Bayesian Analysis of Fingerprint, Face and Signature Evidences with Automatic Biometric Systems

Joaquin Gonzalez-Rodriguez, Julian Fierrez-Aguilar, Daniel Ramos-Castro and Javier Ortega-Garcia

ATVS (Speech and Signal Processing Group), Escuela Politecnica Superior,
Universidad Autonoma de Madrid,
E-28049 Madrid, Spain
Tel.: +34-91-4973142, fax: +34-91-4972235
email: joaquin.gonzalez@uam.es

Abstract— **The Bayesian approach provides a unified and logical framework for the analysis of evidence and to provide results in the form of Likelihood Ratios (LR) from the forensic laboratory to Court. In this contribution we want to clarify how the biometric scientist or laboratory can adapt their conventional biometric systems or technologies to work according to this Bayesian approach. Forensic systems providing their results in the form of LR will be assessed through Tippett plots, which give a clear representation of the LR-based performance both for targets (the suspect is the author/source of the test pattern) and non-targets. However, the computation procedures of the LR values, especially with biometric evidences, are still an open issue. Reliable estimation techniques showing good generalization properties for the estimation of the between- and within-source variabilities of the test pattern are required, as variance restriction techniques in the within-source density estimation to stand for the variability of the source with the course of time. Fingerprint, face and on-line signature recognition systems will be adapted to work according to this Bayesian approach showing both the Likelihood Ratios range in each application and the adequacy of these biometric techniques to the daily forensic work.**

Index Terms— Bayesian, biometrics, face, fingerprint, forensic, signature.

I. INTRODUCTION(*)

This contribution deals with how forensic scientists should report to the judge/jury their conclusions when automatic biometric identification techniques are used, such as fingerprint, face, or signature recognition (figure 1). In this sense, we will firstly note the difference from system characterization, that is, the identification abilities of the recognition technique in use, with respect to the characterization of the forensic system that will provide objective results to the Court. This is the key issue of this contribution as forensic scientists should not arrogate the role of the

(*) This work has been partially supported under MCYT Projects TIC2000-1683, TIC2000-1669, TIC2003-09068, TIC2003-08382 and Spanish Police Force "Guardia Civil" Research Program

judge/jury in taking decisions. Making use of thresholds allows the forensic scientist to take the actual decision, that belongs to the court. So, scientists must know how to submit their objective results in order to comply with all the conditions of the judicial procedures, converting the system identification scores in meaningful values useful to the Court.

While commercial biometric verification systems performance, oriented to acceptance or rejection decisions, are widely assessed through different classical decision-based criteria, as type I and II errors or ROC/DET plots [1], an intense debate among forensic practitioners have taken place during the last decade in order to achieve a common framework for the evaluation of evidence and its interpretation in Court, and then how to assess the performance of forensic systems. Nowadays, the Bayesian approach [2][3][4], which propose to submit results to Court in the form of Likelihood Ratios, has been proposed as a theoretical framework valid for any forensic discipline, where systems providing its results according to this approach, from the large experience gained in DNA-based forensic individualisation, can be assessed through Tippett plots as a way to represent forensic systems performance. Firstly used in [5], Tippett plots are based on experiments in paint flakes performed by Tippett et al., where distributions of LRs computed were presented separately in matching and non-matching experiments. In this paper, we will show the different nature of the outputs that automatic biometric systems can provide respectively in commercial and forensic approaches, even if the systems use the same core technology, and subsequently the need for different assessment tools specially suited for their corresponding applications.

The paper is organized as follows. In Section II we will show the different objectives of commercial biometric systems and forensic biometric identification and the different needs of specific characterization techniques, as the scores provided by both commercial and forensic systems will be completely different. In Section III we will introduce the Bayesian approach for evidence analysis and forensic reporting which will perfectly suit both the needs of the Court and those of the forensic scientist, and in Section IV we will show how to assess the performance of any forensic system according to this Bayesian approach in the form of Tippett plots. Once the Bayesian approach is understood, we will show in Section V how any biometric system can be adapted to provide its results in the form of Likelihood Ratios (LR) in this Bayesian environment, being so converted into a forensic identification system, focusing in different procedures for between- and within-source density estimation, the generalization properties of these estimation techniques, and the problems induced from the usual lack of data in forensic cases for density estimation. Finally, in Section VI a detailed description of the adaptation of three different biometric systems, namely fingerprint, face and on-line signature recognition, to the forensic environment according to the Bayesian approach is shown. Results in the form of Tippett plots will show the expected LR values range in each application and the adequacy of these video-based techniques to the forensic biometrics tasks. Some conclusions will be finally extracted in Section VII, concluding with a full list of references to allow a deep insight into any of the aspects of this contribution.

II. BIOMETRIC SYSTEMS AND CLASSICAL FORENSIC REPORTING

A. Assessment of Biometric Systems

The typical objective of commercial biometric systems is to accept true users and to reject impostors, usually minimizing some type of cost function as false acceptance (also known as false alarm in detection theory) and false rejection (missed detections) may occur. The usual operation mode of any biometric system is the following: in the presence of an unknown pattern and a claimed user identity, the system will compare a reference pattern/model from the claimed user with the input pattern, providing a matching score which will be compared with a predetermined threshold for that specific user for the acceptance or rejection of the input pattern as belonging to the claimed user.

Commercial, score-based biometric systems can perform two different tasks, namely *identification* and *verification*. In the former, the trace is compared to N suspect models, leading to N similarity measures (*scores*), one per suspect. Theses measures can be ordered, and selection of one or more "most-similar-models" would imply the establishment of a threshold. On the other hand, a verification task imply that the trace is compared to a claimed identity, and a threshold decides whether the trace belongs or not to the claimed identity. Some forensic disciplines could work in a similar way as a biometric identification system does, as they select one unique or a subset of individuals from a relevant population. Biometric verification systems can be viewed as decision systems classifying a trace in one of two classes, namely *accepted* or *rejected*, regarding the claimed identity. In both cases, the selection of a threshold is needed for taking decisions.

In order to assess the identification abilities of any biometric system, the system must be tested with known users and impostors, task which is usually performed using databases of the corresponding input patterns (fingerprints, faces, signatures...). Two types of error can occur in a verification (i. e., detection) system: false rejection (type I error), when a true user is rejected, and false acceptance (type II error), when an impostor is accepted. The probability of any of these two errors is a trade-off: this means that if the threshold is increased, the false acceptance will be reduced but the false rejection will be increased, and vice versa. As the same system or technology could work in different operating conditions, it is usual to show all possible operating points. This has been done classically in detection tasks by means of ROC (Receiver Operating Characteristic) curves, showing the tradeoff between missed detection (false rejection) and false alarm (false acceptance). In order to have a single value characterizing the global performance of the system, the Equal Error Rate (EER) is usually given, which is the point where the probability of a missed detection equals the probability of a false alarm.

However, as biometric systems performance increases and errors decrease, comparison of systems have become extremely difficult with this representation, as curves from different systems are extremely close to the lower left corner. This problem was overcome with the introduction of the DET (Detection Error Tradeoff) curve [1], which allows an almost linear representation of system performances, improving the observation of system contrasts (real ROC and DET plots for the same system in different conditions are shown in figure 2).

We want to note that this type of performance assessment (ROC/DET) perfectly suits the objective of the assessed systems, that is, to accept or reject users, because it directly shows both types of possible errors (missed detections and false alarms). Obviously, the core technology in use within any forensic biometric system can also be used in a

commercial biometric system and assessed through ROC/DET curves or EER value, in order to see how the system discriminate customers from impostors.

B. Is Acceptance/Rejection the Objective of Forensic Biometrics?

In the last years, the value of the different types of forensic evidence (even traditionally firmly established areas such as fingerprint identification) have been severely attacked, questioning their scientific status, as is shown in influential books [6][7] and papers [8] in the field specially *"...after several highly publicized miscarriages of justice in which forensic expertise played a crucial role*" [9].

Classically, there have been two different approaches to forensic reporting in "individualization of the source" areas, which includes areas as fingerprint, firearms and toolmarks. The first approach has been to provide just "identification" (*acceptance*) or "exclusion/elimination" (*rejection*) decisions, which results in a very high percentage of non-reporting cases. This approach has two main drawbacks: the first one is related with the use of subjective thresholds, as these techniques does not provide absolute identifications, specially in forensic conditions, and all that the system/technique can provide is a score or a probability. Then, if the forensic scientist takes the (subjective) decision of identification or exclusion/rejection, he will be ignoring the prior probabilities related to the case (independent of the evidence under analysis), usurping the role of the Court in taking this decision, as "*...the use of thresholds is in essence a qualification of the acceptable level of reasonable doubt adopted by the expert*" [10]. This fact is well known by the court mainly in DNA analysis, although it is a problem concerning all forensic disciplines [11]. The second drawback is the possibility of existence of a large amount of non-reporting cases that this "identification" or "exclusion" process can induce, when *"...there is no logical reason to suppress probability statements ... because ... any piece of evidence is relevant if it tends to make the matter which requires proof more or less probable than otherwise"* [10]. The second classical approach to forensic reporting in this area consists in the use of a verbal scale of identification probabilities (typically "identification" / "very probable" / "probable" / "not conclusive" / "elimination"). This approach falls in the same errors as has just been noted, as it makes use of several subjective thresholds, but again ignores the prior probabilities (or usurp the judge/jury role if assigns them) relative to every case.

