J Forensic Sci, November 2006, Vol. 51, No. 6
doi:10.1111/j.1556-4029.2006.00253.x
Available online at: www.blackwell-synergy.com

Ivo B. Alberink,[1] Ph.D.; Arnout C. C. Ruifrok,[1] Ph.D.; and Hartmut Kieckhoefer,[2] M.Sc.

Interoperator Test for Anatomical Annotation of Earprints

ABSTRACT: As part of the Forensic Ear Identification (FearID) research project, which aims to obtain estimators for the strength of evidence of earmarks found on crime scenes, a large database of earprints (over 1200 donors) has been collected. Starting from a knowledge-based approach where experts add anatomical annotations of minutiae and landmarks present in prints, comparison of pairs of prints is done using the method of Vector Template Matching (VTM). As the annotation process is subjective, a validation experiment was performed to study its stability. Comparing prints on the basis of VTM, it appears that there are interoperator effects, individual operators yielding significantly more consistent results when annotating prints than different operators. The operators being well trained and educated, the observed variation on both clicking frequency and choice of annotation points suggests that implementation of the above is not the best way to go about objectifying earprint comparison. Processes like the above are relevant for any forensic science dealing with identification (e.g., of glass, tool marks, fibers, faces, fingers, handwriting, speakers) where manual (nonautomated) processes play a role. In these cases, results may be operator dependent and the dependencies need to be studied.

KEYWORDS: forensic science, earprint identification, interoperator effects, anatomical annotation, template matching, classification

In recent years, expert court testimony on earmarks found at crime scenes relied on the assessment of expert witnesses on the biological uniqueness of characteristics found on the marks, cf. (1). Examples of these are overall shape and size, and details such as Darwinian tubercles, creases, moles, piercings, or scars. Because of a relative lack of scientific basis, and the subjective nature of the assessments by the experts, the reliability of earmark identification has been under fire. A good review article on the (lack of) scientific research up to 1999 with respect to earmark identification can be found in (2). This has, e.g., resulted in rejection of earmark evidence in the State versus Kunze case in the United States, see (3), and the calling of a retrial in the Regina versus Mark Dallagher case in the United Kingdom; see (4) and (5).

To solidify the scientific basis for earprint/earmark identification, the EU-financed Forensic Ear Identification (FearID) project was started in nine institutes, including police academies, universities, The Netherlands Forensic Institute, and two commercial partners, over Italy, The Netherlands, and the United Kingdom. The project aims at obtaining estimators for the strength of evidence of earmarks found on crime scenes and the development of methods to match and classify earprints. The ultimate goal of this is verification (one to one matching) and individualization (one to many matching). In the three countries, a training database has been gathered of 1227 donors, donating three left and three right earprints each. Next to this, an operational system was developed allowing for scanning and storing of earprints and that may process them in different ways.

An example of this processing is that a user, from here on *operator*, adds a polyline to the digitized earprint image following the imprint of the ear. From this a connected structure is determined that is supposed to represent the imprint, and which is referred to as a *superstructure*. An example is given in Fig. 1. On the basis of this superstructure, further analysis is performed using various image-processing techniques. Results of this analysis can be found in (6).

Another example of manual annotation is that of anthropological experts adding anatomical annotations of so-called minutiae and landmarks present in prints. On the basis of this knowledge-based approach, comparisons of pairs of prints are then made. The anthropological workpackage of the FearID project decided on a set of possibly present minutiae and landmarks, laid down in De Conti et al. (7), and the prints gathered in the database were annotated accordingly. In this way, an effort was made to objectify (necessarily subjective) expert witness opinions, and lay a basis for classification of earprints on the basis of biological knowledge (as opposed to information acquired by (semi)automatic image processing).

As both clicking of initial axes and anatomical annotation of earprints are subjective processes, the question arises how much the output of the process depends on the operator performing the manual annotating. An experiment was set up to test for these interoperator effects. For polyline clicking, unpublished data showed no significant operator effects. The paper at hand studies interoperator effects for anatomical annotation as in Fig. 2.

The structure of the report is as follows: in the next section a description of the experiment is given. Here, we describe in turn

- the data collection;
- the process of anatomical annotation;
- the performance measure used (equal error rate);
- the method used to compare annotated prints (vector template matching); and
- the statistical technique to analyze the results (binary logistic regression).

In the "Results" section, first annotation frequencies over different operators are compared, after which the analysis of

[1]Department of Digital Technology, Netherlands Forensic Institute, PO Box 24044, 2490 AA, Den Haag, The Netherlands.
[2]Department of Mechanical Engineering, University of Huddersfield, Queensgate, Huddersfield, HD1 3DH, U.K.

Received 13 July 2005; and in revised form 29 Mar. 2006; accepted 23 April 2006; published 30 Oct. 2006.

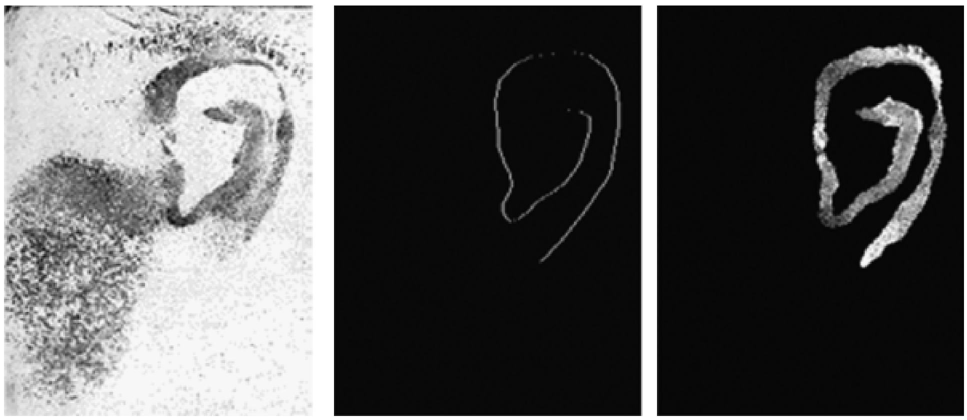

FIG. 1—*Example: original print, clicked polyline, and calculated superstructure.*

the data with respect to interoperator effects is presented. Finally, conclusions are drawn on implications of the results.

