Proceedings of Business and Economic Studies, 2022, Volume 5, Issue 2
http://ojs.bbwpublisher.com/index.php/PBES
ISSN Online: 2209-265X
ISSN Print: 2209-2641

Innovation Research on Fingerprint and DNA Identifications

Jiaer Fei*

Zhejiang Technical Institute of Economics, Hangzhou 310018, Zhejiang Province, China

***Corresponding author:** Jiaer Fei, 670010@zjtie.edu.cn

Abstract: In this study, two models for fingerprint and DNA identifications are constructed based on modern technologies, while offering significant advances over prior models. Our models have high credibility, obtaining relatively accurate results under different circumstances. Under different assumptions, this model tests the probability of the validity in the statement that human fingerprints are unique to be 93.94%. In other words, the percentage of misidentification is 6.06%. This model is a robust fingerprint identification method that can tolerate highly nonlinear deformations. The model is tested on the basis of a self-built database, proving that the model has high credibility, and convincing results are obtained from sensitivity analysis. In order to estimate the odds of misidentification by DNA evidence, we emphasized on two factors that might contribute to misidentification: random match probability and the probability arising from laboratory errors. Then, a model is developed using Bayes' theorem to reveal the inherent relationship between them, which carries some reference value. The probability of matching by DNA evidence is estimated based on the changes in the significant level. Finally, the probabilities of misidentification by both fingerprint evidence and DNA evidence are compared using numerous data. We found that the probability of the former is 6.06% and that of the latter is smaller than 4.0×10^{-10}. Therefore, it can be concluded that DNA identification is far better than that of fingerprint identification.

Keywords: Sensitivity analysis; Minutiae triplets; Triangle features; Corresponding triangle; Bayes' theorem

Online publication: April 28, 2022

1. Fingerprint identification

1.1. Assumptions

(1) We assume that every country has its own template fingerprint database.

(2) When the corresponding triangles are obtained, we assume that the information contained by the vertex and the corresponding minutiae points are identical.

According to several studies [1-4], the structure of the model is shown in **Figure 1**.