III. Bayesian Analysis of Forensic Evidence

Fortunately, the Bayesian (or *Likelihood-Ratio*, LR) approach has been proposed as a theoretical framework valid for any forensic discipline [2][3][4]. As an example, there are eight Working Groups (DNA, Fibers, Fingerprint, Firearms, Handwriting, Tool Marks, Paint and Glass, Speech and Audio) in ENFSI (European Network of Forensic Science Institutes) dealing with individualization of the source. Some of them [12], in discussions open also to non-European participants, have dealt or are dealing with the Bayesian approach at source (DNA) or activity (Fibers, Paint and Glass) levels [13], looking for common standards and procedures.

In this Bayesian framework, the roles of the scientist and the judge/jury are clearly separated, as the Court wants to know the odds in favor of the prosecution proposition (C), ("the biometric trace belongs to the suspect"), given the

circumstances of the case (I) and the observations made by the forensic scientist (E). These odds in favor of C are obtained from Eq. (1):

$$O(C|E,I) = \frac{Pr(E|C,I)}{Pr(E|\overline{C},I)} \cdot O(C|I) \qquad (1)$$

Expressed in words, *Posterior Odds = Likelihood Ratio x Prior Odds*, where the prior odds concern to the Court (background information relative to the case) and the likelihood ratio (LR):

$$LR = \frac{Pr(E|C,I)}{Pr(E|\overline{C},I)} \qquad (2)$$

is provided by the forensic scientist. As a reference, in [2] a scale of likelihood ratios (LR) in the framework of DNA analysis is proposed with their respective linguistic qualifier suggesting the strength of verbal support for the evidence. This scale, as shown in Table 1, is actually in use in DNA laboratories and is being extended to other identification areas and labs all over the world.

The use of the Bayesian approach is recommended because "*...assists scientists to assess the value of scientific evidence, help jurists to interpret scientific evidence, and clarify the respective roles of scientists and of members of the Court*" [10]. In this way, the scientist alone cannot infer the identity of the author/source of the questioned biometric pattern from the analysis of the scientific evidence, but gives the Court the likelihood ratio of the two competing hypothesis (usually C, "the biometric trace belongs to the suspect", and $\overline{C}$, "the biometric trace belongs to someone else but not the suspect").

This Likelihood Ratio, or Bayes factor, must be determined by the forensic scientist. In order to compute these numerator and denominator probabilities, population data allows the forensic scientist to determine objective probabilities (figure 3). For score-based systems, as all biometric techniques are, distribution of measurements can be modeled from data, both within and between sources, as this LR is in this case a ratio of probability density functions evaluated at the evidence score. Also, Bayesian analysis allows to easily include subjective opinions in the form of subjective probabilities [14].

Moreover, the Bayesian approach allows to combine different types of evidence present in a case (voice, fingerprint, ...) and even the incorporation of subjective probabilities related to uncertain events, as shown in [4], providing an unified approach to the joint analysis of any type of evidence.

IV. Assessment of Forensic Biometric Systems

In order to test the abilities of systems providing their results in the form of LR values (as it is described in [15], [16], [17] and [18]), some system calibration experiments have to be performed. Based on the work by Tippett et al. [19], in [5] a useful representation for between-source comparisons in any forensic discipline, the so-called Tippett plots, is provided, representing proportion of cases with "LR values greater than…" (figure 4). Then, we will draw in each Tippett plot simultaneously two curves, one for the C hypothesis (the pattern belongs to the suspect – targets),

where the system should provide high LR values (LR>>1), and another one for the $\overline{C}$ hypothesis (the pattern does not belong to the suspect – non-targets), where the system should provide low LR values (LR<<1). In this way, for any *x*-axis value each curve shows proportion of cases with LR greater than *x*. Then, the greater the separation between curves, the higher the strength of the prior to posterior inference by means of the LR value, and the better the forensic system (in an ideal system the curves should adjust respectively to the upper-right and lower-left margins of the plot). Additionally, good performance of both curves around LR=1 is desired, that is, the proportion of targets with LR<1 and non-targets with LR>1 should be as small as possible. Moreover, in forensic applications it is convenient, in order to guarantee the presumption of innocence, that non-target suspects do not obtain LRs greater than one, even if this condition leads to worse performance (smaller separation between target and non-target curves in Tippett plots).

As no threshold should be involved in forensic reporting under the Bayesian methodology, ROC/DET curves remains useful and perfectly suited to assess performance in detection tasks, but another way of performance assessment has to be used in a Bayesian forensic environment, where no threshold is established. Tippett plots are ideal to present system performance in such methodology.

V. Computation of Likelihood Ratios in Forensic Biometrics

Unfortunately, there is no closed solution to the problem of Likelihood Ratio (LR) computation in all the different classical identification-of-the-source forensic areas (DNA, Fibers, Fingerprint, Firearms, Handwriting, Tool Marks, Paint and Glass, Speech). While it is assumed that the numerator of the LR calls for an assessment of the intra-variability of the studied feature for the putative source, the denominator is the random match probability of the same feature over the relevant population of sources. Both can be obtained from objective or subjective measures. There exist two different alternatives for the evaluation of the evidence, depending on the existence of categorical data (e.g., the probability of the suspect blood type) or continuous data (e.g., the refractive index of a piece of glass found in the suspect clothes). In case of biometric evidence types, we are always dealing with continuous data as our "evidence" will be scores obtained from the comparison of a known pattern from the suspect (the "model") with the questioned biometric pattern from the crime scene. Those scores are the output of generic biometric systems, which compared to a threshold give a decision of acceptance or rejection. in commercial biometric systems. In forensic systems, where we do not look for an acceptance or rejection decision, those scores need to be referenced into the within-source variability and between-source variability in order to provide to Court the likelihood ratio of the competing hypothesis.

In [20] a solution is proposed for evaluation of continuous data from glass evidence, but we want to focus in the alternative shown in [21] for the problem of forensic speaker recognition using automatic speaker recognition techniques. In this paper we have adapted that proposal for its use in any forensic biometric area, as any automatic biometric system will provide scores relating identity models with questioned biometric patterns.

The approach to the computation of the Likelihood Ratio in the different biometric disciplines (fingerprint, face recognition, and on-line signature will be considered in this paper) is shown in figure 5. Once the laboratory has a suspect and a questioned biometric pattern, the classical approach consists in obtaining some reference patterns from

the suspect (called here biometric controls). In the case of automatic biometric systems, these reference patterns will be used for obtaining a biometric model (e.g., a minutiae pattern for fingerprint recognition or a Hidden Markov Model for signature recognition), which will be finally tested with the questioned pattern, obtaining a probability (likelihood, score) of the questioned pattern coming from that model (person). This value, which will be called here the evidence, is directly used in the classical forensic approach to submit a decision or a qualification of the decision in form of probability scale (both inadequate because of the incorrect use of thresholds as shown above).

However, in the Bayesian approach this evidence must be referenced into two different distributions, the within-source and between-source variabilities (figure 6). The within-source variability, which stands for the intravariability of the putative source, looks for the consistency of the biometric patterns of the suspect, obtaining a statistical distribution of the scores resulting from the comparison of biometric control patterns with (a) biometric model(s) obtained from some other control patterns. This distribution is usually assumed to be gaussian, which has been confirmed with experimental data as a good model of this distribution.

The between-source variability, which stands for the intervariability of the source, tries to model the probability of the test pattern coming from anyone from a reference population. Several problems arise here as how to select the correct size and member list of the population, or how to model the available between-source distribution as it will always be composed of just a part of the real possible population. Then, this distribution estimation procedure should be robust to the absence of enough population data and observe good generalization properties, that is, the results with the observed (limited) population should be as similar as possible to those with a bigger (ideally the whole) population.

A. Between-source Density Estimation

This is one of the main problems for LR computation, as we will always have less population available than the real population of candidates. Additionally, the underlying probability density function or even its type (e.g. Gaussian, Rayleigh) is unknown, so we will look for estimation procedures able to adjust to any new unknown data set. Both parametric and non-parametric methods have been described in the literature, and one of each have been proposed for this task [21][22].

Non-parametric estimation via Parzen windows [23], also known as Kernel Density Functions (KDF), is used in [21] to model the underlying distribution of between-source variability. If we perform histogram estimation, dividing the x-axis into successive bins of length h, we can estimate the probability of a sample x being in a bin for each of the bins. If N samples are available and k_N are located in a bin, the corresponding probability is approximated by the frequency ratio k_N/N, which converges to true for $N\to\infty$. The corresponding pdf value is assumed constant through the bin and is approximated by:

$$\hat{p}(x) \equiv \hat{p}(x_0) \approx \frac{1}{h}\frac{k_N}{N}, \qquad |x - x_0| \le \frac{h}{2} \tag{3}$$

where x_0 is the midpoint of the bin. Parzen [23] showed that using smooth functions ϕ (instead of step functions as in the histogram), provided $\phi(x) \ge 0$ and:

$$\int_{x} \phi(x) \cdot dx = 1 \qquad (4)$$

the resulting estimate is a legitimate pdf. Such smooth functions are known as kernel (also potential) functions or Parzen windows. Typical examples are exponentials, gaussians and so forth. In this kind of estimators, for fixed *N* the smaller the *h* the higher the variance, which will give a noisy appearance to the resulting pdf estimate. In the same sense, for a fixed *h*, the variance decreases as the number of sample points *N* tends to infinity.

On the other hand, the authors propose in [22] the use of parametric estimation via a linear combination of *M* gaussian density functions:

$$p(x) = \sum_{m=1}^{M} p_m \cdot b_m(x) \qquad (5)$$

where:

$$\sum_{m=1}^{M} p_m = 1 \qquad (6)$$

In order to obtain the best model that fits the known data, we use Maximum Likelihood estimation via the EM (Expectation Maximization) algorithm [24]. This EM algorithm guarantees that given a model λ of the underlying distribution of the data *X*, the new estimate λ' in each iteration verifies $p(X|\lambda') \geq p(X|\lambda)$. The likelihood function keeps increasing until a maximum (local or global) is reached and the EM algorithm converges. The great advantage of the algorithm is that its convergence is smooth and is not vulnerable to instabilities, which is ideal here for the between-source variability estimation as the population subset available is just a part of the whole reference population

B. Generalization

This property of generalization is highly desirable for the selected density estimation technique, as this is going to be one of our main problems because of the limited size of available databases. Our reference population will always consist in known members from the proper biometric database, but they should be good representatives of the general behavior of the whole population which will be highly dependent on the biometric pattern in use. As this can not be guaranteed for every available group in known databases, we have to check the generalization properties of every method.