Description of the Experiment

Data Collection

To test for differences in annotation behavior between operators, a collection of 135 earprints was presented to three operators from the participating countries (Italy, The Netherlands and the United Kingdom). In each of the countries, the operator would annotate the whole collection, leading to a total number of $3 \times 135 = 405$ anatomically annotated prints.

To allow for testing for eventual effects of factors like donor country, number of donors per country, number of prints per donor (i.e., safeguard the representativeness of the sample), the collection was built up as follows: it consisted of 45 different prints, always three different prints from 15 donors in total. The prints were all repeated for three times, that is, each operator annotated the same *identical* print for three times. This repetition of identical prints was undertaken to test whether different operators are internally consistent compared with consistency among different operators.

The 15 donors, five donors from each of the participating countries, were taken from an earlier validation experiment undertaken to test the stability of the standard operating procedure for the

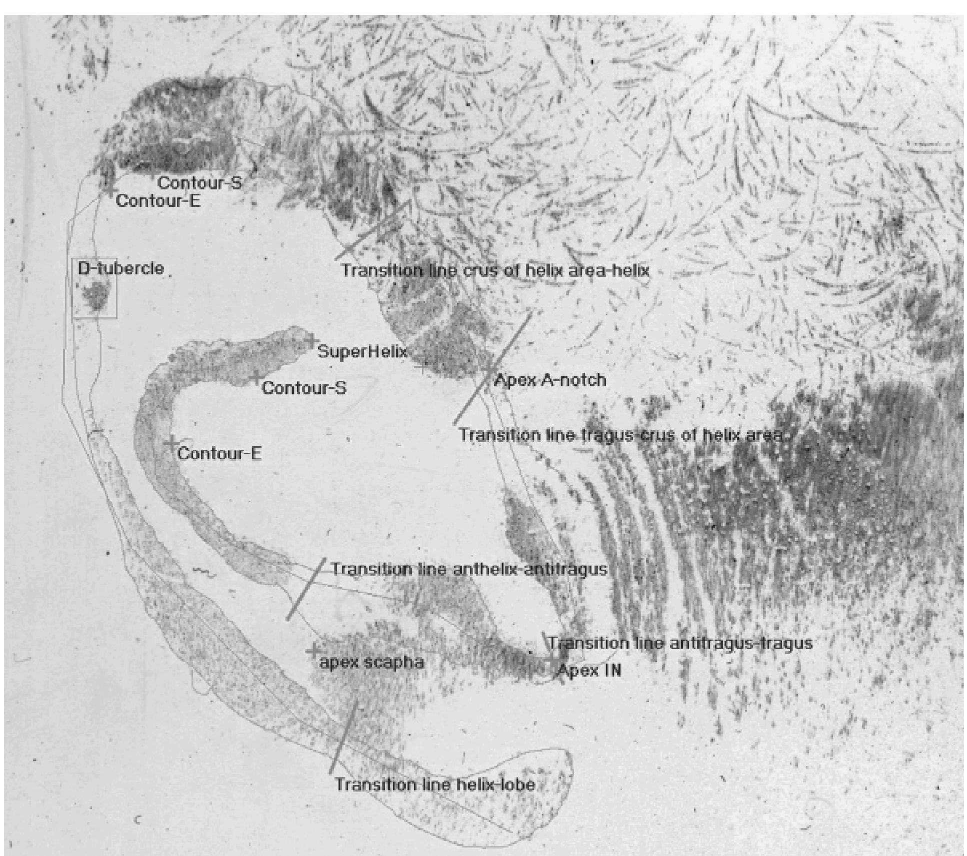

FIG. 2—*Example: earprint with superstructure and anatomical annotations.*

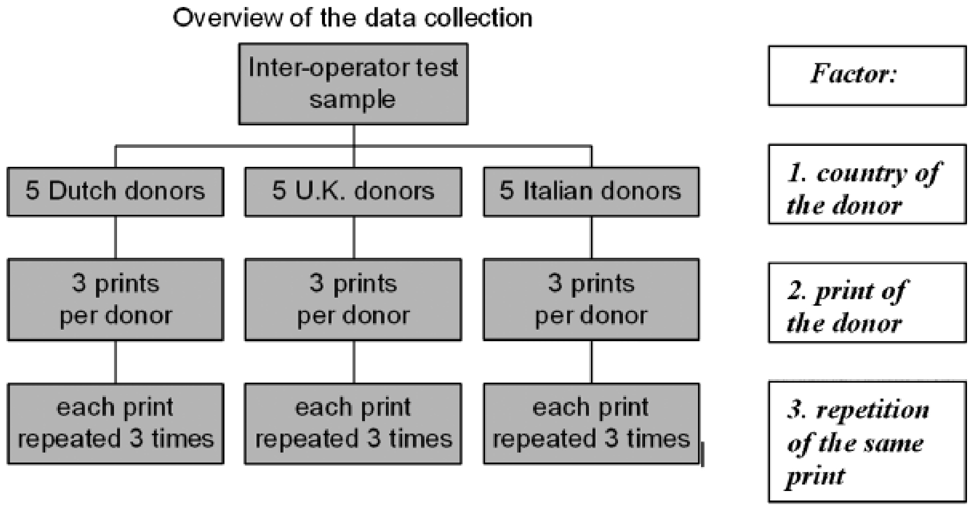

FIG. 3—*Overview of the data collection.*

taking of earprints as formulated by the project; see Johnson (8). The collection of 135 prints was offered in a fixed order, consisting of three consecutive blocks of the same 45 prints, each time put in a random order.