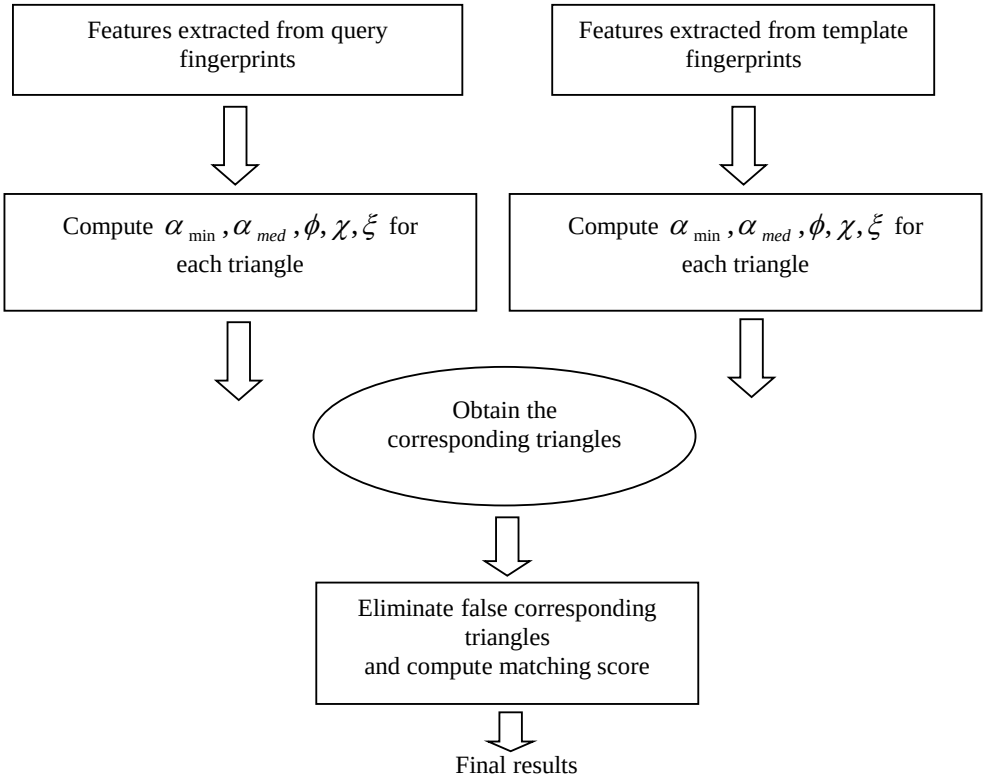

Figure 1. Structure of the fingerprint identification model

1.2. Fingerprint model

(1) Step 1: MATLAB program is used to extract the self-collected fingerprint features [5-8]. The approximate images of the locations of the minutiae are then obtained. Based on the locations of the minutiae, a triangle is formed among the minutiae in one fingerprint image, which is shown in **Figure 2**.

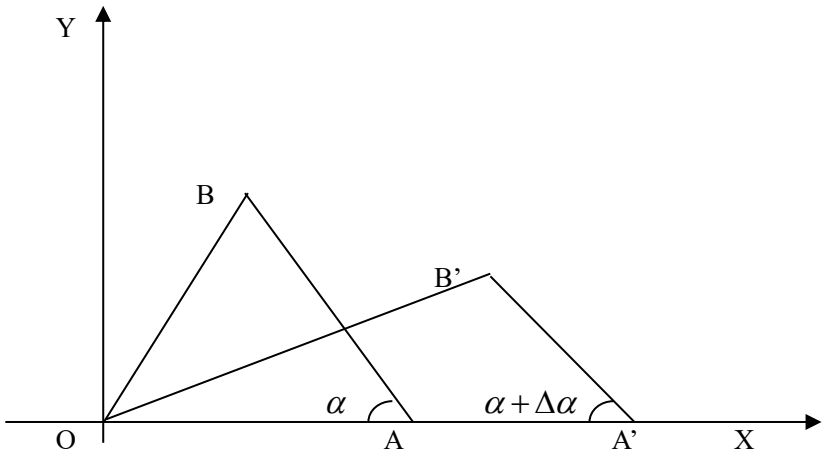

Figure 2. Uncertainty of minutiae locations

According to internet sources and a study by Tan and Bhanu [2,9,10], the factual data are α_{min}, α_{med}, φ, λ, and ξ, which are defined below.

(a) Angles α_{min} and α_{med}

Suppose α_i represents three angles in the triangle, i = 1, 2, 3. Let $\alpha_{max} = \max\{\alpha_i\}$, $\alpha_{min} = \min\{\alpha_i\}$, and $\alpha_{med} = 180° - \alpha_{max} - \alpha_{min}$.

(b) Triangle orientation φ

Let $z_i = x_i + jy_i$ be the complex number ($j = \sqrt{-1}$) corresponding to the coordinates (x_i, y_i) of point p_i, where $i = 1, 2$. Define $Z_{21} = Z_2 - Z_1$, $Z_{32} = Z_3 - Z_2$, and $Z_{13} = Z_1 - Z_3$. Let $\varphi = \text{sgn}(Z_{21} \times Z_{32})$, where sgn is the signum function and × is the cross product of two complex numbers.

(c) Maximum side λ

Let $\lambda = \max\{L_i\}$, where $L_1 = |Z_{21}|$, $L_2 = |Z_{32}|$, and $L_3 = |Z_{13}|$.

(d) Ridge counts ξ

ξ_1 is the ridge count of the side $p_1 p_2$; ξ_i is the ridge count of the side $p_2 p_3$; ξ_3 is the ridge count of the side $p_3 p_1$; ξ is a vector consisting of all ξ_i.