Some experiments on the generalization properties of both non-parametric (Kernel Density Functions) and parametric (Maximum Likelihood via Expectation Maximization) methods for the estimation of the between-source variability with the known population have been performed. Results in the form of Tippett plots and interesting conclusions will be shown with a forensic fingerprint system in subheading VI.A (figures 9 and 10).

C. Data Scarcity for Density Estimation

Another problem related with the limited amount of data in biometric databases is the lack of enough data for reliable estimation of probability density functions, even for single gaussians as usually selected for within-source variability estimation. The problem is usually related with the time course since the test pattern was obtained in order

to generate the suspect controls. Usually single-session biometric controls are highly consistent, shown in likelihood values relative to the biometric model of the suspect highly similar, which usually gives very low variance gaussians.

Different techniques for variance restriction are proposed in the following subheadings, with the intention of retaining the intersession variability of the corresponding biometric pattern of the suspect. Experiments with different restrictions of the distribution variances have been performed both with the forensic face recognition system (figure 15) and the forensic on-line signature recognition system (figures 19 and 20), where the efficiency of those methods will be shown.

VI. Experimental Forensic Evaluation of some Biometric Systems

In order to have an overview about the application of the Bayesian approach to the different forensic biometric disciplines, some sample systems will be shown here. We will show how three different biometric systems (fingerprint, face and on-line signature) have been turned into efficient forensic systems according to the Bayesian approach. In this contribution, a detailed description of the biometric systems will not be provided, as they are out of the focus of this paper, and will be just properly referenced. We will then focus on the process of optimizing a forensic biometric system from a reference biometric system, that is, from an already working system in classical biometric tasks.

For each forensic system we will provide a short description of the system, the corresponding biometric database and the global performance of that system with that database in verification tasks through DET plots. Then, the Bayesian approach is applied with each system, detailing the selection of the selected/available reference population and showing the performance of the forensic system in the form of Tippett plots. Finally, some optimization techniques will be shown in each case, in order to have the best possible forensic system with the same biometric reference system.

A. Fingerprint Evidence Analysis

When enough quality of the fingerprint is available, they are usually used in the judicial process as absolute indicators of identity. However, with classical fingerprint analysis, based on manual comparison of minutiae after automatic selection of the list of N-best candidates, this “decision” cannot be objectively combined neither with the a priori probability of the suspect being the author of that fingerprint (based on other information about the case) nor with other types of evidence present in the case.

The automatic fingerprint recognition system we have used for these experiments have been developed in our ATVS laboratory [25][26], based on minutiae extraction and pattern comparison through dynamic programming [27]. The different image processing steps to obtain the fingerprint minutiae are shown progressively in figure 7 (details in [25]), where from the scanned fingerprint the orientation field is obtained and used for ridge-oriented spatial filtering. Then the ridges are thinned to one pixel width, and minutiae are searched and saved with the last 10 points along the corresponding ridge. For pattern comparison, firstly the two patterns are spatially aligned (translation and rotation) and then the edit distance is computed [28], which is a dynamic programming algorithm that will take into account

substitutions, insertions and deletions and look for the minimum cost edit operation to transform one minutiae pattern into the other.

In order to test the system, a subcorpus of 50 users from the MCYT database [29] has been used. Each user has 10 sample fingerprints, where the first one will be used as reference pattern. When DET curves are shown, every user acts also as impostor to the other users, so 50×9=450 correct user trials and 50×49×9=22050 impostor trials are summed up in the DET plot. Figure 8 shows the performance of our fingerprint recognition system with that 50 users subset of the MCYT database, showing an Equal Error Rate close to 3%, with excellent operating points as 0.1% false acceptance with just 9% of false rejection.

When forensic experiments are performed and summed up in Tippett curves, 5 (out of 9) samples will be used as controls (reference fingerprints from the known suspect), and the remaining 4 are used as test (unknown) fingerprints. In each Likelihood Ratio computation, the reference population is composed of the remaining 49 users of the MCYT database subset. Target curves comprise 50×4=200 Likelihood Ratios and 50×4×49=9800 for non-target.

In order to test the different approaches for between-source estimation described in heading V.A, we have performed two different experiments with the same data and reference population, the first one using Maximum Likelihood (ML) Gaussian Mixture Models (GMM) with different number of mixtures (M), and the second using Kernel Density Functions (KDF), with different bin sizes (k).

In the first experiment we want to test the capability of our LR-based forensic system for different number of gaussian mixtures. We can observe in figure 9 that the better results are obtained for a low number of mixtures (M=1 or 3), as they show excellent performance around LR=1 for non-targets and a good enough performance for targets. It is interesting to note the excellent separation between target and non-target curves with our fingerprint system, showing that can be used to provide Likelihood Ratio results to Court for any fingerprint (if quality conditions are not worse than those present in the database used in this experiment), that is, there are not no-reporting cases because of uncertainties of the method/expert. Additionally, the possibility of absolute identifications with classical methods agrees here with 90% of target users obtaining LR values greater than 10^5.

The same experiment was conducted with KDF estimation of the between-source variability, summing up the results in figure 10. As shown, results are equally good, even for low values of h as enough population data (50 people) is available.

Both experiments have been performed with 50 people in the reference population. However, in lot of cases the reference population could be even smaller, so in the following experiment we will test again both alternatives selecting a subgroup of 10 people (out of the original 50) and observe which technique can better predict with 10 people the performance of 50 people, which should be closer to the whole relevant population.

Now we will focus firstly in the results with KDF estimation (figure 11), where general good performance is observed. However, the results for target users with a reference population of 10 people are better than those with 50, which means that we are having a falsely optimistic prediction.

In figure 12 the results with (ML) estimation are shown for 10 and 50 people as reference population. The performance for non-targets is extremely similar to that observed with KDFs. However, the prediction for targets is far more realistic than that with KDF as both (10 and 50) target curves show similar performance. This relative advantage of ML over KDF for this task is based in the generalization properties of both techniques. With KDF in order to obtain a good estimate of the actual histogram a low value of h is needed, but when this histogram comes from a small amount of data (10 people) the resulting pdf is excessively adjusted to these data and does not represent with enough accuracy the underlying distribution from a higher number of reference persons. On the other hand, the ML estimate of the pdf shows better generalization properties, as it is less accurate to the histogram for the 10 people case, but it is a better model of a greater unknown population as is shown here for 50 people.

B. Face Evidence Analysis

The automatic analysis of face evidence in court [30] is not as usual as other characteristics as fingerprint, but the underlying principles shown here for an automatic face recognition system could be used for other face recognition systems, so there are a lack of face databases under forensic conditions. However, as biometric systems are becoming more and more available, we could think in possible crimes committed after a false acceptance of a non-user (impostor) accessing to a restricted area through a face-verification based access control system. Later we should show in Court if the face of the suspect corresponds or not to the photo of the person who accessed cheating the system. Of course this comparison could be also performed subjectively by a human, but that person cannot obtain an objective Likelihood Ratio from his observation to relate with other parts of the process (a priori probabilities or other evidences, even of different kind).

The face recognition system that has been used has been developed in our ATVS laboratory and is based in feature extraction through a combination of eigenfaces and fisherfaces [31][26]. Pattern comparison will be performed through the normalized scalar product of feature vectors. Before dimensionality reduction, geometric normalization is performed through face warping. A sample face preprocessing is shown in figure 13 (details in [31]), where the original face, the warped face with the reference points, and the final selection with an elliptic patch are shown. The dimensionality of that elliptic patch (64×64) is first reduced to 250 through eigenfaces, and then down to 180 through fisherfaces.

The different experiments that have been performed have used the configuration 1 of Lausanne Protocol described in [32] for the XM2FDB database (a subcorpus of XM2VTS), with 200 users and 95 different impostors with two faces per session and 4 photo sessions. This database contains data typically used for verification of identity, and not on data available in forensic conditions. However, it is useful to show the LR computing process. For DET assessed experiments, 50 users out of 200 have been selected, with 3 sample faces used for training and the remaining 5 as test, so 50×5=250 user trials and 50×49×5=12250 impostor trials are shown in the DET curve. In figure 14, the performance of our face verification system in those conditions is shown, where an Equal Error Rate close to 2% is obtained, with very interesting operating points for very low false acceptance rates. This performance is even better than in the case of the reported fingerprint experiment, due to the controlled conditions of the face database in use.

For the forensic evaluation, assessed with Tippett plots, the whole XM2FDB database has been used. Three samples are used to model each suspect, three are used as suspect face controls, and the remaining two are used as test (unknown) faces. The reference population for each Likelihood Ratio computation will be composed of the 199 remaining users. Each one of the 8 sample faces of the 95 impostors subset will be used as non-targets faces. Then, 200×2=400 target Likelihood Ratios are obtained and 200×8×95=152000 for non-targets.

In the following forensic experiment, we will firstly show the performance of our automatic face recognition system with the users and reference populations of the database described above. One of the usual problems of available databases is the small amount of control material available, in this case different unquestioned faces from the same suspect. This will lead to inconsistencies in the estimation of the within-source variability of the suspect, which could lead (as in the following experiment) to target evidences with scores clearly above the scores obtained by the population that incorrectly give low or very-low likelihood ratios, as described in heading V.C (solid line in figure 15, where 10% of targets obtain LR values smaller than one).

