score for biometric characteristic (i.e. individual fingerprint) $k$, between enrolment visit $i$ and verification visit $j$. If multiple biometric captures happen at a single visit, average the match scores across all combinations of enrolment templates from visit $i$ and verification images from visit $j$ for biometric characteristic $k$.
- Identify visit time blocks that may be binned together (see Fig. 2), for
  - same visit subsets within a short-delay;
  - between visit subsets within a short-delay: in our case visits separated by only a few weeks are binned together;
  - short-delay visit subsets over a longer delay: in our case the $2 \times 2$ visits separated by two weeks may be binned together when compared to visits years apart.
- Average each subject-finger score over the elements of each subset of visits binned together, to calculate $\bar{G}_{\tilde{i},\tilde{j},k}$, where $\tilde{i}$ and $\tilde{j}$ represent represent each visit when considering binning.

So, for example, given a sequence of four visits occurring respectively between 2012-02-12 and 2012-03-03; between 2012-03-12 and 2012-03-31; between 2013-03-06 and 2013-03-25; and between 2013-04-05 and 2013-04-27, the 'Year1U' (one year interval, upper triangle) would consist of all those matches between enrolment templates obtained during the two visits spanning 2012-02-12 to 2012-03-31, and verification presentations obtained during the two visits spanning 2013-03-06 to 2013-04-27; while the 'Year1L' (one year interval, lower triangle) would consist of the corresponding 'backward' matches between verification presentations obtained in the

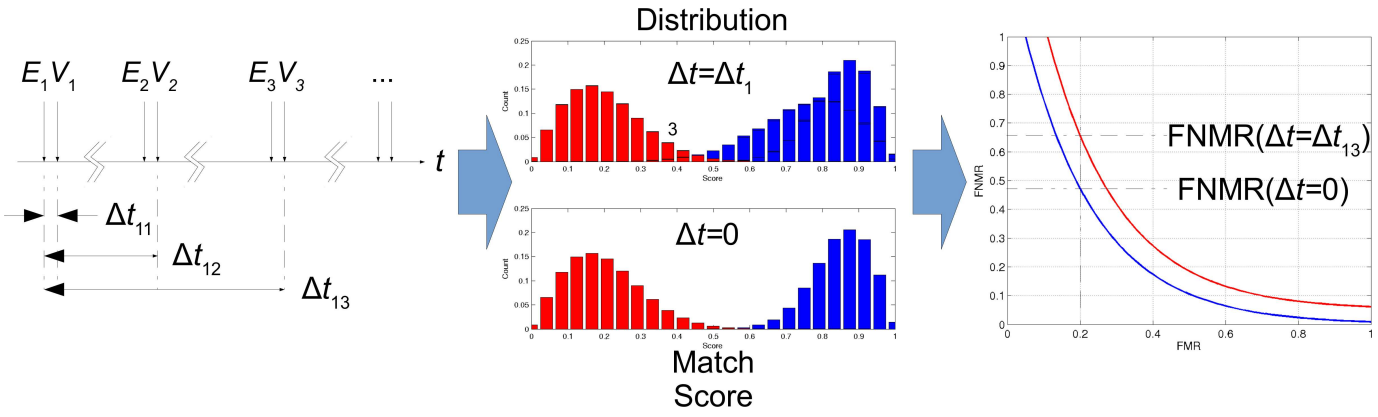


Fig. 1. Conceptual overview of our method. *Left:* Over time, a sequence of biometric capture events occur at which enrolment ($E$) and verification ($V$) events occur. The time between visits $a$ and $b$ is defined as $\Delta t_{ab}$. *Right:* Using only biometric data for a given value $\Delta t$, various DET curves can be calculated; here curves for $\Delta t = \Delta t_{13}$ and $\Delta t = 0$ (i.e. from the same visit) are shown. *Biometric Permanence* at a given $\Delta t$ is then defined in terms of the change in FNMR (at a chosen FMR) beween $\Delta t = 0$ and the considered interval.