(2) Step 2: Using the same method, the query fingerprint is processed. It is computed and the feature data is obtained as follows: α_{min}, α_{med}, φ, λ, and ξ.

(3) Step 3: The features of every triangle of the query fingerprint are sequentially compared to the features of each triangle of the fingerprint that is stored in the database (features of 12 fingerprints). When the two smaller angles, α_{min} and α_{med}, are approximately equal to each other and the ridge counts are also approximately equal to each other, we can assume that the two fingerprints weakly correspond to each other. If the longest lengths are equal to each other, and the directions of the two vectors of the largest angle are also equal, we can assume that the two fingerprints strongly correspond to each other. The number of the corresponding triangles is then determined.

1.3. Model testing

(1) Step 1: A fingerprint from the database is selected as the query fingerprint and is numbered n. MATLAB is then used to extract the features of the query fingerprint and obtain the approximate locations of the minutiae. Triangles are then developed among the minutiae and the feature data is computed as follows:

$$\alpha_{min}, \alpha_{med}, \varphi, \lambda, \text{ and } \xi$$

In view of the complexity of the data, some feature data of the triangles are selected.

(2) Step 2: The uncertainties of α_{min}, α_{med}, φ, λ, and ξ, are determined.

(a) The angles of the triangles are allowed an uncertainty of ±2°. The uncertainties of φ, λ, and ξ are kept unchanged while changing the uncertainty of the angle. For example, taking ±1°, ±2°, and ±4°, respectively, as the uncertainties of the angles, the number of corresponding triangles is obtained. The results are then analyzed as shown in **Figure 3**.

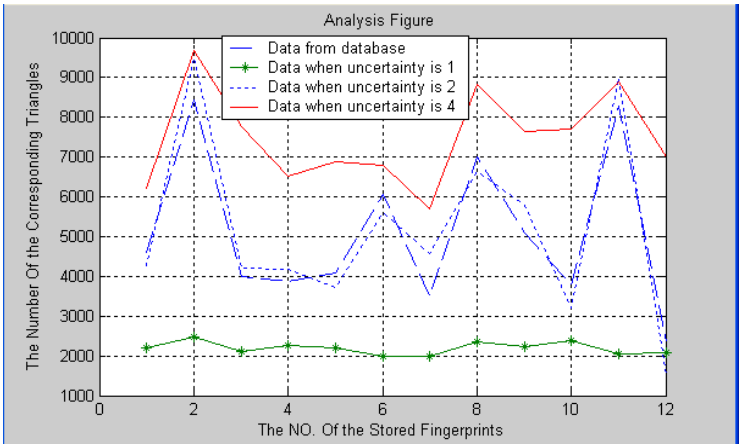

Figure 3. Analysis of the number of stored fingerprints

From **Figure 3**, it can be seen that the uncertainties of the angles have significant influence on the number of corresponding triangles. If the uncertainty is too small, the number of corresponding triangles is approximately the same among different pairs of fingerprints (the query fingerprint and the one from the database against which the query fingerprint is compared with are assumed as a pair; however, due to the multitude of fingerprints found on the database, there are many pairs); hence, it is difficult to distinguish. Therefore, $\Delta\alpha = \pm 2°$ is chosen.

(b) In the same way, the uncertainties of α, φ, and ξ are kept unchanged, while changing the uncertainty of λ. Taking 0.1 (pixel), 1 (pixel), and 2 (pixel) as the uncertainty of λ, the number of corresponding triangles is obtained. The results are analyzed as shown in **Figure 4**.