Two different, but similar, approaches have been used in this case to avoid this problem, both of them related with the establishment of a minimum variance in the pdf (gaussian) estimate of the within-source variability. The first approach consisted in computing the standard deviation of all within-source variabilities in the database, and the mean of those standard deviations is used as minimum for every user, where the underlying idea is that the multisession variability for all users should be similar or at least have a common minimum.

The second approach was to compute this minimum from the mean of the standard deviations of the between-source variabilities assumed to be single-gaussian (this assumption is just used for this minimum computation; later on, the between-source variability is estimated as usual with Kernel Density Functions or Gaussian Mixture Models). In this case, the assumption is that the multisession variability of the user can be predicted from the variability of the test face relative to the reference population.

The likelihood ratios obtained with those three systems are summed up in the Tippett plots of figure 15. As shown, when no restrictions in the variance are imposed (solid line), about 10% of targets are obtaining likelihood ratios smaller than one. However, when variance restrictions are applied, the performance of target and non-target curves is excellent with both proposed alternatives. Looking with higher detail, a slightly better performance is obtained with the second alternative, $\sigma_{W,min}$= mean($\sigma_{B,i}$) with dotted line, where the first alternative, $\sigma_{W,min}$= mean($\sigma_{W,i}$) with dashed line, performs slightly worse for targets in the area around LR=1.

C. On-line Signature Evidence Analysis

Similar to the case of automatic face recognition systems, up to date is weird to have a case in Court involved with on-line signature analysis (the signature is captured with a especial pen and board, obtaining instantaneous position, pressure, azimuth and altitude as shown in figure 16). However, a few years ago, some companies use dynamic signature for the identity check of their bank customers. A case of ID theft involving dynamic signature could be a realistic case for a court. We have used this technology in these experiments because it is an available technology in our laboratory, that could be used e.g. for validating on-line transactions. We are actually developing in our laboratory an off-line signature recognition system (a written signature in any document is scanned and converted

into a gray-level image) [33]. Once we had ready our off-line signature recognition system, the approach will be exactly the one shown below, and this is an extremely frequent case in Courts. Anyway, as biometric systems are more and more present in society, it would not be strange to have soon cases where fraud have been committed with an on-line signature system, and exactly the below shown approach should be used in Court.

The on-line signature recognition system [34] have also been developed in our laboratory [35], taking advantage of the dynamic nature of the signature process. Instantaneous information (x-position, y-position, pressure, azimuth, altitude) is sampled every 10 ms., and from those raw parameters some derived ones are obtained (instantaneous trajectory angle, log-curvature radius, instantaneous displacement). From that 8 coefficient vector, velocity (delta) and acceleration (delta-delta) are computed, deriving the final 24 parameter vector. Dynamic modeling of the signature process is obtained by means of Left-to-Right Hidden Markov Models (HMM) [35], a powerful tool to solve pattern recognition problems when temporal or context information is present. Figure 16 shows pen azimuth and altitude, and the dynamic vector flow characterizing the on-line acquisition of writing. Figure 17 shows a sample signature (lower right part) and the corresponding raw time-dependent parameters.

We have again selected for the experiments another subset of the MCYT database [29] containing 50 signers with 15 signatures per signer. Additionally, every signer is deliberately mimicked by three different signers in 5 different signatures per impostor signer. Different types of tests have been performed both with casual impostors (any signature in the database) and skilled impostors (signers mimicking another signer signature). Signers are modeled with 6 sample signatures, and the remaining 9 will be used for testing, so 50×9=450 user trials are used in DET assessed experiments, with 50×49×9=22050 casual impostor trials or 50×3×5=750 skilled impostor trials in each case. When the forensic system is tested, the 9 samples available excluding training are divided in 5 for controls and 4 for test (unknown) signatures, which results in 50×4=200 Likelihood Ratios for targets and 50×49×4=9800 for non-targets.

As just said, we have just available in that database a small amount (fifteen) of skilled forgeries of every signer trying to imitate a known signature. In that case, we cannot use reference populations because of lack of data in order to obtain the between-source variability of each signature. However, we have a big amount (all those included in the database) of different signatures which could be used as casual impostors. Of course, it could be said that the performance is not going to be comparable when the system is tested with skilled (imitating a signature) or casual (using his own signature) impostors, which would be completely true if a human expert performs the comparisons. However, as the automatic system does not rely in the resulting image (the signature) but in the time-dependent sequence of positions and pressure, a well designed system could be robust enough both if casual or skilled impostors are used. In figure 18 we show with a solid line the performance of the system, calibrated with casual impostors, that is going to be used in the following forensic experiments, which shows an Equal Error Rate (EER) slightly lower than two percent. We also show for comparisons two optimizations that have been obtained in our laboratory, both for casual impostors (dashed line) with EER about unity, and skilled impostors (dotted line) with EER close to two percent. We want to note then that if we had enough skilled impostors for each signer in the database, we could use

the skilled-impostor-based system, but testing the forensic system with a casual-based impostor approach will show a highly similar performance to that of skilled impostors if enough impostors per signature were available.

The forensic on-line signature experiment that has been designed shows again the same problem of lack of data for the proper estimation of the within-variability of the suspect signature. Then, we have compared the raw results, with solid line in figure 19, with those obtained imposing a variance restriction. In that case, we have modeled both within- and between-source variabilities with single Gaussians, and then we will apply the same variance restrictions to both distributions. As shown in the Tippett plots of figure 19, about 20% of target users obtained likelihood ratios smaller than one (solid line), which has been solved imposing that variance restriction (dashed line).

But the resulting system, even perfectly separating the target and non-target curves, does not obtain very high values for target users (LR are always below 500). This could be the best case in other biometric or non biometric technologies where better separation of the curves could not be possible because of no better technology was available, which would be shown in DET curves with much higher Equal Error Rates. However, in that case this can be improved using a better estimation of the between-source variability (remember it was obtained in this experiment with a single Gaussian with variance restriction), which is going to be performed here both with Kernel Density Functions and with Gaussian Mixture Models. The results are shown in figure 20 relative to the best case of last experiment (single Gaussian with variance restriction). In this case the variance restriction is just applied to the within-source distribution, and in both cases (KDF and GMM) the results improve significantly. From the curves, it seems that KDF is performing extraordinarily better than GMM (look both target curves), but from previous experiments it seems that the KDF estimation is being too optimistic from the reduced population available, as happened in the experiments reported in figure 11 (section VI.A). However, in that case we cannot check that hypothesis as we have used the full size of the population for that experiment, so the performance with bigger populations cannot be obtained and be used as reference.

VII. CONCLUSION

In this contribution, the role of the forensic scientist for the analysis of biometric evidences has been clarified, providing a unifying approach for the analysis of any type of evidence and their combination in a single objective number to be provided to Court through the use of Likelihood Ratios. The key point of this contribution is that we provide a method for adapting any existing biometric system into a forensic system according to the Bayesian approach. Additionally, the performance of the forensic system providing its results in the form of likelihood ratios cannot be assessed as classical identification systems through ROC or DET plots, and needs a different kind of specific representation such as Tippett plots. We have strengthen the need for reliable estimation techniques with good generalization properties both for the between-and within-source variability, and some simple but effective procedures as the use of variance restrictions have been shown to avoid singularities in the computation of Likelihood Ratios. Finally, detailed examples have been provided using fingerprint, face and on-line signature recognition systems, where the likelihood ratio range, both for targets and non-targets, and the forensic adequacy of these video-based techniques have been shown.

ACKNOWLEDGMENT

Authors wish to thank all people from both Speech and Signal Processing Group (ATVS) at Universidad Politecnica de Madrid for their excellent work, and the Video and Audio Laboratory of Spanish Police Force "Guardia Civil" for extensive testing, suggestions, and close collaboration. J. F-R and D. R-C also thank Consejería de Educación de la Comunidad de Madrid and Fondo Social Europeo for supporting their doctoral research.

REFERENCES

[1] A. Martin, et al., "The DET curve in assessment of detection task performance", Proc. EuroSpeech'97, pp. 1895-1898, Rhodes (Greece), 1997.

[2] I.W. Evett, "Towards a Uniform Framework for Reporting Opinions in Forensic Science Casework", Science & Justice 1998: 38(3), pp. 198-202.

[3] C. Champod, "Overview and Meaning of Identification", Encyclopedia of Forensic Sciences, pp. 1077-1084, Academic Press, 2000.

[4] C.G.C. Aitken and F. Taroni, "Statistics and the Evaluation of Evidences for Forensic Scientists 2nd. Ed.", John Wiley and Sons, 2004.

[5] I.W. Evett and J.S. Buckleton, "Statistical Analysis of STR (short tandem repeat) data", Advances in Forensic Haemogenetics, A. Carracedo, B. Brickmann, and W. Bär, Editors. Springer-Verlag: Heidelberg, pp. 79-86, 1996.

[6] M. J. Sanks.et al., "Lessons from the Law's Formative Encounters with Forensic Identification Science", Hastings Law Journal, 49 (4), 1069-1141, 1998.

[7] Epstein, R., "Fingerprints meet Daubert: The Myth of Fingerprint "Science" is Revealed", Southern California Law Review, 75, 605-655, 2002.

[8] F. Taroni, C.G.G. Aitken, "Forensic Science at Trial", Jurimetrics Journal 37, 327-337, 1997.

[9] A.P.A. Broeders, "Forensic Speech and Audio Analysis: the State of the Art in 2000 AD", Proc. of SEAF-2000 (1st National Conference of the Spanish Forensic Acoustics Society), Ed. J. Ortega-García, Madrid (Spain), 2000.