Thus, to keep the sample as representative and balanced as possible, and at the same time allow investigation of inter- and intraoperator effects connected to the annotation of prints, the sample was divided up along the following lines:

1. the *country* from which the donor was taken (three possibilities);
2. *specific donor*: given any country, five donors were selected;
3. *earprint number*: given any donor, three prints were selected; and
4. *specific annotation* of a particular earprint: three annotations by any operator per print.

The above, for example, allows to test whether interoperator effects are different for donors from different countries. The sample is summarized in Fig. 3.

Anatomical Annotation of Earprints

After print collection, earprints were annotated in a three-stage process for which FearID's Earprint Storage and Analysis System was used. The first stage was carried out at the collection sites Centrex NTC (U.K.), LSOP (The Netherlands), and Padova University (Italy). Stages two and three were carried out independent of the collection site, in a different country from the site where they were collected. All annotations were based on a standard set and agreed upon by the anthropological analysis group that consisted of Glasgow University, Leiden University Medical Centre, and Padova University.

The first annotation stage consisted of operators clicking the initial axis of the superstructure or *super-helix* structure used to represent the visible features of the ear palm (or *pinna*); see Figs. 1 and 4.

The superstructure is a spiral-like structure, starting at the *anthelix,* or *anti-helix*, and ending at the *lobule*, chosen as appropriate for subsequent computerized analysis of print patterns. The initial axis is a piecewise linear path linking points placed along the superstructure, with additional specific points representing the lowest point of the lobe and, in the anthelix area, apex and indentation markers, indicating the number and extent of anthelix branches; see Fig. 5.

Only apex and indentation markers and the lowest point of lobe were included in the point matching set of a print, as they were linked to specific locations in the earprint while the points along the superstructure were less specific.

During the second stage of print annotation, operators added print transition lines. These indicate the transition between print segments that one encounters when following the superstructure path, see Fig. 6, and contain information about superstructure segmentation.

In the third and final stage, the anthropological analysis group annotated *anatomical features*, *minutiae*, *landmarks,* and some other characteristics. Here

- an *anatomical feature* refers to anything to be seen in a print, including gross features, but also other characteristics like, for instance, a Darwinian tubercle or a crease. It is a very general term,
- *minutiae* are characteristic anatomical details that may match a print uniquely to a particular live ear, such as, for instance, a mole, a Darwinian nodule, a piercing, a scar, or a particular crease formation at a particular position, and
- the term *landmark* is used for a particular predefined point on a gross feature, or on the outline of a lacuna, unlikely to change through time.

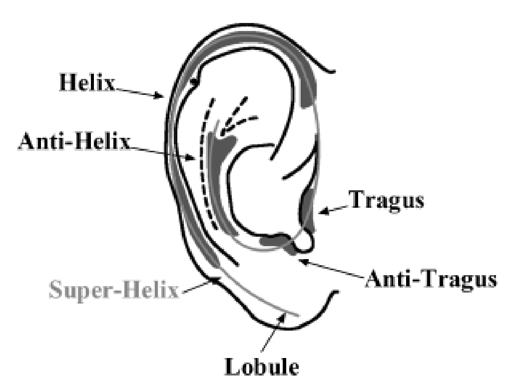

FIG. 4—*Initial axis of earprint superstructure.*

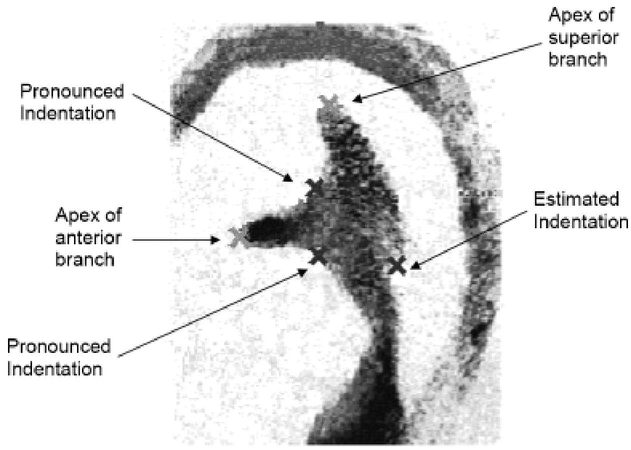

FIG. 5—*Anthelix points.*

The main categories were divided up as follows:

- Anatomical features and minutiae:
  1. Mole
  2. Knob of helix
  3. Notch of helix
  4. Darwinian tubercle (nodule and enlargement)
  5. Piercing
  6. Crus of helix posterior
- Landmarks:
  1. Apex of scapha
  2. Apex of anterior notch
  3. Apex of intertragic notch
- Other characteristics:
  1. Skin complexion and detail
  2. Scar
  3. Hairstyle
  4. Glasses
  5. Crease
  6. Pimple

FIG. 6—*Earprint transition lines and print orientation vector.*

Further subdivision along these lines and inclusion of features from the first and second annotation stage led to a total of 104 features used in the subsequent analysis. A full description of the instructions for marking minutiae and related landmarks in earprints is given in (7).