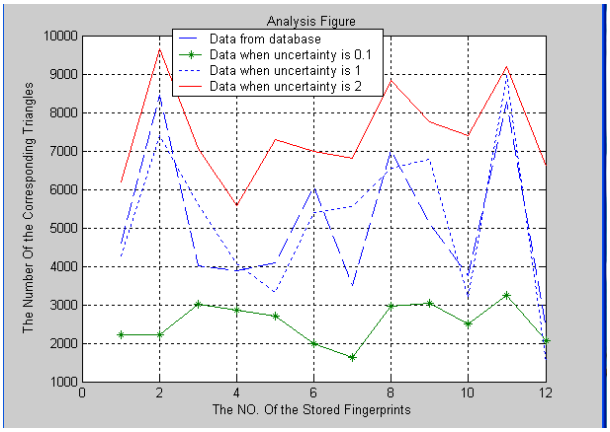

Figure 4. Analysis of the number of stored fingerprints

(c) From **Figure 4**, it can be seen that when $\Delta\lambda = \pm 1$ (pixel), the number of corresponding triangles can be clearly distinguished. Hence, $\Delta\lambda = \pm 1$ (pixel) is chosen.

(d) Using the same method, $\Delta\xi = \pm 1$ (pixel) is determined.

(e) Step 3: Five fingerprints that have more corresponding triangles than others when compared with the query fingerprint are selected. We defined Td as the two points to be corresponding points (same position). When the Td is too large, the results of Step 2 is not as accurate, whereas if the Td is too small, some corresponding triangles may be omitted when in fact, they should not be. Taking into account the two possibilities, the results when Td = 1, 10, and 20 (pixel) are analyzed. The results are shown in **Figure 5**.

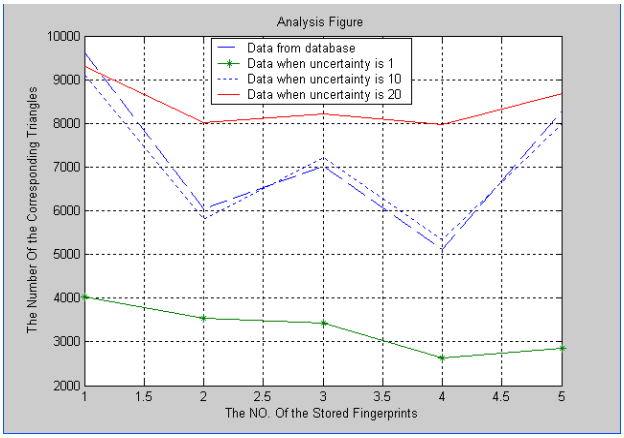

Figure 5. Analysis of the number of stored fingerprints

From the analysis above, the use of Td = 10 yields the best result. Hence, Td = 10 (pixel) is chosen.

1.4. Probability of misidentification

In order to determine whether the statement that each fingerprint is unique is true, we grouped several pairs of fingerprints that are made up of two separate fingerprints. The number of corresponding triangles for each pair of fingerprints is then determined.

In order to determine whether the fingerprint is unique, we analyzed the data found on the database, and the results are obtained.

We defined a = (1/2) * (TO / TA + TO / TB) * 100%, where TO is the number of corresponding triangles, TA is the total number of triangles of A, and TB is the total number of triangles of B.

(1) Suppose that the average percentage, a, is above 50%, the two fingerprints match. Six data (71.75%, 57.43%, 67.43%, 78.26%, 62.61%, and 65.81%) out of 66 data were found to be above 50%, thus obtaining the result as follows:

$$a = (66 - 6) / 66 = 0.9091$$

In the sense that the percentage is 90.91%, the statement that each fingerprint is unique is true.

(2) Suppose that the percentage, a, is above 60%, the two fingerprints match. Three data (71.75%, 67.43%, 78.26%, and 65.81%) out of 66 data were found to be above 60%, thus obtaining the result as follows:

$$b = (66 - 4) / 66 = 0.9394$$

In the sense that the percentage is 93.94%, the statement that each fingerprint is unique is true.

(3) Suppose that the percentage, a, is above 70%, the two fingerprints match. Two data (71.75% and 78.26%) out of 66 data were found to be above 70%, thus obtaining the result as follows:

$$c = (66 - 2) / 66 = 0.9697$$

In the sense that the percentage is 96.97%, the statement that each fingerprint is unique is true (**Figure 7**).