[10] C. Champod and D. Meuwly, "The Inference of Identity in Forensic Speaker Recognition", Speech Communication, vol. 31, pp. 193-203, June 2000.

[11] Redmayne, M., "Appeal to Reason", The Modern Law Review, 65, 19-35, 2002.

[12] D. Meuwly, "Current Discussions of the ENFSI-WG About the Use of the Bayesian Approach for the Interpretation of Evidence", Meeting of the Speech and Audio Group of ENFSI –European Network of Forensic Science Institutes-, Paris (France), 2001.

[13] Cook, R., I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert, "A Hierarchy of Propositions: Deciding Which Level to Address in Casework". Science and Justice, 1998. 38(4): 231-240.

[14] F. Taroni et al., "De Finetti's Subjectivism, the Assessment of Probabilities and the Evaluation of Evidence: A Commentary for Forensic Scientists". Science and Justice, 2001. 41(3): 145-150.

[15] Meuwly, D., “Reconnaissance de locuteurs en sciences forensiques - l'apport d'une approche automatique”, PhD: Institut de Police Scientifique et de Criminologie; Université de Lausanne.

[16] Champod, C., I.W. Evett, and B. Kuchler, “Earmarks as Evidence: A Critical Review”. Journal of Forensic Sciences, 2001. 46(6): 1275-1284.

[17] Champod, C., “Identification / Individualization: Overview and Meaning of ID.” in Encyclopedia of Forensic Science, J. Siegel, P. Saukko, and G. Knupfer, Editors. 2000, Academic Press: London, 1077-1083.

[18] Champod, C. and I.W. Evett, “A Probabilistic Approach to Fingerprint Evidence. Journal of Forensic Identification”, 2001. 51(2): 101-122.

[19] C.F. Tippett et al., “The evidential value of the comparison of paint flakes from sources other than vehicles”, Journal of the Forensic Science Society, vol. 8, pp. 61-65, 1968.

[20] K.A.J. Walsh et al., “A practical example of glass interpretation”, Science and Justice 1996; 36: 213-218.

[21] D. Meuwly, A. Goode, A. Drygajlo, J. Gonzalez-Rodriguez and J. Lucena-Molina, “Forensic Speaker Recognition based on a Bayesian Framework and Gaussian Mixture Modeling”, Forensic Science International, Vol. 136 Suppl. 1, September 2003, pp.364.

[22] J. Gonzalez-Rodriguez, J. Fierrez-Aguilar and J. Ortega-Garcia, “Forensic Identification Reporting Using Automatic Speaker Recognition Systems”, Proc. IEEE 2003 International Conference on Acoustics, Speech and Signal Processing (ICASSP’03), vol. II, pp. 93-96, Hong-Kong (China), 2003.

[23] E. Parzen, “On the estimation of a probability density function and mode”, Ann. Math. Stat., Vol. 33, pp. 1065-1072, 1962.

[24] A.P. Dempster et al., “Maximum Likelihood from incomplete data via the EM algorithm”, J. Royal Statistical Society, Vol. 39(1), pp. 1-38, 1977.

[25] D. Simon, J. Ortega-Garcia, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, “Image Quality and Position Variability Assessment in Minutiae-Based Fingerprint Verification”, IEE Proc. Vision, Image and Signal Processing, vol. 150, n. 6, pp. 402-408, December 2003.

[26] J. Ortega-Garcia, J. Gonzalez-Rodriguez et al. “From Biometrics Technology to Applications Regarding Face, Voice, Signature and Fingerprint Recognition Systems”, Chapter 12 in Biometric Solutions For Authentication in an E-World, D. Zhang (ed.), pp. 289-337, Kluwer Academic Publishers, 2002.

[27] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, “An Identity-Authentication System Using Fingerprints”, Proceedings of the IEEE, Vol. 85, No. 9, pp. 1365-1388, September 1997.

[28] F.J. Damerau, “A technique for computer detection and correction of spelling errors”, *Commun. ACM*, Vol. 7(3), pp. 171-176, 1964.

[29] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez-Rodriguez et al., “MCYT Baseline Corpus: A Bilmodal Biometric Database”, IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet, vol. 150, n. 6, pp. 395-401, December 2003.

[30] Peacock, C., A. Goode, and A. Brett, “Automatic Forensic Face Recognition from Digital Images”. Science and Justice, 2004. 44(1): 29-34.

[31] S. Cruz-Llanas, J. Fierrez-Aguilar, J. Ortega-Garcia and J. Gonzalez-Rodriguez, "A comparative Evaluation of Global Representation-Based Schemes for Face Verification", in Proc. of IEEE International Conference on Image Processing, vol. 3, pp. 905-908, Barcelona (Spain) 2003.

[32] J. Luettin, G. Maître. "Evaluation Protocol for the XM2FDB Database", IDIAP-COM 98-05, October 1998.

[33] J. Fierrez-Aguilar, N. Alonso-Hermira, G. Moreno-Marquez and J. Ortega-Garcia, "An Off-Line Signature Verification System Based on Fusion of Global and Local Information", in Proc. European Conference on Computer Vision, Workshop on Biometric Authentication, Springer LNCS, vol. 3087, Prague (Czeck Republic), 2004.

[34] R. Plamondon and G. Lorette, "Automatic Signature Verification and Writer Identification - The State Of The Art", Pattern Recognition, vol. 22, n. 2, pp. 107-131, 1989.

[35] J. Ortega-Garcia, J. Fierrez-Aguilar, J. Martin-Rello and J. Gonzalez-Rodriguez, "Complete Signal Modeling and Score Normalization for Function Based Dynamic Signature Verification", in Proc. International Conference on Audio- and Video-Based Person Authentication (AVBPA), Springer LNCS, vol. 2688, pp. 658-667, Guilferd (UK), 2003.

Authors Short CV

Joaquín González-Rodríguez, received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación), in 1994; and the Ph.D. degree "cum laude" also in electrical engineering (Doctor Ingeniero de Telecomunicación), in 1999, both from Univ. Politécnica de Madrid, Spain.

He is an Associate Professor from January 2002, and from October 2004 he is at the Computer Science Department, Universidad Autónoma de Madrid, Spain. From 1995 to 2001 he was Assistant Professor also at Universidad Politécnica de Madrid. His research interests are focused on signal processing, biometrics and forensics: speaker recognition, forensic biometrics, data fusion in biometrics (face, fingerprint, signature and speaker recognition), robustness in speech/speaker recognition and speech enhancement (auditory criteria, microphone arrays, binaural modeling). He has published diverse international contributions, including book chapters, refereed journal and conference papers.

Dr. González-Rodríguez is an invited member of ENFSI (European Network of Forensic Science Institutes), giving an invited tutorial in the Weisbaden Meeting (May'02) to all Forensic laboratories across Europe. He has acted as member of the Scientific Committees of ICSLP'02 (International Conference on Spoken Language Processing), EuroSpeech'03 and Odyssey'04-The Speaker Recognition Workshop to be held in Toledo, Spain, in 2004, where he acted also as Vice-chairman.

Julián Fiérrez-Aguilar, received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación), in 2001, from Universidad Politécnica de Madrid, Spain.

Since 2002, he is a scholar of the CAM program at the Computer Science Department, Universidad Autónoma de Madrid, Spain, where he is currently working towards the Ph.D. degree. His research interests are focused on signal and image processing, pattern recognition and multimodal biometrics. He has published contributions in international conferences and refereed journals and has been the recipient of the Best Poster Award at International Conference on AVBPA 2003.

Daniel Ramos-Castro, received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación), in 2001, from Universidad Politécnica de Madrid, Spain.

He is Associate Professor in Since 2003, he is a scholar of the CAM program at the Computer Science Department, Universidad Autónoma de Madrid, Spain, where he is currently working towards the Ph.D. degree. His research interests are focused on speaker recognition and forensic science. He has published several contributions in international conferences and collaborated in book chapters.

Javier Ortega-García, received the M.S. degree in electrical engineering (Ingeniero de Telecomunicación), in 1989; and the Ph.D. degree "cum laude" also in electrical engineering (Doctor Ingeniero de Telecomunicación), in 1996, both from Univ. Politécnica de Madrid, Spain.

He is associate professor at the Computer Science Department, Universidad Autónoma de Madrid, Spain. From 1999 to 2003 he was Associate Professor at the Audio-Visual and Communications Engineering Department, Universidad Politécnica de Madrid, Spain. From 1992 to 1999 he was Assistant Professor also at Universidad Politécnica de Madrid. His research interests are focused on biometrics signal processing: speaker recognition, face recognition, fingerprint recognition, on-line signature verification, data fusion and multimodality in biometrics. His interests also span to forensic biometrics, acoustic signal processing, signal enhancement, and microphone arrays. He has published diverse international contributions, including book chapters, refereed journal and conference papers.

Dr. Ortega-García has chaired several sessions in international conferences. He has participated in some scientific and technical committees, as in EuroSpeech'95 (where he was also Technical Secretary), EuroSpeech'01, or Odyssey'01-The Speaker Recognition Workshop. He has been appointed as General Chair at Odyssey'04-The Speaker Recognition Workshop to be held in Toledo, Spain, in 2004.