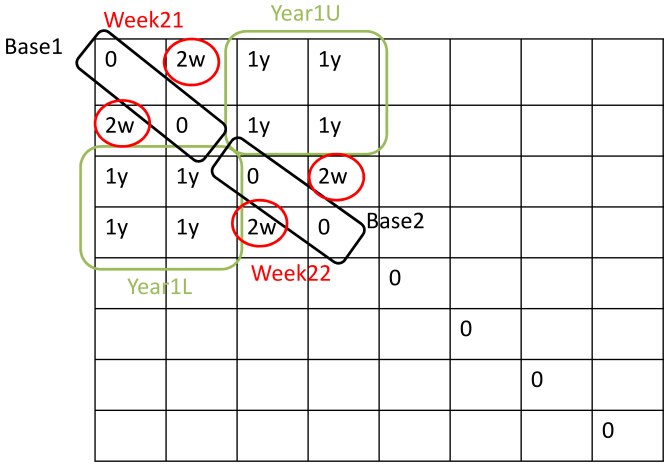


Fig. 2. Matrix of match scores for a single subject over the protocol. Each row and column represents a visit (with enrol, $E_i$ and verification, $V_i$ records). In our testing protocol, each round of testing has a pair of visits separated by two weeks. The upper triangle represents match scores "forward in time" ($E_i$ vs. $V_j$, $i < j$), while the lower represents the corresponding match scores "backward in time" ($E_i$ vs. $V_j$, $i > j$). Match scores on the diagonal are from the same visit (i.e. $\Delta t = 0$).

earlier year, 2012-02-12 to 2013-03-31, against enrolment templates obtained in the later year, 2013-03-06 to 2013-04-27. The corresponding '2w' (short-delay) matches would be evaluated between enrolment templates obtained during 2012-02-12 to 2012-03-03 and verification presentations obtained during 2012-03-12 to 2012-03-31, and between enrolment templates obtained during 2013-03-06 to 2013-03-25 and verification presentations obtained during 2013-04-05 to 2013-04-27, along with their corresponding backward matches below the diagonal.

At this point, each subject-finger score, $\bar{G}_{\tilde{i},\tilde{j},k}$, is associated with a time interval. We then identify the corresponding averaged baseline scores $\bar{G}_{\tilde{i},\tilde{i},k}$, and $\bar{G}_{\tilde{j},\tilde{j},k}$ corresponding to the time interval $\tilde{i}, \tilde{j}$. We then:

- Average the forward- and backward-time subject-finger scores for each major interval, to calculate $\frac{1}{2}(\bar{G}_{\tilde{i},\tilde{j},k} + \bar{G}_{\tilde{j},\tilde{i},k})$; for example, 'Year1U' and 'Year1L' for the '1-year' interval in our case (Fig. 2).
- Subtract the average of the two corresponding base mean scores for each subject-finger, $\frac{1}{2}(\bar{G}_{\tilde{i},\tilde{i},k} + \bar{G}_{\tilde{j},\tilde{j},k})$.
- Finally, average over subject-fingers to get a mean matched delta score.

Thus we calculate $\Delta G_{\tilde{i},\tilde{j}} = \Delta G_{\tilde{j},\tilde{i}}$, where

$$\Delta G_{\tilde{i},\tilde{j}} = \frac{1}{2} \mathop{\mathrm{E}}_{k} \left[ \bar{G}_{\tilde{i},\tilde{j},k} + \bar{G}_{\tilde{j},\tilde{i},k} - \bar{G}_{\tilde{i},\tilde{i},k} - \bar{G}_{\tilde{j},\tilde{j},k} \right] \quad (2)$$

where the expectation, $E$, is over all subject fingers, $k$. Since expectation is linear, the exact order in which the various averages are applied is largely a matter of computational convenience.

The mean match deltas, $\Delta G_{\tilde{i},\tilde{j}}$, are then used to shift a single representative genuine distribution, constructed from the subset of match scores along the diagonal (i.e. the distribution of baseline scores) and $P_B$ is calculated from this shifted composite distribution using the definition of Eq. 1 in the normal way.

We make a practical assumption – justified in our case by visual examination of the match score histograms – that the imposter score distributions are not significantly affected by aging, and hence only process the genuine match scores in the above way. Note that we do not require individual imposter matches to be unaffected; rather, we argue that the effect that reduces the match score for a particular imposter is equally likely to increase it for another. This assumption is merely a computational convenience and the method we describe here could be extended to the imposter scores if required – for example, for a different biometric which did, in fact, show aging in the imposter distributions.