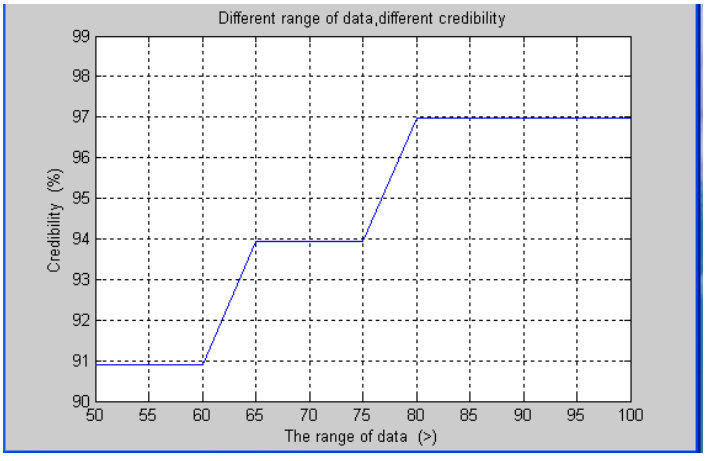

Figure 7. Different range of data with different credibility

It is generally safe to say that the fingerprint of each human being who has ever lived is different. This model tests the probability of the validity in the statement that human fingerprints are unique to be greater than 90% under various assumptions. The model has been put to test using a self-built database, proving that the model has high credibility.

1.5. Strengths and weaknesses

1.5.1. Strengths

(1) This model serves as a robust fingerprint identification method that can tolerate highly nonlinear deformations.
(2) In this model, the triangles represent the features of the fingerprint, while the number of corresponding triangles represent the similarity of two fingerprints, thus simplifying the fingerprint identification process.
(3) This model presents high credibility in that each fingerprint is unique, and the identification probability is high.

1.5.2. Weaknesses

(1) Due to the lack of data, the selection of uncertainties is not as accurate, resulting in some errors in the model.
(2) In determining potential corresponding triangles, the multitude of triangles existing in the query fingerprint and the sample fingerprint poses a great challenge. It increases the complexity of the model and degrades the performance of the algorithm

1.6. Future work

This model is only based on corresponding triangles. We presume that the corresponding minutiae points have been identified after defining the corresponding triangles. In other words, we merely looked at the number of minutiae points between two fingerprints, which is not very accurate. In order to achieve better matching results, we need to examine the features of the minutiae points in addition to their quantity.

2. Identification of DNA

2.1. Analysis

If two DNA samples are the same, it is generally assumed that the two samples are from the same person [11,12]. When they do come from the same person, the DNA evidence is strong. However, when using DNA evidence for identification, it is conceivable that in certain cases the DNA samples that are reported to be the same are not really from the same source, thus resulting in misidentification, as is often the case with forensic identification.

2.2. Model design

According to Thompson WC, Taroni F, and other internet source, the probability of two alternative propositions is assessed under the conventional expression of Bayes' theorem [13-15].

$$\frac{p(S \mid R)}{p(\bar{S} \mid R)} = \frac{p(S)}{p(\bar{S})} \times \frac{p(R \mid S)}{p(R \mid \bar{S})}$$

S refers to the specimen from a suspect, $\bar{S}$ refers to the specimen that did not come from a suspect, and R is the report of a match.

$$\frac{p(R \mid S)}{p(R \mid \bar{S})} = \frac{1}{RMP + [FPP \times (1 - RMP)]}$$

RMP is the random match probability, and FPP is the false positive probability on the DNA evidence.

The following equation can then be derived:

$$\frac{p(S \mid R)}{p(\bar{S} \mid R)} = \frac{p(S)}{p(\bar{S})} \times \frac{1}{RMP + [FPP \times (1 - RMP)]} \tag{1}$$

The equation can be modified as such:

$$RMP + [FPP \times (1 - RMP)] = \frac{p(R \mid \bar{S})}{p(R \mid S)} = \mathrm{K} \tag{2}$$

From the above equation, it can be seen that *RMP* and *FPP* have an inherent relationship. In order to clearly demonstrate the inherent relationship, let K = 10^{-2}, 10^{-3}, 10^{-4}, …, 10^{-10}, as shown in **Figure 8**.