FIGURES

Figure sizes have been magnified to ease review.
Actual figures size fits in every column of a two column page format

Full paper length in two column format: 11/12 pages

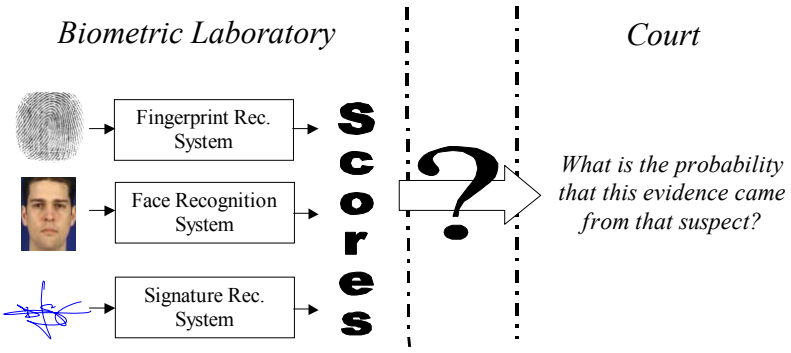

Fig. 1. The problem of biometric score submission to Court.

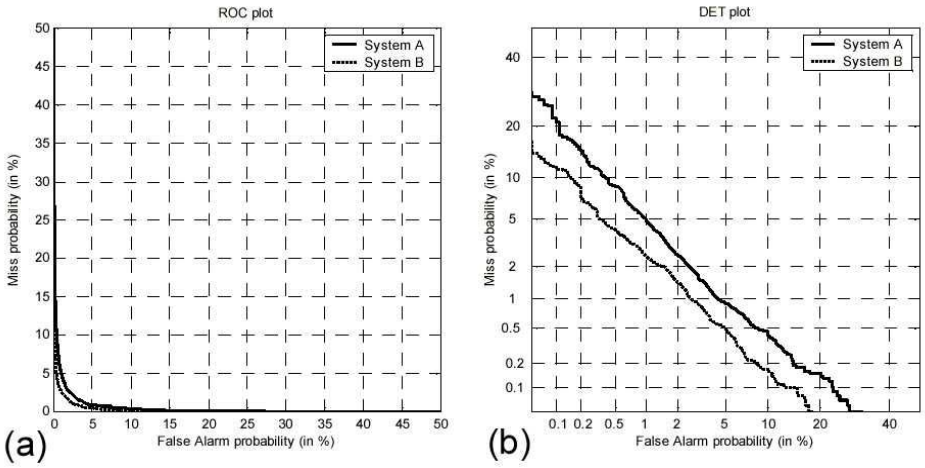

Fig. 2. Comparison of system performance in two different conditions with (a) ROC detection curve, and (b) DET plots

TABLE I
LR SCALE WITH CORRESPONDING LINGUISTIC QUALIFIERS

Likelihood Ratio (LR)	Verbal Equivalent
>1 to 10	Limited evidence to support
10 to 100	Moderate evidence to support
100 to 1000	Moderately strong evidence to support
1000 to 10000	Strong evidence to support
>10000	Very strong evidence to support

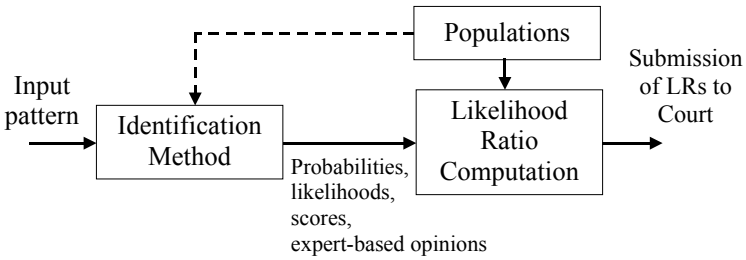

Fig. 3. System architecture for LR Computation in the Bayesian framework.

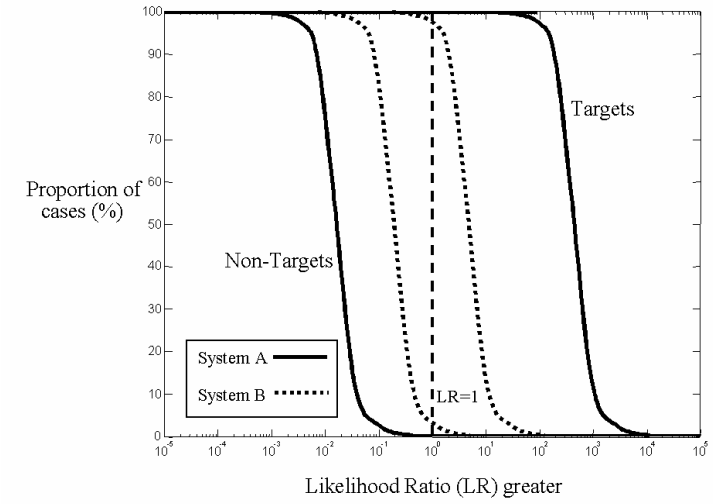

Fig. 4. Example of Tippet curves for two competing systems.

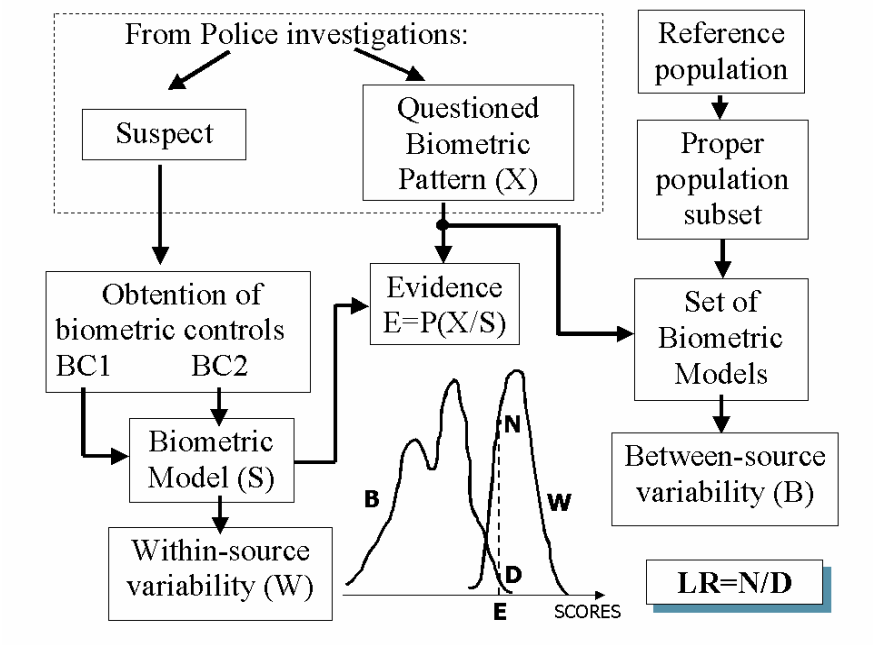

Fig. 5. LR computation in Forensic Analysis of Biometric Evidences.

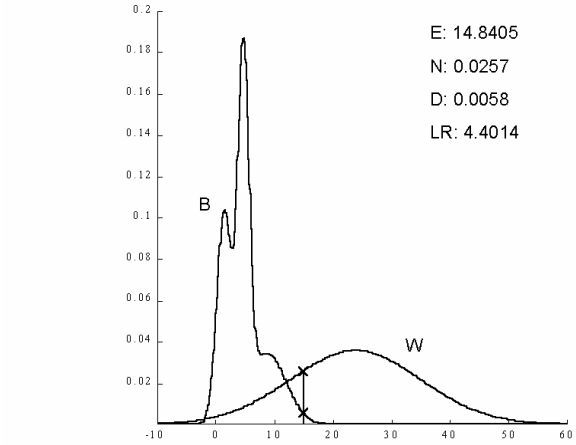

Fig. 6. LR Computation of an user Likelihood Ratio.

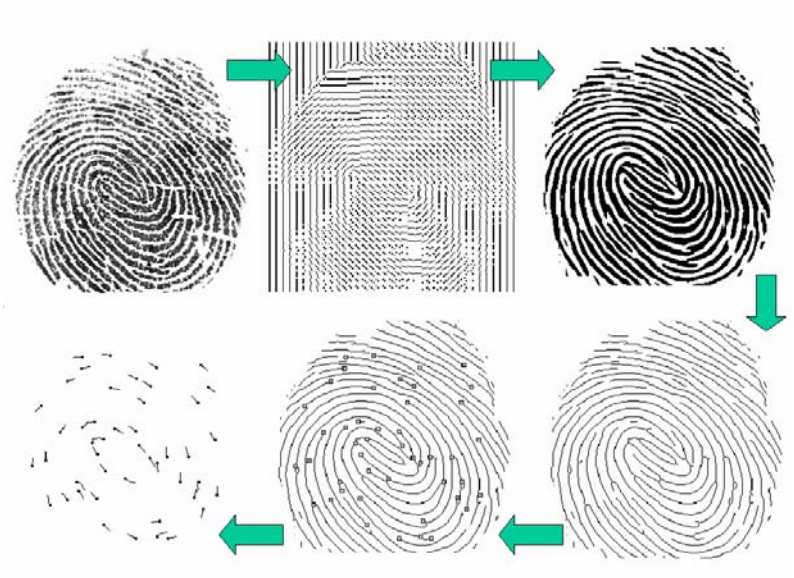

Fig. 7. Different image processing stages in the minutiae computation process from a fingerprint.