## III. Simulation

We have defined biometric permanence, $P_B$ using, first, a model based on verification calculations alone ("reference metholology", the value calculated this way is $P_{B,R}$); and next, a "matched delta methodology" ($P_{B,M}$) in which visit variability is modelled and removed by calculating the shift in the genuine distribution. Our expectations are:

- In the absense of visit-to-visit variability, $P_{B,M}$ is an unbiased estimate of $P_{B,R}$.
- $P_{B,M}$ is a lower variance estimator of $P_{B,R}$ (since a single parameter is estimated from data before the DET calculation, which is known to be noisy).
- In the presence of visit-to-visit variability, $P_{B,M}$ will show more plausible results than $P_{B,R}$.

In this section, we develop a numerical model to evaluate these expectations.

We assume that, for a particular product, there is an underlying genuine match score for the $D^{th}$ finger ($D = 1, 2, \ldots, 10$) that may be described by a function $S^D(\Delta t)$ where $\Delta t$ is the time difference between the finger's enrolment ($E$) and verification ($V$) visits. Here, $S^D$ is distributed over experimental subjects, i.e. for each subject $n$ there is a single function $s^D(\Delta t)$, which will, in general, be subject to measurement noise which we discuss below.

In the most general formulation, the measured scores have some unknown functional dependence on the enrolment visit $E_i$ and verification visit $V_j$, i.e. we are only able to observe $\tilde{S}^D_{ij} = f_{ij}(S^D(\Delta t))$. In order to proceed however, we need to develop a tractable model of the visit dependence, which we model as:

$$f_{ij}(S^D(\Delta t)) = S^D(\Delta t) + a_i + b_j \tag{3}$$

where $a_i$ and $b_j$ represent enrolment visit $i$ and verification visit $j$ biases, respectively. A more complete model would use biases $a_i + b_j + c_{ij}$ representing both visit-specific and inter-visit terms, but we do not consider it here.

In addition to the visit-dependent bias, we include in the model a measurement noise term $W^D_{n;i,j}$ in the sense that repeated presentations of the same subject-finger within the same enrolment-verification visit pair will produce different scores. We might expect this term to vary from finger to finger, and to be anti-correlated to the mean score because of the (sometimes strong) non-linearity of some vendors' matching algorithms. In particular, once the quality of the enrol- and verify-fingerprints exceeds a threshold, the match scores for certain devices will saturate, such that the within-finger variability tends to zero. Without loss of generality we can choose $W$ to be zero-mean i.e. $\langle W^D_{n;i,j} \rangle = 0$ where the angled brackets denote averaging over presentations of the same finger. In the case of our experiment, the averaging is over six verification presentations, with each enrolment consisting of three presentations. The latter are aggregated into a single biometric information record (BIR) or *template* in a vendor-dependent way, and are not available for explicit averaging during generation of the match scores. We index the within-finger variability as $W^D_{n;i,j}$ to remind ourselves that it varies over enrolment-verification visit pairs $i, j$, although likely only implicitly; that is, via the mean match score itself.

Hence we can write the observed score for the $D^{th}$ finger of subject $n$ as

$$f_{ij}(S^D(\Delta t)) = S^D(\Delta t) + a_i + b_j + W^D_{n;i,j} \tag{4}$$

We take as our starting point a pair of canonical distributions for the underlying false and genuine match scores, where scores range between 0 and 1. In this work we have chosen to model the false scores as a simple Rayleigh distribution, and the genuine scores as a "flipped" Rayleigh distribution. In suitably normalized form the probability densities become

$$p_I(s) = \frac{s}{\beta_I^2} e^{-s^2/2\beta_I^2}; s \geq 0 \tag{5a}$$

$$p_G(s) = \frac{1-s}{\beta_G^2} e^{-(1-s)^2/2\beta_G^2}; s \leq 1 \tag{5b}$$

where the scale parameters $\beta_{I,G}$ are related to the mean scores by $\mu_I = \beta_I \sqrt{\frac{\pi}{2}}$ and $\mu_G = 1 - \beta_G \sqrt{\frac{\pi}{2}}$. To model the sequence of visits, we then make two assumptions, namely