Performance Measure

We turn to the measure used to test the performance of the system. With respect to evidential value of earprint comparisons, there are two key concepts: that of *verification*, or one to one comparison, and that of *individualization*, or one to *n* comparison. Of the corresponding performance measures, *Equal error rate* (ERR) and *hitlist behavior*, we shall further concentrate on the first, Equal error rate, to express the performance of the system given different circumstances with respect to operators.

A verification system is a classification system with two classes of outcomes: matching (or positive or acceptance) and nonmatching (or negative or rejection). Given the features in a system, for any comparison of two prints, a single value is constructed optimally summarizing the matching information. Classification takes place according to whether the outcome does or does not exceed some threshold t.

Common performance parameters with this type of system are the probabilities of making a wrong judgment, expressed in the false rejection rate (FRR) (cases in which the system declares a nonmatch in case of matching prints) and false acceptance rate (FAR) (cases in which the system declares a match in case of nonmatching prints). As the FRR and FAR are threshold-dependent, we concentrate on the EER, which is the (common) probability of misclassification starting from the threshold t for which $\text{FAR}(t) = \text{FRR}(t)$. As an example of this, we depict the end results of the project based on three different feature extraction methods, including the current comparison, cf. (6), in Fig. 7.

Vector Template Matching (VTM)

Depending on the nature of a particular earprint feature, operators used either a point, line or area marker to highlight landmarks and minutiae. Point minutiae were included in the point matching set of a print. Line and area landmarks were processed to extract salient points such as extremities and bifurcation points of creases and barycenters of papules, thus reducing annotated line and area features to characteristic points as well. The coordinates were stored for each print, including the feature names (label), and formed the earprint point pattern used for matching earprints. The patterns were represented as lists of the form

$$P \equiv [p_1, p_2, \ldots p_i, \ldots p_{N_P}],$$

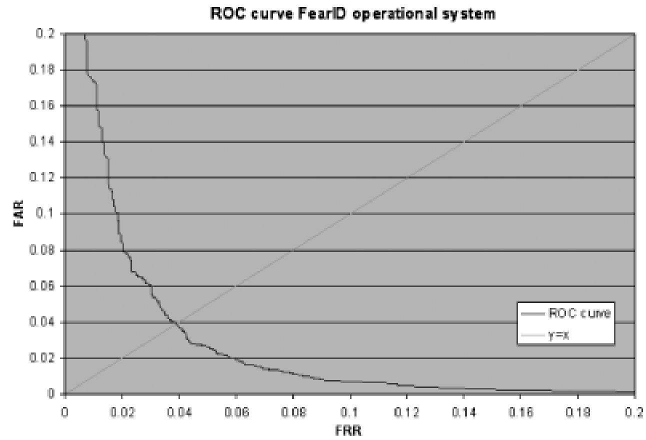

FIG. 7—*Plot depicting false acceptance rate (FAR) versus false rejection rate (FRR) resulting from the analysis of the Main sample of the forensic ear identification (Fear 1D) project. Results are on the basis of three feature extraction methods. The equal error rate (c. 4%) can be seen at the intersect of the curve and the line.*

where list item $p_i = \langle x_i^{(P)}, y_i^{(P)}, \ell_i^{(P)} \rangle$ gives the Cartesian coordinates of a point with respect to some arbitrary axis derived from print digitization, together with a label indicating the nature of the point. The ordering of items in the points list is arbitrary.

The coordinates of points in different prints from the same ear will differ because of having different origin and rotation, so the comparison mechanism needs to be invariant to translation and rotation. Our numerical analysis is based on a method called VTM.

As described, each print has a template consisting of labeled points representing annotated earprint landmarks and minutiae, distinguished into different classes. Prints are compared by assessment of the similarity between their templates.

In the case of prints originating from the same ear, the same labels are expected to turn up, although because of translation and rotation of the ear not being on the same coordinates. Comparison of the templates takes place in the following way. For any vector in print 1, all vectors in print 2 are determined sharing the same, anatomically meaningful, labels. In Fig. 8, this would, for example, mean starting at the vector with the *A* and *B* in print 1, and comparing this with the corresponding vector found below in print 2.

The angle between the vectors, in this case *c*. 26°, is determined and stored. In order to minimize comparison of vectors that *do* have matching labels but do not correspond, the ratio of the length of the vectors—which is anatomically supposed to be close to one—is supposed to be inside the fixed interval (0.90; 1.11), which is the result of independent training of the method. The

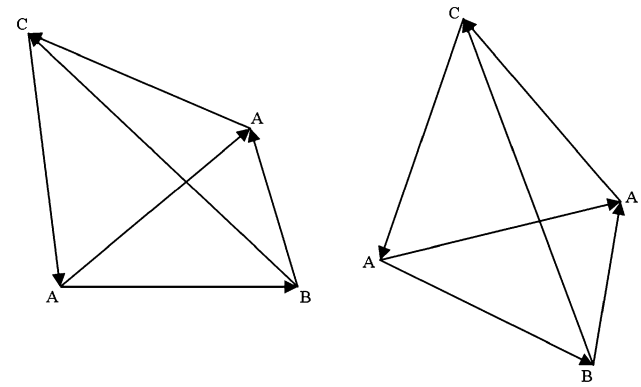

FIG. 8—*Example of two labeled point patterns, differing only in translation (26°).*

above is done for all combinations of vectors from prints 1 and 2 sharing the same labels and approximate length, and a histogram is made of the observed outcomes.