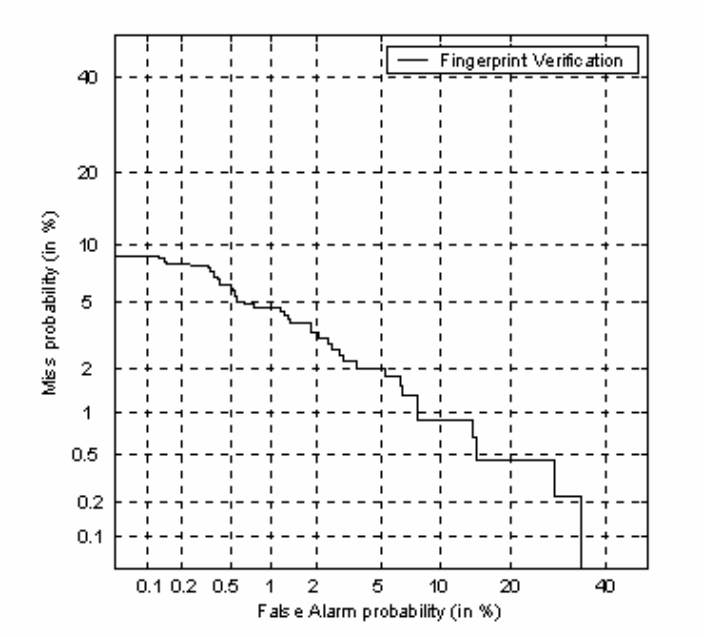

Fig. 8. DET plot of the reference fingerprint verification system.

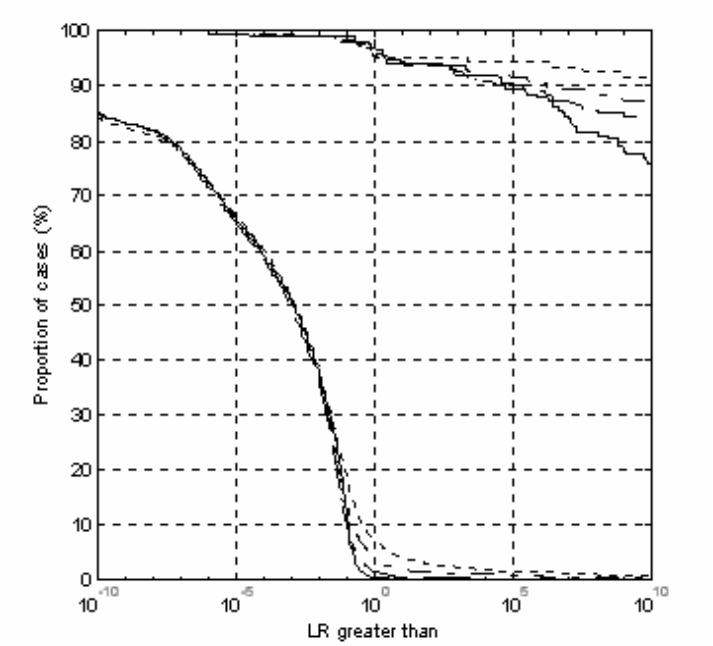

Fig. 9. Tippet plots for the forensic fingerprint system with between-source estimation via Maximum Likelihood with M gaussian mixtures; M=1 (solid), M=3 (dashed), M=10 (dash-dot) and M=30 (dotted).

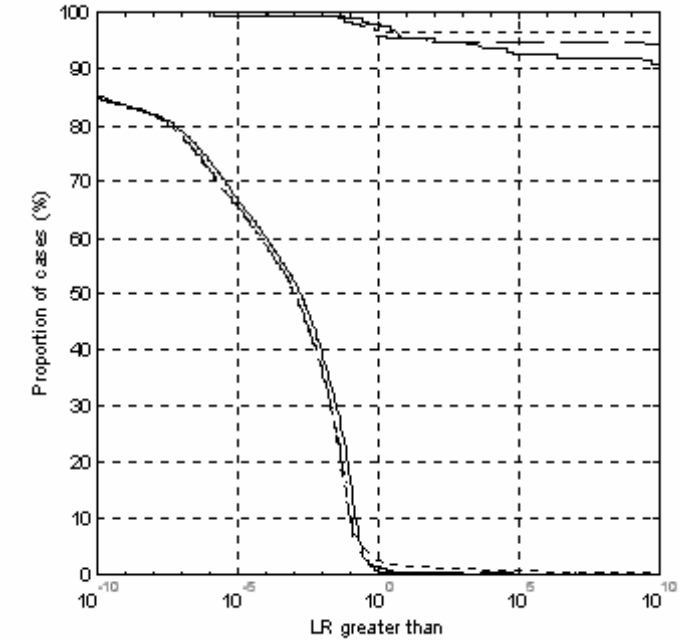

Fig. 10. Tippet plots for the forensic fingerprint system with between-source estimation via Kernel Density Functions with bin size h=10 (solid), h=3 (dashed), h=1 (dotted).

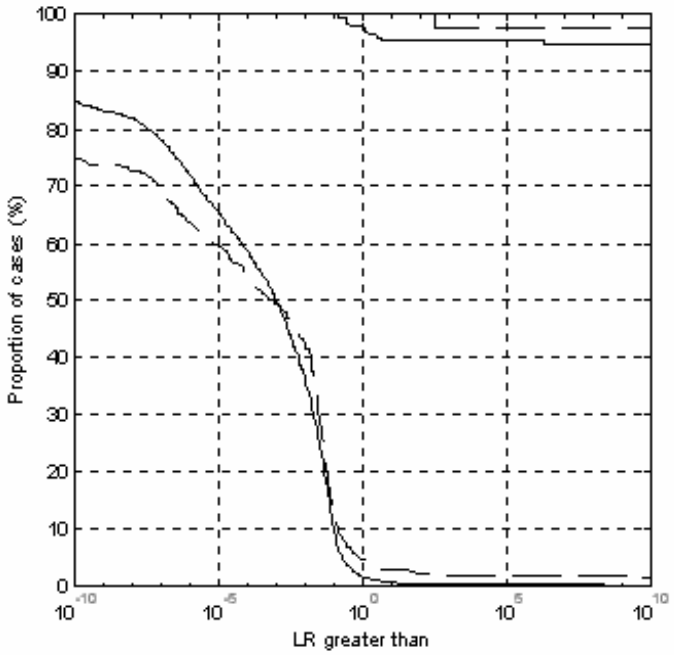

Fig. 11. Analysis of the generalization abilities via Kernel Density Functions estimation (h=3) for different sizes (L) of the population; L=50 (solid), L=10 (dashed).

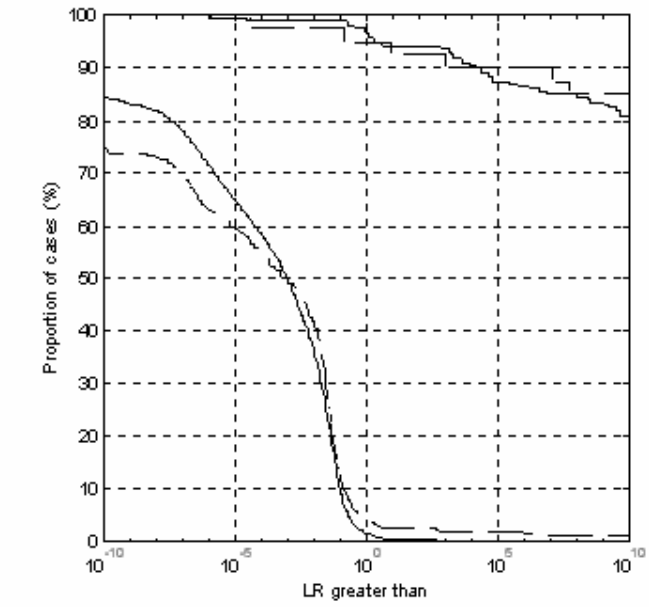

Fig. 12. Analysis of the generalization abilities via Maximum Likelihood estimation of a gaussian mixture (M=2) for different sizes (L) of the population; L=50 (solid), L=10 (dashed).

Fig. 13. Geometric normalization previous to dimensionality reduction.

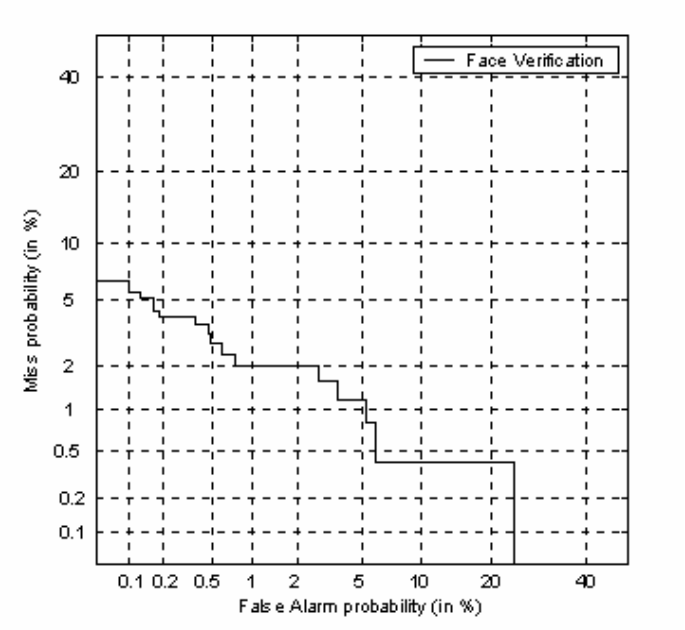

Fig. 14. DET plot of the reference face verification system.