- *The imposter distribution remains constant over time.* Since the imposter distribution represents non-matched samples, it will not be significantly sensitive to changes due to elapsed time. This is roughly equivalent to the assertion that for any pair of non-identical fingers whose match score is improved by some mechanism, there is another pair whose match score is correspondingly reduced.
- *With each visit $k$ we can associate a pair of systematic biases $a_k, b_k$ (for enrolment and verification, respectively) that affect all presentations equally.* We then model the effect of a match score between a verification presentation from visit $j$ against a template recorded in visit $i$ as a shift in the genuine distribution equal to $(a_i + b_j)$

We then model the template aging as a further, time-dependent, shift of the genuine distribution. In this work we choose a simple time-symmetric assumption, namely $\delta(\Delta t) = -\alpha(1 - \exp(-\kappa|\Delta t|))$ i.e. a term that asymptotically approaches $\delta = -\alpha$ with time-constant $1/\kappa$.

To each variate generated according to Eq. 5 we apply an additional zero-mean Gaussian noise term $W_{I,G}$ to represent the natural variation in match score over repeated presentations of the same finger against the same template (effectively a kind of 'measurement noise').

### A. Simulation of a single sequence of visits

To make a baseline qualitative evaluation of the efficacy of our method in removing the experimental biases, we constructed a single realization of a sequence of eight visits, with bias values $a_k, b_k$ taken from a Gaussian distribution with standard deviation 0.025. We then generated canonical sample distributions according to Eq. 5 for each of two fingers for a sample size of 17500 subjects

