

J. H. Liu,¹ Ph.D.; C. H. Lin,¹ LL.M.; J. W. Osterburg,¹ M.P.A.;
and J. D. Nicol,¹ M.S.

Fingerprint Comparison. II: On the Development of a Single Fingerprint Filing and Searching System

REFERENCE: Liu, J. H., Lin, C. H., Osterburg, J. W., and Nicol, J. D., "Fingerprint Comparison. II: On the Development of a Single Fingerprint Filing and Searching System," *Journal of Forensic Sciences*, JFSCA, Vol. 27, No. 2, April 1982, pp. 305-317.

ABSTRACT: A FORTRAN program has been written to compare the minutiae coded in an "unknown" fingerprint with the minutiae of fingerprints stored in a data base. The criteria for matching are scores based on the type and number of minutiae matched. Minutiae of low relative frequency have higher scoring weights than those that are more common. The matching mechanism is tested by using a fingerprint coded several times, first by a single individual and then by six other individuals who have no previous knowledge about fingerprints. These tests yield satisfactory results.

KEYWORDS: criminalistics, fingerprints, identification systems, minutiae, probability of identification

Although latent fingerprint identification is based on the comparison of minutiae in single fingerprints, current fingerprint filing and searching systems [1] are based on the pattern and ridge count of ten fingers. This method of filing creates a problem for identifying a latent fingerprint, even if the fingerprints of the person are on file.

The development of automatic fingerprint identification systems has been a subject of considerable attention. Gaffney et al [2] proposed some approaches to implementing an automatic fingerprint identification system (AFIS) to compare the encoded reproduction of a single print with many prints taken from a very large file and then indicate which prints are most similar to the print to be identified. Wegstein and Rafferty [3] also proposed similar approaches. AFIS was further developed along several lines, including such systems as fingerprint automatic classification techniques [4], a "videofile" information system [5], the New York State Intelligence and Information System [6], and the Federal Bureau of Investigation's automatic fingerprint identification system [7]. Although all systems were primarily designed for a ten-finger system, the FBI's system is based on ridge direction and minutiae (ridge endings and bifurcations only) and is most suitable for single fingerprint comparison.

The purpose of this study is to develop a computerized matching mechanism for comparing the minutiae of a questioned fingerprint with the minutiae of fingerprints stored in a data base. Similarity of fingerprints is judged by comparison scores derived basically from the "probability of identity" studies [8-10]. It is hoped that the development of this matching mechanism leads to a practical single fingerprint filing and searching system.

Received for publication 28 May 1981; accepted for publication 22 Sept. 1981.

¹Assistant professor, graduate student, professor, and professor, respectively, Department of Criminal Justice, University of Illinois at Chicago Circle, Chicago, IL 60680.

Core and Baseline Setting

The minutiae coding and comparisons rely on rules for reproducibility in setting a core and a baseline in each fingerprint. Core setting is based on the Henry System [1]. Since there are no cores in arches, new conventions were designed for setting cores in various types of arches as follows. The top of the lowest arch-ridge is defined as the core of a plain arch (Fig. 2a). Tented arches are further classified into angle, upthrust, and resemblance-to-loop arches. For an angle arch, the core is set at the vertex of the angle (Fig. 2b); for an upthrust arch, the core is set at the ending point (Fig. 2c); for a resemblance-to-loop arch, the core is