In the case of two matching prints, for the histogram this will result in a peak near the actual rotation of the one print with respect to the other, and low variation in outcomes. In the case of nonmatching prints, the histogram is expected to be noisy; see Fig. 9.

Two point patterns are shown to be similar by assessing the dominant mode of the distribution of the angles between pairs of vectors from the two patterns. For this, for any VTM comparison, nine features were extracted from the resulting histogram:

1. *N*, total number of comparisons of vectors in the procedure;
2. *Std*, standard deviation of the resulting histogram;
3. *IQR*, inter quartile range of the resulting histogram;
4. *Peak,* peak value of the resulting histogram;
5. *Peak/N*;
6. *Peak/Std*;
7. *Peak/IQR*;
8. *Ntot*, total number of possible comparisons of vectors in the VTM procedure, not considering either labels or vector length ratio; and
9. *N/Ntot.*

For details about the implementation and training (with respect to histogram bin width and vector length ratio) of the VTM method, see (9). The analysis of the outcomes is based on the statistical method of binary logistic regression (BLR).

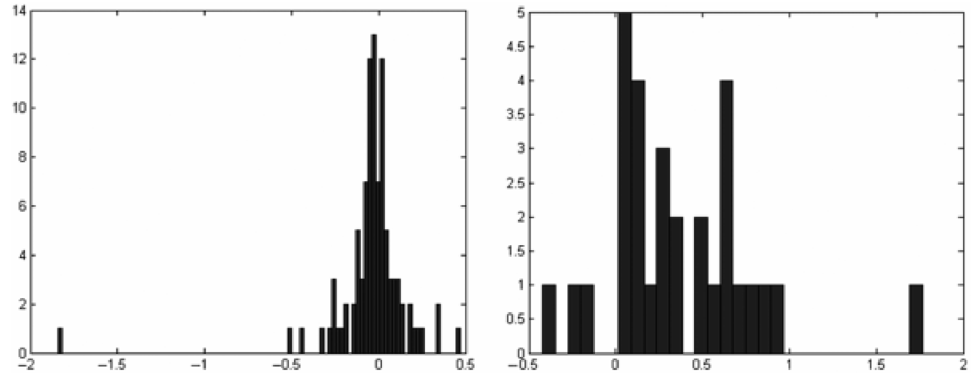

FIG. 9—*Example of histograms corresponding to a pair of matching prints (on the left) and a pair of nonmatching prints. In the case of matching prints, we see a compact histogram, in this case around rotation angle 0. In the case of nonmatching prints, the histogram is more "noisy."*

Logistic Regression

We aim to obtain a score function that minimizes the number of features used, but is still close to optimizing the EER. The reasoning behind this is that the nine VTM parameters are all highly correlated, and there is a risk of overtraining of the system on the data. The analysis of the outcomes is based on the statistical method of Binary Logistic Regression, or BLR, cf. (10). BLR is used rather than linear discriminant analysis as normality assumptions on the features are violated and some features are discrete, in which case BLR outperforms linear discriminant analysis. Based on the training data, the BLR method extracts a linear combination of certain of the used features, optimally separating pairs of matching from pairs of nonmatching prints.

For reasons of symmetry, we added the inverse values and natural logarithms to the mentioned features, thus ending up with 27 features per comparison of prints. However, no more than one instance of the same feature was used in the end model for the score function: for example, either *N*, *1/N,* or *log N*. Comparisons for which the VTM algorithm led to $N = 0$ or $N = 1$ were filtered out as in that case the histogram lacks informative value. (Note that in practice, this is what one would do as well.) Approximately 6900 comparisons (1.1% of the total number) were filtered out this way.

For the training sample, there were a total of circa 5.3 million possible comparisons of nonmatching prints and 3084 of matching prints, on the basis of which the BLR score was trained. Because of filtering of bad, double, and empty prints, as well as comparisons with VTM results for which $N < 2$, the number of valid comparisons further decreased to 2727. As BLR is not robust against differences in size between separable groups, four nonoverlapping subsamples of around 4700 (filtered) combinations of nonmatching prints were taken and trained against the fixed sample of (all) 2727 matching prints. In this way, four training samples of matching and nonmatching prints were made, with noncorrelated nonmatching parts.

The BLR analysis was performed such that first the system of 27 features was trained, using low thresholds for removal of features. For this, we used the SPSS module for BLR, Backward: LR method, with the parameter settings "probability for stepwise entry and removal" both at 0.01. The outcomes for the four different training samples were compared. This led us to one "steady" feature of the original nine, namely *peak/IQR*. Besides, features connected to *N*, *Std* (in the case of *N*: *1/N* and *log(N)*) regularly emerged, as well as log(*N/Ntot)*. Training the system on the basis of all of these features together, an EER of 6.8% was achieved for the model, where the BLR method for all four training samples removed the features connected to *Std*. Trying out the possible permutations of remaining features, it turned out that *Peak/IQR* and ln(*N/Ntot)* together led to a steady EER (over all four training samples) of 6.6%.

As we aim for simplicity of model and it leads to no significant loss in EER it was decided to further use the discriminant score

$$D = 1/(1 + \exp(-(14.7 + 0.017\text{Peak}/\text{IQR} + 3.1\ln(N/Ntot)))).$$

The implementation and training (with respect to histogram bin width and vector length ratio) of the VTM method, as performed in Kieckhoefer (9), were on the basis of this score.

Results

Annotation Frequencies

From here on, in box plots, the boxes denote the inter quartile range, their middle lines denoting the median. Whiskers show the distance from the end of the box to the largest and smallest observed values less than 1.5 box lengths from either end of the box. Circles denote outliers (between 1.5 and three box lengths from the end of the box), and stars extreme points (more than 3 box lengths from the end of the box).