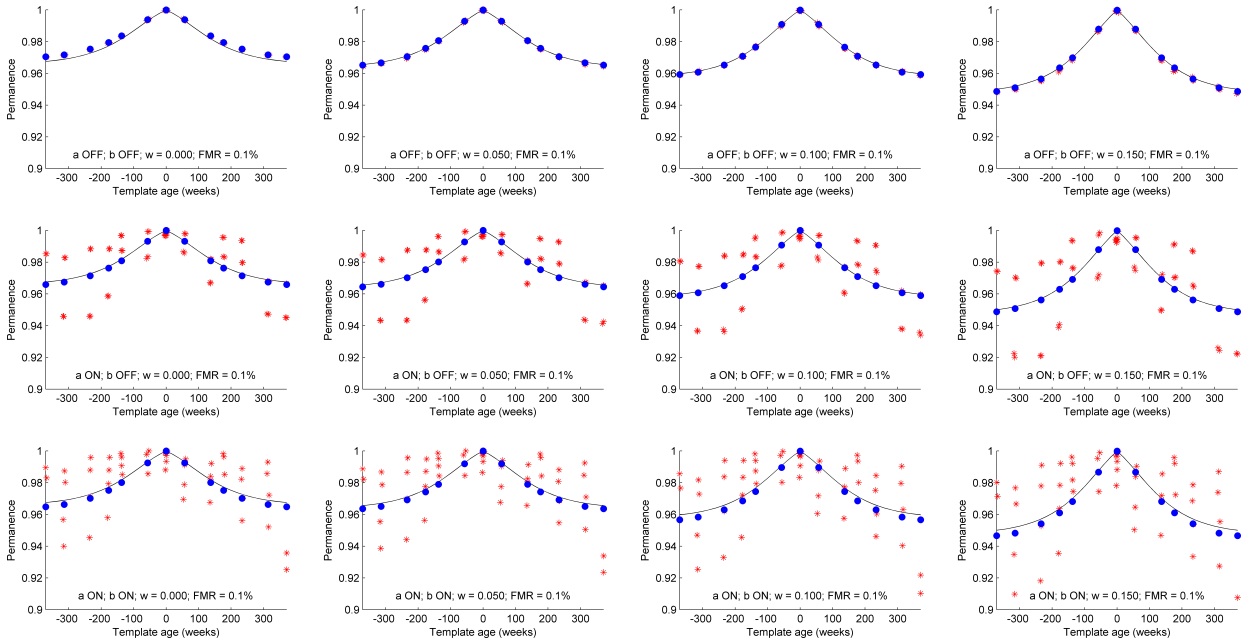


Fig. 3. Exploration of the effect of visit biases. Each figure shows $P_B$ (vertical axis) vs. elapsed time (horizonal axis) in weeks between enrolment and verification. The {ON,OFF} status of parameters $a$ (enrolment) and $b$ (verification) represents whether the corresponding biases are simulated or not. Thus "a:ON b:ON" indicates simulation of both enrolment and verification biases. $w$ is the standard deviation of simulated presentation noise. *Top row*: noise only; *middle row*: noise and enrol visit biases; *bottom row*: noise, enrol and verify bias. The red stars are the reference method of Section II-A while the blue circles are our matched delta method of Section II-B. The black curve is derived from the analytical tail integrals of Eq. 7.

with $\mu_I = 0.2$ and $\mu_G = 0.85$ for each finger. Next, for each of the $8 \times 8 = 64$ enrol-verify visit combinations, we modify the finger scores according to:

$$S_{i,j}^{1,2}(\Delta t) = S_0 + a_i + b_j - \alpha(1 - \exp(-\kappa|\Delta t|)) \quad (6)$$

Finally, we synthesize six independent presentations of each finger by adding a zero-mean Gaussian noise term with variance $\sigma^2 = w^2/6$: this scaling is a convenience, so that we can identify $w^2$ with the sample variance of the presentation-averaged experimental scores. We then processed the scores in two ways: (i) a simple direct calculation according to Section II-A; and (ii) our matched delta method as described in Section II-B.

We ran the simulation with three scenarios: first, with no visit biases ($b_k = a_k = 0$); second with bias in the creation of enrol templates only ($b_k = 0; a_k \neq 0$); and third with bias in both enrol template creation and verification presentation ($b_k \neq 0, a_k \neq 0$) (Fig. 3). The value of the presentation-averaged sample standard deviation $w$ was varied in the range 0 to 0.075 units and the aging parameters were $\alpha = 0.1$ unit and $\kappa = 0.01$ week$^{-1}$. The reference FMR for the permanence calculation was 0.001 (0.1%).

Taken together, the visit bias terms and aging correspond to a modification $(1 - s) \rightarrow (1 - a_i - b_j - \delta(t) - s)$ in the genuine score distribution of Eq. 5. The densities then become convolved with the Gaussian density — one can show that the resulting tail integrals for the FMR and FNMR at some threshold score $\theta$ become:

$$FMR(\theta) = \frac{\beta_I}{\beta'_I} e^{-\theta^2/2\beta'^2_I} \quad (7a)$$

$$FNMR(\theta; i, j; t) = \frac{\beta_G}{\beta'_G} e^{-(\chi(t;i,j)-\theta)^2/2\beta'^2_G} \quad (7b)$$

where $\beta'^2_{I,G} = \beta^2_{I,G} + \sigma^2_{I,G}$ (with $\sigma^2_{I,G}$ being the variances of the imposter and genuine presentation noise terms respectively) and

$$\chi(t; i, j) = 1 - a_i - b_j - \delta(t) \quad (8)$$

represents the mean degradation in genuine match score due to the per-visit biases and template aging. The solid black curves in Fig. 3 correspond to Eq. 7 , 8 with $a_i = b_j = 0$ i.e. they represent the 'ideal' (unbiased) aging behavior that would be observed due to presentation noise only: the deviation from this curve can be interpreted as the residual effect of bias that is not removed by the technique.

### *B. Simulation of an ensemble of visit sequences*

Over an ensemble of experiments (that is, sequences of enrol-verify visits) with experimental biases taken independently from some zero-mean distribution(s), one would like to show that the mean of the reference permanence measure (Eq. 1) does indeed converge to the value obtained by our new technique. Accordingly we ran the same procedure up to 180 times to simulate multiple independent realizations of our experiment. At intervals of 5 simulated experiments we calculated the mean deviation (across time differences $\Delta t_m$)

between the permanence calculated according to our method (Section II-B) and reference method (Section II-A) (Fig. 4).