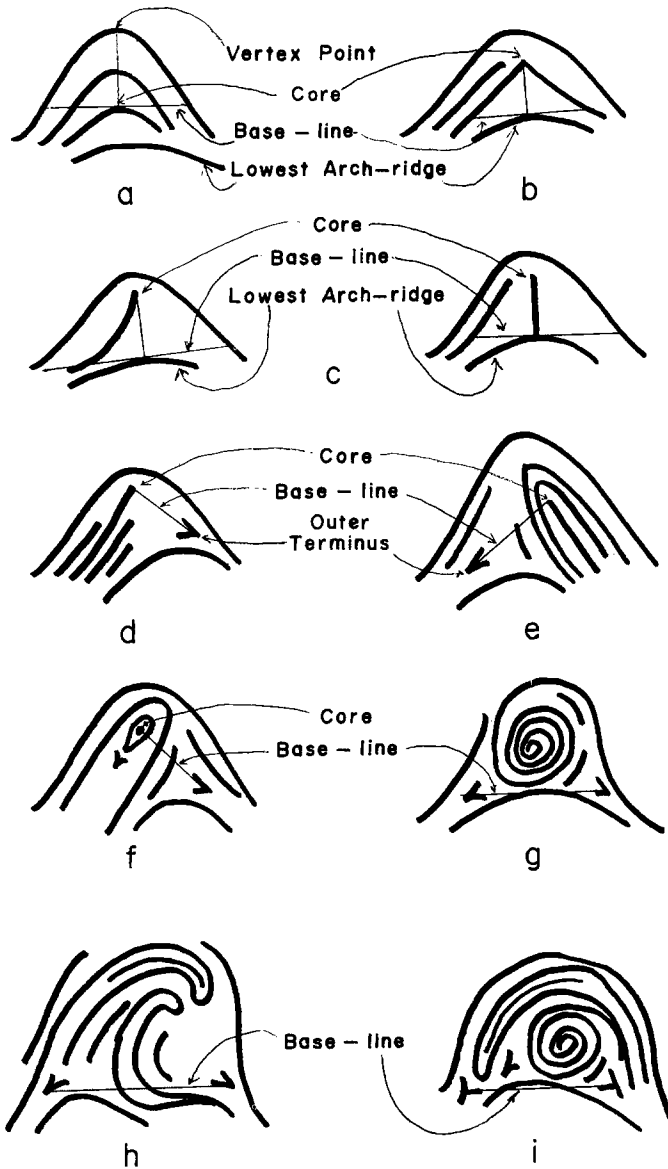


FIG. 2—Conventions for setting baseline in various fingerprint patterns.

set in the same way used for a loop (Fig. 2d). The convention [12] used to determine baselines for various fingerprint patterns is described in Table 1 and illustrated in Fig. 2.

Minutiae Comparison

The flow chart of a computer program written to compare a questioned fingerprint with those in the data base is shown in Fig. 3. The program first counts the frequency of occurrence of each minutia in all fingerprints in the data base. An empty cell has informational content and in this study is considered the same as a fingerprint characteristic. A probability parameter is estimated for each type of minutia. The probability parameter of a minutia is obtained by dividing the frequency of occurrence of that minutia by the total number of grid cells within the fingerprint area. The negative log of the probability parameter [9] is used as the weight for minutiae comparison scoring purposes.

The major parts of this program concern the consecutive location of minutiae in all cells in the questioned fingerprint, followed by the location of the same minutiae in the same or neighboring positions in the known fingerprint against which the questioned print is to be compared. If the same minutia is found in the same cell position or in a cell that differs by one row or one column, or both, it is considered a match. A score, which is the total of all weights of the particular matching minutiae, is compiled after all cells are compared. (Once a minutia in a fingerprint is matched, it may not be used for matching again.) The comparison is continued with the next fingerprint in the data base until all fingerprints in the data base are examined.

The program then proceeds to rank the scores obtained from comparing the questioned fingerprint with each fingerprint in the data base. A predetermined number of best matching scores are printed. Codes and legends of fingerprints that give these best matching scores are also identified, so that they may be pulled from the file for visual comparison.

Results and Discussions

The occurrence frequency, probability parameter, and weight of each minutia in the data base used for this study are listed in Table 2. The weights of most of these minutiae reported by Osterburg et al [9] are also included in Table 2. These two sets of weights are generally in good agreement, considering the small size of sample used in both studies and some differences in defining these minutiae. The frequency of occurrence of minutiae obtained by this study is further compared to those reported previously [8] (Table 3). The discrepancies

TABLE 1—*Conventions used to determine baselines in various patterns of fingerprints.*

Pattern	Convention
Plain arch, Fig. 2a	Connect all vertex points of the arch-ridges in the central part of the fingerprint. Draw a line through the core (see definition in the text) and perpendicular to the line connecting the vertex points.
Tented arch, Fig. 2b,c	Draw a vertical line through the core of T to the lowest arch-ridge. Through this point on the lowest arch-ridge, draw a line (the baseline) perpendicular to the vertical line.
Ulnar loop, Fig. 2d; radial loop Fig. 2-e; and, central pocket loop, Fig. 2f	Draw a line (the baseline) through the core and the outer terminus.
Plain whorl, Fig. 2g; double loop, Fig. 2h; and accidental, Fig. 2i	Draw a line (the baseline) through the left-most and right-most outer termini.