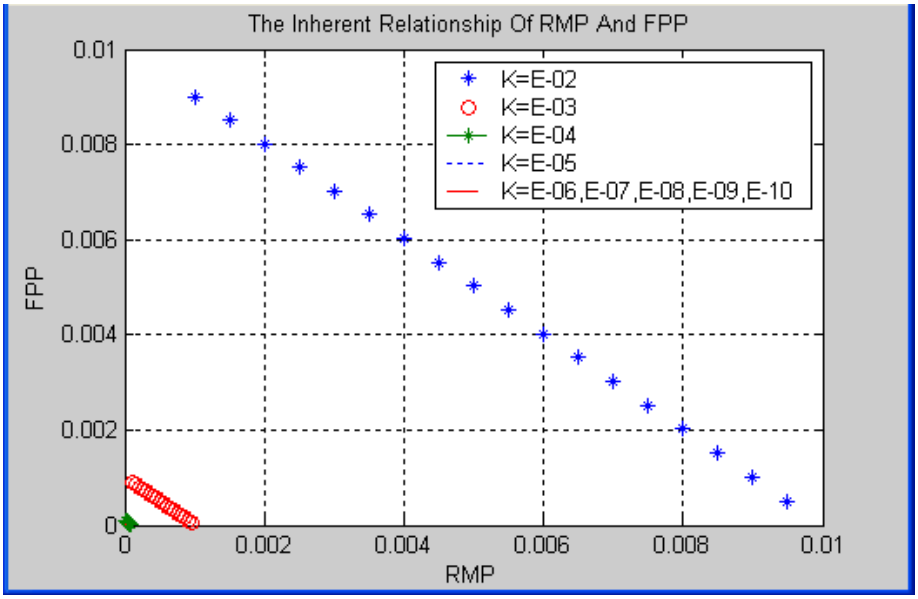

Figure 8. Inherent relationship of *RMP* and *FPP*

There is difficulty in accessing the DNA database, which presents a challenge for our task. Nonetheless, our aim is to determine the odds of misidentification by DNA evidence, which is based on *RMP* and *FPP* in this model. Unfortunately, no accurate results are obtained; only the inherent relationship between the two components is revealed. Therefore, we can only estimate the odds.

$$P \leq 1 - (1 - \alpha)^{\frac{1}{N}} \tag{3}$$

P is the total match probability, α refers to the significant level, and *N* is the number of a random group of people.

It can be assumed that the DNA image is unique among a group of *N* people. If two DNA images are identical, it can be considered that they are from the same person. Suppose the American population has 260,000,000 people, and given α = 0.01, then the credibility = 99%, and the maximum *P* is 3.9×10^{-11}. Therefore, if *P* is smaller than 3.9×10^{-11}, it can be considered that when the credibility is 99%, the DNA image is unique among the population. **Table 1** shows the total match probability, *P*, under different significant levels.

Table 1. Total match probability at different significant levels

Significant level, α	Total match probability, P
0.001	4×10^{-12}
0.005	1.9×10^{-11}
0.01	3.9×10^{-11}
0.05	2.0×10^{-10}
0.1	4.0×10^{-10}

Generally speaking, the significant level α is usually less than 0.1. As seen in **Table 1**, when α is smaller than 0.1, the total match probability P under all circumstances is smaller than 4.0×10^{-10}, and there is an inverse relationship between P and α. Therefore, the conclusion that $P \leq 4.0 \times 10^{-10}$ can be drawn.