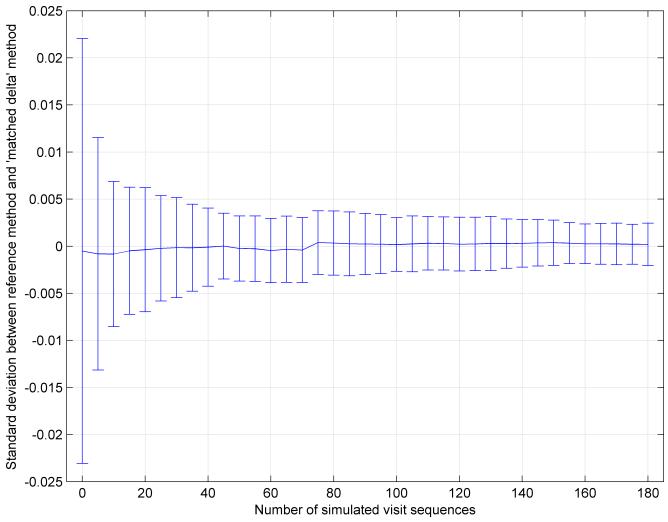


Fig. 4. Difference ($\pm$SD) between the reference measure Section II-A and the matched delta method of Section II-B as the size of the experiment ensemble is increased.

## IV. Discussion

Slow changes in biometric features over time are typically referred to as "template aging", and the performance of large-scale systems can be influenced by this effect. Unfortunately template aging is hard to measure, because it is very sensitive to the visit-to-visit variability inherent in such a study (e.g. test personnel, test equipment and weather).

In the case of fingerprints, Uludag et al. [15] addressed the case of typicality an/or variability between presentations of the same biometric using novel template selection algorithms, based either on clustering or on mean distance. They then used this template selection to evaluate a number of template update schemes. They found that a scheme in which an original template was updated selectively using later presentations ("AUGMENT-UPDATE") outperformed one in which the original template data were discarded altogether ("BATCH-UPDATE"). From this, we might infer that the magnitude of the template aging effect was not significantly greater than that of the intraclass variance, at least over the relatively short interval of their study (approximately four months).

A comparable study into face recognition was recently conducted by Manjani et al. [10]. Fingerprint aging might be expected to share some of the same physiological factors that they identified for face aging – in particular, skin textural changes and loss of tissue elasticity. They evaluated both 2D and 3D facial recognition algorithms on a dataset of sixteen participants acquired over a period of ten years, comparing genuine acceptance rate (GAR) at 0.1% false acceptance rate (FAR) for short-term intervals (less than three months between enrolment and verification) versus long-term intervals (more than five years between enrolment and verification). Unlike the present work, the intervals were not blocked into absolute acquisition times i.e. all intervals greater than five years were taken together. They were able to reject at $\alpha = 0.05$ the null hypothesis that the short- and long-term genuine scores were drawn independently from normal distributions of equal mean and variance (t-test), or from the same continuous distribution (Kolmogorov-Smirnov test). In the case of the algorithm that performed best over the long-term intervals ("3D Region Ensemble: Product"), they found weak evidence against the corresponding hypotheses for the imposter scores: this is consistent with our model, in which the imposter distribution was assumed to be constant.

Template aging in the iris modality was addressed recently by Hofbauer et al. [5]. They noted some controversy about its existence, and discussed the difficulty of controlling confounding factors independently – in particular, the cases of illumination and pupillary dilation. There was only a single long-term time interval – in this case of four years – while the study consisted of data from 47 subjects. The authors considered two schemes for re-normalization of pupil diameters: a "rubber sheet model" (RSM) and a "biomedical model" (BMM). They showed that while such re-normalizations were somewhat effective in improving long-term match accuracy, there was still a decrease in performance between intra-year and inter-year comparisons. This suggests that while systematic changes in pupillary diameter are a factor in iris template aging, they are not the only such factor.