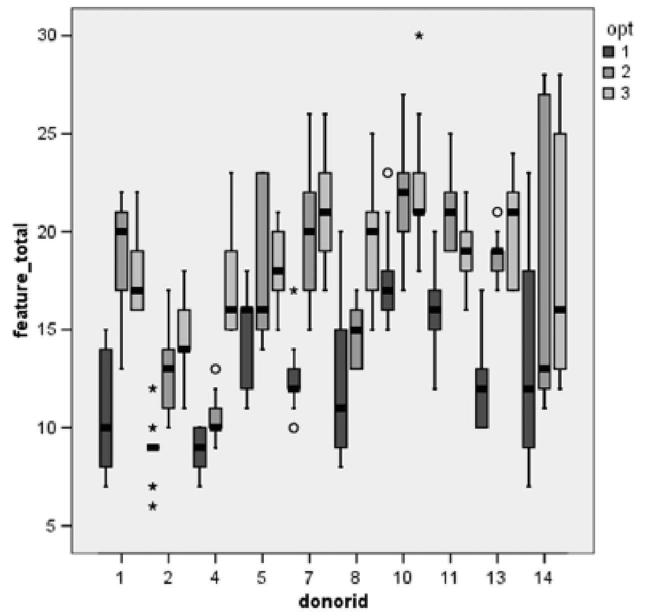

FIG. 10—*Box plots of total number of clicked features for 10 donors (on the x-axis) by different operators.*

We start by looking at annotation frequencies over the three operators. As there were 104 possible labels that could be used per clicked point, it is not very interesting to study frequencies per label (as they are usually 0), so we concentrate on the total number of annotated points per print. First, we present the results per donor, divided up as to operator (for each donor, all three operators annotated nine prints); see Fig. 10.

The above, summarized for all donors, is combined in Fig. 11.

We see substantial differences in clicking frequency between operators, operator one annotating on average around 12 points per print, whereas operator three is annotating 17. Although there is also an overlap between the boxes, the differences are significant. This is formalized by a two-way analysis of variance, starting from three (per operator) times 15 (per donor) samples of size

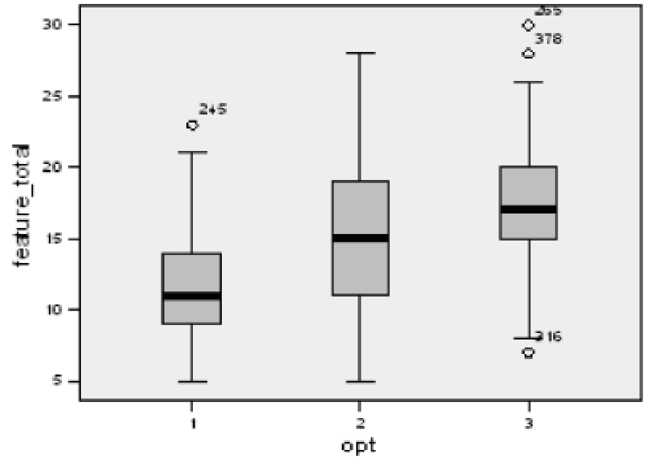

FIG. 11—*Box plots of total number of clicked features per operator.*

three (mean number of clicked features per print of a certain donor by a certain operator).

Interoperator Effects

We study the discriminating power of the generated outcomes, that is, the EER performance of the BLR model. To this, we divide the possible comparisons between annotated prints into three groups:

1. comparison of pairs of annotated versions of the *same earprint;*
2. comparison of pairs of *different prints from the same ear;* and
3. comparison of pairs of prints originating *from different donors*.

Distances between outcomes for matching and those of nonmatching prints are supposed to increase among the groups, and interoperator effects will be informative for groups 1 and 2 only.

In (9), the discriminating power of the BLR outcomes on the basis of the VTM method was summarized in an EER of 6.6% for comparisons of controlled prints. This means that there is an outcome above which precisely 6.6% of the nonmatching, and below which precisely 6.6% of the matching comparisons end up. Here, by the term "controlled prints," we mean lab quality prints, taken following the standard operating procedures laid down in Johnson (8), and per donor annotated by the same operator.

Regarding the discriminating power of the BLR outcomes in the current experiment, we compare discriminant scores that emerge for the groups 2 and 3. To resemble a real-life scenario, for group 2 we use only prints clicked by different operators and illustrate the process in Fig. 12. Here, on the left approximately 38,000 nonmatching comparisons are gathered into a box plot; on the right, 1200 and 600 matching ones are gathered, respectively, annotated by different operators and identical ones.

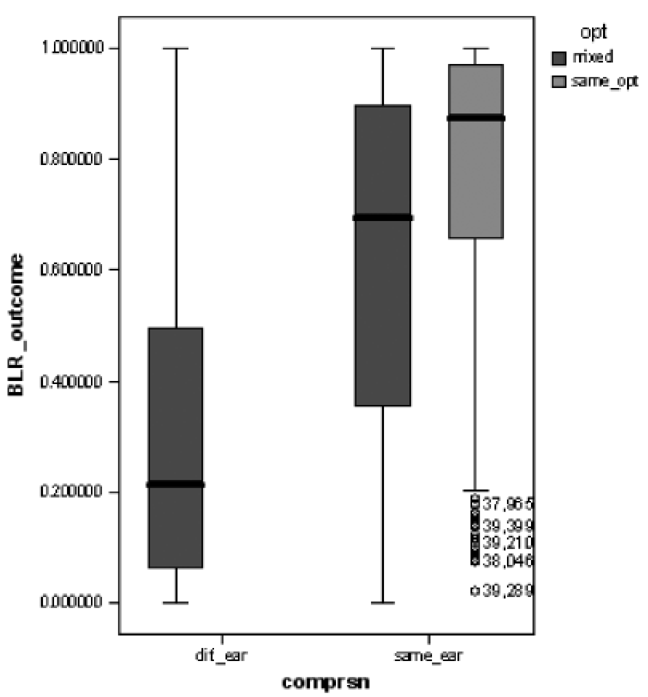

FIG. 12—*Comparison of (binary logistic regression based) discriminant scores for pairs of prints, starting from (on the left) different ears, and (on the right) different prints from the same ear, clicked by different operators.*