3. Conclusion

In the first model, when the percentage of (PA + PB) / 2 is above 50%, we assume that the two fingerprints match. In this way, the percentage of identification is 90.91%, and the percentage of misidentification is (1 – 90.91%) = 9.09%. When (PA + PB) / 2 is above 60%, we assume that the two fingerprints match. In this way, the percentage of identification is 93.94%, and the percentage of misidentification is (1 – 93.94%) = 6.06%. When the percentage of (PA + PB) / 2 is above 70%, the percentage of identification is 96.97%, and the percentage of misidentification is (1 – 96.97%) = 3.03%. Based on the analysis above, the average percentage of misidentification is as follows:

$$P = (0.0909 + 0.0606 + 0.0303) / 3 = 0.0606 = 6.06\%$$

Hence, in the first model, the percentage of misidentification is 6.06%.

In the second model, by analyzing the data in **Table 1**, we find that all the total match probabilities, P, are smaller than 4.0×10^{-10}. Hence, the percentage of misidentification is far smaller than that in the first model. From this, we may infer that DNA identification is far superior to fingerprint identification, and that DNA identification is employed more frequently in practice.

Disclosure statement

The author declares no conflict of interest.

References

[1] Derakhshani R, Schuckers SAC, Hornak LA, et al., 2003, Determination of Vitality from a Non-Invasive Biomedical Measurement for Use in Fingerprint Scanners. Pattern Recognition, 36(2): 383-396.

[2] Tan X, Bhanu B, 2003, A Robust Two Step Approach for Fingerprint Identification. Pattern Recognition Letters, 24(13): 2127-2134.

[3] Que C, Tan KB, Sagar VK, 2001, Pseudo-Outer Product Based Fuzzy Neural Network Fingerprint Verification System. Neural Netw, 14(3): 305-323.

[4] Leis T, Repp R, Borkhardt T, et al.,1998, A New Fingerprint Method for Sequence Analysis of Chromosomal Translocations at the Genomic DNA Level. Leukemia, 12: 758-763.

[5] Sujan VA, Mulqueen MP, 2002, Fingerprint Identification Using Space Invariant Transforms. Pattern Recognition Letters, 23(5): 609-619.

[6] Ross A, Jain A, Reisman J, 2002, Proceedings of 16th International Conference on Pattern Recognition, August 11-15, 2002: A Hybrid Fingerprint Matcher. IEEE, Quebec City, QC, Canada, 795-598.

[7] Rossa A, Jaina A, Reisman J, 2003, A Hybrid Fingerprint Matcher. Pattern Recognition, 36(7): 1661-1673.

[8] Brodeur MB, Debruille JB, 2003, Visual Evoked Potentials to Line- and Luminance-Defined Triangles. Vision Research, 43(3): 299-306.

[9] Tahmasebi AM, Kasaei S, 2002, A Novel Adaptive Approach to Fingerprint Enhancement Filter Design. Sig Proc: Image Comm, 17(10): 849-855.

[10] Bhanu B, Tan X, 2004, Computational Algorithms for Fingerprint Recognition, Springer, Netherlands.

[11] Dziuk P, 2003, Positive, Accurate Animal Identification. Reprod Sci, 79(3-4): 319-323.

[12] Choubey D, Gutterman JU, 1996, The Interferon-Inducible Growth-Inhibitory P202 Protein: DNA Binding Properties and Identification of a DNA Binding Domain. Biochemical and Biophysical Research Communications, 221(2): 396-401.

[13] Thompson WC, Taroni F, Aitken CGG, 2003, How the Probability of a False Positive Affects the Value of DNA Evidence. J Forensic Sci, 48(1): 47-54.

[14] Taroni F, Biedermann A, Garbolino P, et al., 2004, A General Approach to Bayesian Networks for the Interpretation of Evidence. Forensic Sci Int, 139(1): 5-16.

[15] Wan QH, Qian KX, Fang SG, 2003, A Simple DNA Extraction and Rapid Specific Identification Technique for Single Cells and Early Embryos of Two Breeds of Bos Taurus. Animal Reproduction Science, 77(1-2): 1-9.