In this paper we have developed and defined a measure of template aging which we call biometric permanence $P_B$, based on the change in FNMR (at a given FMR) between the template aging interval under test, and a short-time test. While intuitive, this definition of $P_B$ is practically difficult to apply to estimate small changes in permanence in a longitudinal study subject to experimental error and visit-to-visit systematic biases. To address this issue, we have introduced the matched delta method. Comparisons of these methods were performed using simulated data, and it was determined that the new method showed dramatically reduced sensitivity to systematic biases. Simulations were designed to evaluate two aspects of the proposed robust calculation method in comparison to the calculation of $P_B$ from Eq. 1. First, simulations test the first and second order statistical properties (i.e. bias and variance); and, second, the sensitivity of the methods to visit-to-visit biases.

Fig. 3 compares the two methods to the analytical values (Eq. 7). For two different values of presentation noise (columns), the presence or absence of visit biases is evaluated. A single sample of visit biases is evaluated in each case; multiple values at a single time interval indicate different ways in which the given time offset can be calculated. Without visit biases, the methods perform similarly, while their presence dramatically impacts the values calculated using Eq. 1. Meanwhile Fig. 4 investigates the statistical properties of the methods. As sample number increases, the variance decreases and bias between methods decreases towards zero, as expected

from our analytical model.

We are in the process of applying this methodology to data collected in a multi-year, multi-vendor experimental fingerprint acquisition and matching study, involving over 350 participants, with a gallery size in excess of 12,000 ISO/IEC standards-compliant two-finger biometric enrolment templates and obtained with a variety of commercially-available fingerprint sensor technologies.

## REFERENCES


[1] M. E. Brockly. The role of test administrator and error. Master's thesis, Purdue University, West Lafayette, IN, December 2013.

[2] M. Drahansky, M. Dolezel, J. Urbanek, E. Brezinova, and T. Kim. Influence of skin diseases on fingerprint recognition. *Journal of Biomedicine and Biotechnology*, 2012. Article ID 626148.

[3] S. Elliott, E. Kukula, and N. Sickler. The challenges of the environment and the human/biometric device. In *Proc. International Workshop on Biometric Technologies*, 2004.

[4] S. P. Fenker and K. W. Bowyer. Experimental evidence of a template aging effect in iris biometrics. In *Proc. 2011 IEEE Workshop on Applications of Computer Vision (WACV) (WACV '11)*, pages 232–239, 2011.

[5] H. Hofbauer, I. Tomeo-Reyes, and A. Uhl. Isolating iris template ageing in a semi-controlled environment. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, Sept 2016.

[6] ILO. ILO seafarers identity documents biometric testing campaign report (part i). Technical report, International Labour Organization (ILO), November 2004.

[7] ISO. ISO/IEC 19795-1:2006 - Information technology – Biometric performance testing and reporting – Part 1: Principles and framework. Technical report, 2016.

[8] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Biometrics*, 2004.

[9] D. Maltoni, D. Maio, A. Jain, and D. S. Prabhakar. *Handbook of fingerprint recognition*. Springer Professional Computing, 2009.

[10] I. Manjani, H. Sumerkan, P. J. Flynn, and K. W. Bowyer. Template aging in 3d and 2d face recognition. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, Sept 2016.

[11] Y. Matveev. The problem of voice template aging in speaker recognition systems. In *Proc. 15th International Conf. Speech and Computer (SPECOM) 2013*, Pilsen, Czech Republic, September 2013.

[12] J. Ryu, J. Jang, and H. Kim. Analysis of effect of fingerprint sample quality in template ageing. In *NIST Biometric Quality Workshop II, November 2007*. NIST, 2007.

[13] N. C. Sickler and S. J. Elliott. An evaluation of fingerprint quality across an elderly population vis-à-vis 18–25 year olds. In *Security Technology, 2005. CCST '05. 39th Annual 2005 International Carnahan Conference on*, pages 68,73,, October 2005.

[14] R. F. Stewart, M. Estevao, and A. Adler. Fingerprint recognition performance in rugged outdoors and cold weather conditions. In *IEEE Int. Conf. Biometrics: Theory, Applications and Systems (BTAS09)*, pages 28–30, Washington DC USA, September 2009.

[15] U. Uludag, A. Ross, and A. Jain. Biometric template selection and update: a case study in fingerprints. *Pattern Recognition*, 37:1153–1542, 2004.

[16] A. Wolf. Template aging in speech biometrics. *Proc. Biometrics: Challenges Arising from Theory to Practice (BCTP)*, 2004.