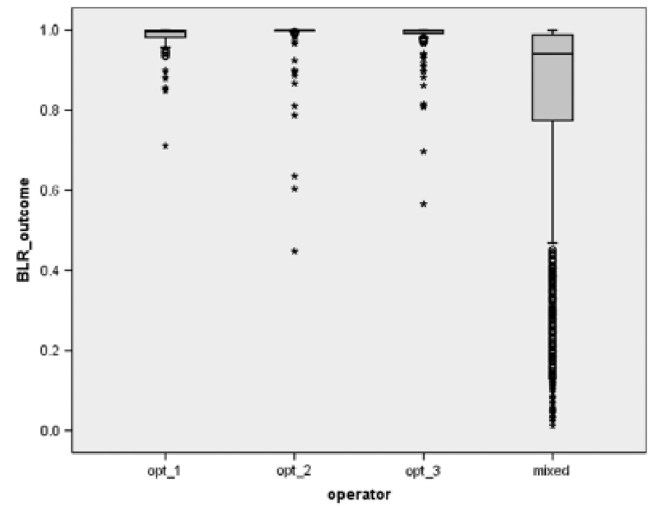

FIG. 13—*Box plots describing the (binary logistic regression based) discriminant scores encountered when comparing identical prints (group 1), either annotated twice by the operator, the three participating countries (box plots one up to three), or by two different operators (box plot four).*

The picture illustrates that whereas BLR outcomes for matching and nonmatching comparisons seem to be well separated with the same operator annotating the matching prints, with an EER of 20%, matching prints denoted by *different* operators are much less easy to separate, with an EER of 29%. We note that in a real-life scenario, different operators will be annotating the prints, so here the 29% is the percentage of interest.

Next, we look at the interoperator effects for matching comparisons. In Fig. 13, box plots are depicted describing the (BLR based) discriminant scores encountered when comparing identical prints (group 1), either annotated twice by the operator from Italy, The Netherlands, or the United Kingdom, or by two different operators. The latter (outcomes for identical prints annotated by different operators) is on the basis of C. 1200 outcomes. Boxes per operator are on the basis of 135 outcomes.

In Fig. 14, box plots are depicted describing the discriminant scores encountered when comparing different prints from the same ear (group 2) or from different ears (group 3). For the first group, first box plots for outcomes are given with the same operator annotating both prints, and then (in the fourth box) with mixed operators. The fifth box depicts outcomes for comparisons of prints of nonmatching ears. Boxes per operator are again on the basis of 135 outcomes, for mixed operators annotating the same ear on basis of 207 comparisons.

The EER of 20% corresponds to the difference between the first, second, and third box plot versus the fifth, the EER of 29% to that between the fourth box plot versus the fifth. In both cases shown in Figs. 13 and 14 , the interoperator effect is quite clearly visible.

Conclusions

The objective of the validation experiment at hand was to study the stability aspects of manual anatomical annotations that, like minutiae in fingerprints, function as the starting point for comparison of earprints. In our analysis of the main FearID database, on the one hand the situation was such that print gathering took place by the same one or two operators per country. Moreover, two manual annotation procedures formed the basis of the comparison process for which the operators per donor were not varied.