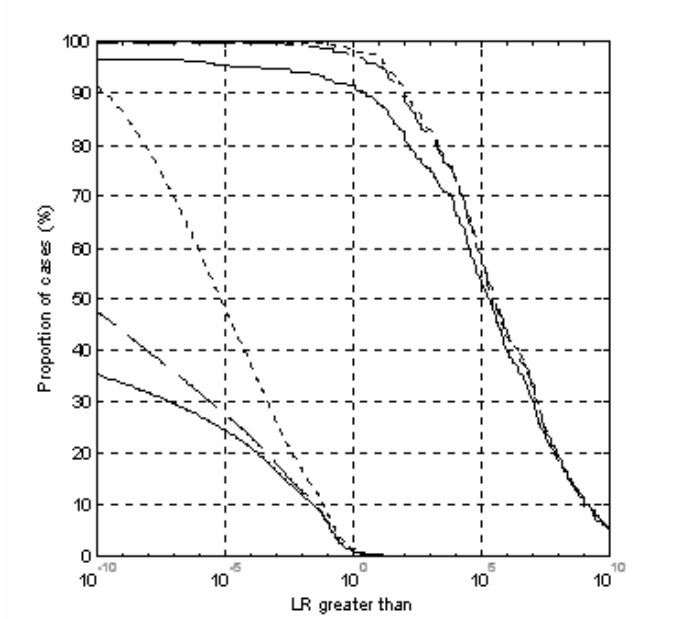

Fig. 15. Tippet plots for the forensic face recognition system with different variance restrictions in within-source-estimation; no-restriction (solid), $\sigma_{W,min}$= mean($\sigma_{W,i}$) (dashed), $\sigma_{W,min}$= mean($\sigma_{B,i}$) (dotted).

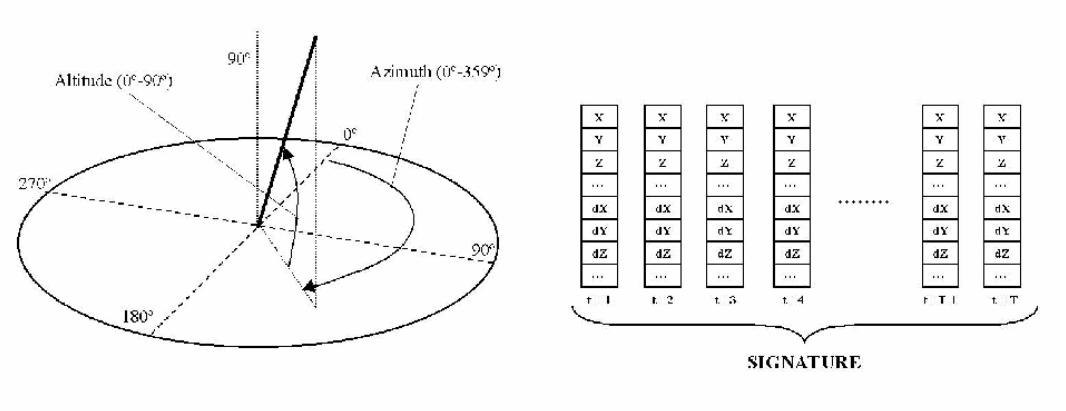

Fig. 16. Pen azimuth and altitude, and characteristic vector time series representing a signature.

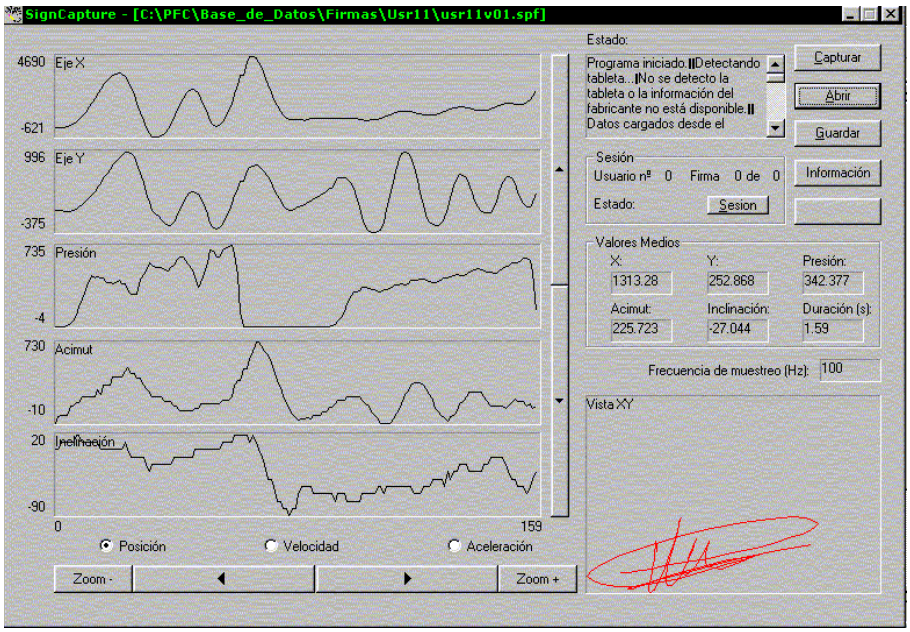

Fig. 17. Sample signature and raw time dependent coefficients (x-position, y-position, pressure, azimut and altitude).

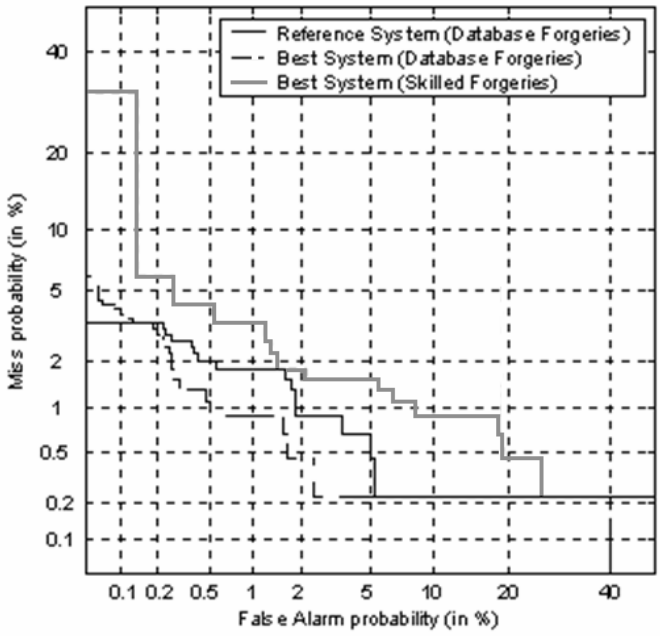

Fig. 18. DET plot of the reference on-line signature verification system (used in the following forensic experiments) with random impostors from the database (solid, black), our best system in the same condition (dashed), and our best system with skilled impostors forging the true signature (solid, grey).

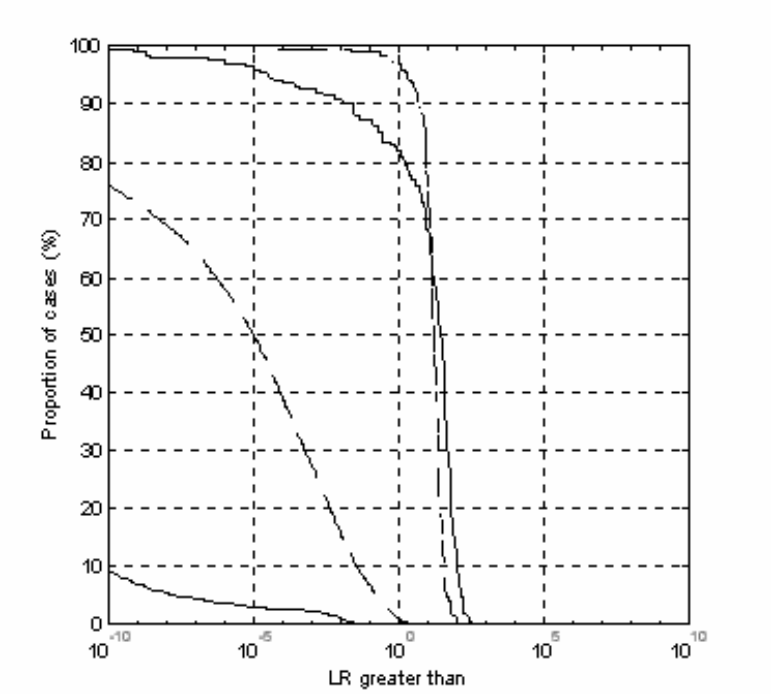

Fig. 19. Tippet plots for the forensic signature system with single gaussian between- and within-source estimation; no restriction (solid), $\sigma_{W,min}= \sigma_{B,min}= \text{mean}(\sigma_{B,i})$ (dotted).

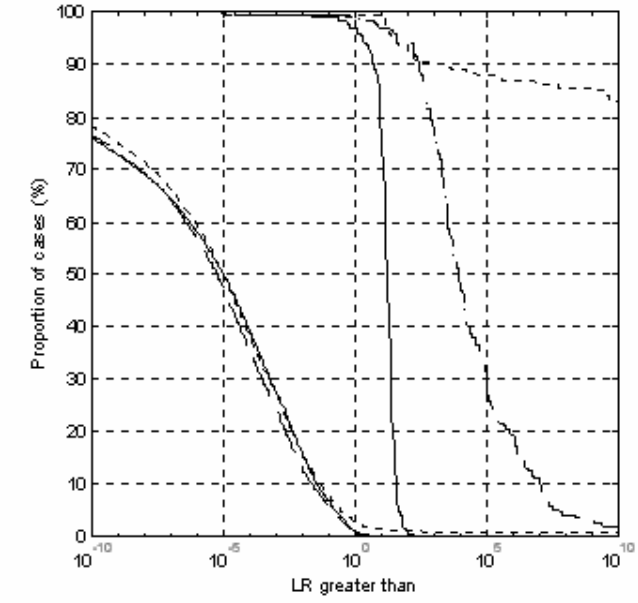

Fig. 20. Tippet plots for the forensic signature system; single gaussian between- and within-source estimation with $\sigma_{W,min}= \sigma_{B,min}= \text{mean}(\sigma_{B,i})$ (solid), between-source estimation via Maximum Likelihood with M=3 (dashed), and between-source estimation via Kernel Density Functions with h=3 (dotted).