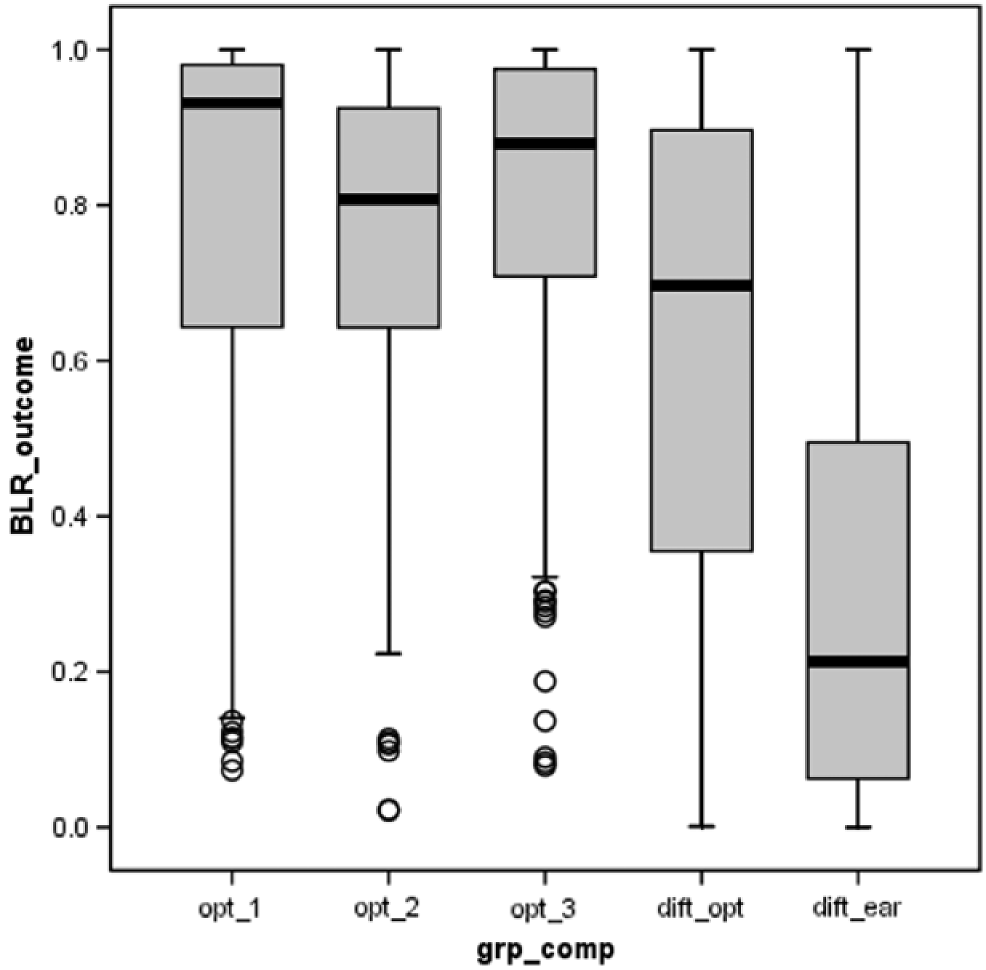

FIG. 14—*Box plots describing the (binary logistic regression [BLR] based) discriminant scores encountered when comparing different prints of the same ear (group 2), versus different prints of separate ears (on the right). The first three box plots describe outcomes for comparisons of matching ears annotated twice by the same operator (from the different countries). The fourth box plot describes outcomes for comparisons of matching ears annotated by different operators. The fifth describes comparisons between nonmatching prints.*

The stability of anatomical annotation that the current paper is about is quite disappointing. Equal error rates are significantly increasing (from 20% to 30%) when one starts looking at prints annotated by different operators. Hence, the results reported in Alberink et al. (6) and Kieckhoefer et al. (9) will not hold in practice, as in a real-life scenario different operators will be annotating the prints—and should be, as the aim is to objectify the current practice of (subjective) print comparison.

In the above, only a small sample of 15 donors was involved, with prints of worse quality than those in the FearID main sample: EER results that should be comparable read as 20% for the current test and 6.6% for the analysis of the FearID main sample. However, the data suggest that the end results on the FearID main sample using VTM comparisons are in reality, in which operators vary, worse than reported on the basis of the analysis of the main sample. Operator variation on both clicking frequency and choice of annotation points themselves suggests that implementation of the above is not the best way to go about objectifying earprint comparison.

It is important to keep processes like these in mind for any forensic science that deals with identification, be it of glass, tool marks, fibers, faces, fingers, handwriting, or speakers, where manual (nonautomated) processes play a role. For operator effects in fingerprint recognition, see, e.g., Evett et al. (11). In cases like these, the results may well be operator dependent and these dependencies need to be studied.

Acknowledgments

This study was carried out within the FearID project, which is a shared-cost RTD project funded under the 5th Framework Programme of the European Community, within the Competitive and Sustainable Growth Programme, Measurement and Testing Activity, Contract G6RD-CT-2001-00618. The authors would like to thank Francesca De Conti (University of Padova), Lynn Meijerman (University of Leiden), and Sarah Sholl (University of Glasgow) for annotation of the sample, and the referees for useful comments that helped to improve the paper.

References

1. Lugt van der C. Earprint identification. Den Haag: Elsevier Bedrijfsinformatie, 2001.
2. Champod C, Evett IW, Kuchler B. Earmarks as evidence: a critical review. J Forensic Sci 2001;46:1275–84.
3. State v. David, Kunze Wayne. Court of Appeals Washington, Division 2, 97 Wash. App. 832, 988 P.2d, 977 (1999).
4. BBC online network. Ear print catches murderer, http://news.bbc.co.uk/1/hi/uk/235721.stm, BBC online network, December 15, 1998.

5. Times Online. Man convicted of murder by earprint is freed, http://www.timesonline.co.uk/article/0,,1-973291_1,00.html, Times Online, January 22, 2004.
6. Alberink IB, Ruifrok ACC. Performance of the FearID earprint identification system. For Sci Int, In press.
7. De Conti F, Giacon M, Meijerman L, Sholl S. Instructions for marking minutiae and related landmarks in earprints. Internal FearID report, 2004, http://forensic.to/fearid/Labelling_minutiae_and_characteristic_landmarks.pdf.
8. Johnson Z. Standard Operating Procedure for the taking of earprints. Internal FearID report, 2003, http://forensic.to/fearid/Procedure.doc.
9. Kieckhoefer H, Ingleby M, Alberink IB. Vector template matching of earprints using labelled point features. Internal FearID report, 2005, http://forensic.to/fearid/VTMfinal.doc.
10. Pampel FC. Logistic regression: a primer. Sage university pagers series on quantitative applications in the social sciences, series no. 07-132. Thousand Oaks, CA: Sage Publications, 2000.
11. Evett IW, Williams RL. A review of the sixteen points fingerprint standard in England and Wales. J For Ident 1996;46:49–73.

Additional information and reprint requests:
Ivo B. Alberink, Ph.D.
Netherlands Forensic Institute
Department of Digital Technology
PO Box 24044
2490 AA, Den Haag
The Netherlands
E-mail: ivo@holmes.nl