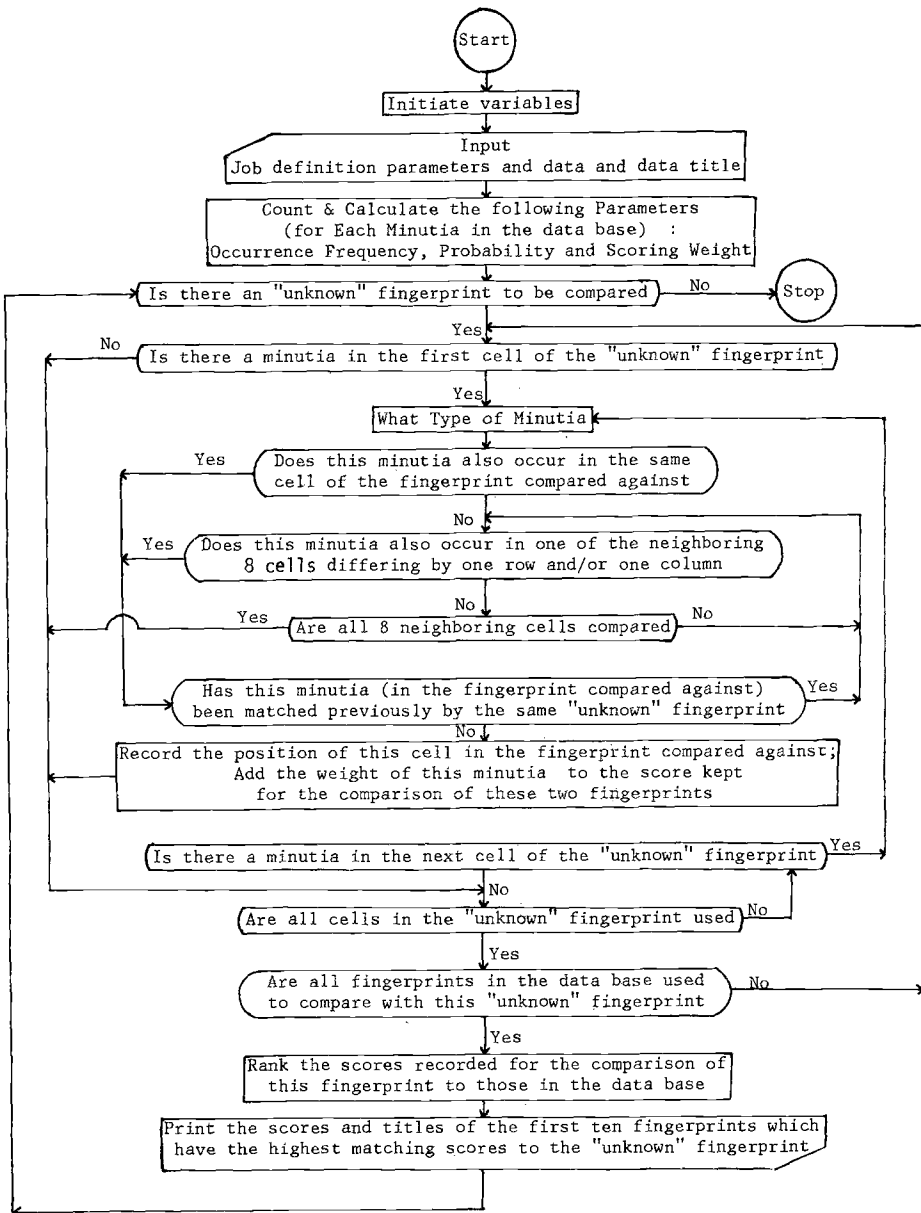


FIG. 3—Flow chart of the minutiae comparison program.

noted between each study are probably a result of the small population samples used in all of these studies and of different definitions for each minutia.

Recently, there have been several interesting "probability of identity" studies [8-10], which are based on the modeling of minutiae occurrence and the computation of probabilities. In these studies, the occurrence frequency of each minutia was first calculated from a data base. A probability parameter based on the frequency was assigned to each minutia. Various minutiae configurations that would constitute the lower limits of "legal identifica-

TABLE 2.—Occurrence frequency, probability parameter, and weights of the minutiae found in the 38 pairs of fingerprints and weights reported by Osterburg et al [9].

	Ridge Ending	Short Ridge	Dot	Bifurcation	Spur	Double Bifurcation	Eye	Broken Ridge	Angle Ridge	Empty Cell
Occurrence frequency	3571	370	191	967	179	101	65	443	65	31 063
Probability parameter	9.647	0.9996	0.5160	2.613	0.4836	0.2729	0.1756	1.197	0.1756	83.92
Weights	1.016	2.000	2.287	1.583	2.316	2.564	2.756	1.922	2.756	0.0761
Weights by Osterburg et al [9]	1.08	...	1.82	1.42	2.13	2.85	2.19	1.86	...	0.116

TABLE 3—Comparison of minutiae frequencies reported by different authors.

	Osterburg et al [13]	Santamaria [14]	Kingston [8]	This Study
Ending ridge	0.471	0.534	0.459	0.600
Short ridge	0.062
Dot	0.096	0.076	0.083	0.032
Bifurcation	0.146	0.151	0.341	0.163
Spur	0.066	0.030
Double bifurcation	0.009	0.017
Eye	0.025	0.042	0.032	0.011
Broken ridge	0.074
Angle ridge	0.011
Bridge	0.072	...	0.019	...
Tri-radii	0.008	...	0.017	...
Trifurcation	0.007
Other	0.100	0.197	0.031	...

tion" were computed. The probability of identifying a wrong person was discussed in these studies.

It has been shown [9] that the common practice in matching twelve characteristics represents various levels of probability, depending on the types of minutiae used. For example, the matching of three trifurcations has a smaller probability than that of twelve ridge endings. It was also shown [9] that the common practice of using twelve characteristics to constitute a "legal identification" represents a probability of 10^{-20} or less. Therefore, any configuration resulting in a probability less than 10^{-20} constitutes an identification.

Since the matching score is the sum of the weights (or negative log of the probability parameter) of each minutia matched, any configuration resulting in a sum greater than 20 would, according to accepted practice, yield an identification. Problems, however, may be encountered in applying this theoretical calculation to a computerized comparison algorithm because of difficulties in coding minutiae correctly and consistently. There are at least two sources of coding errors:

1. A minutia may not be consistently coded in the same cell because of the variation in distance or direction in relation to the core. This variation may be caused by different printing conditions or the slight variation in positioning the core and setting the baseline.
2. A minutia may not be coded consistently either because of a difference in interpretation or simply by being overlooked.

The error in coding a minutia in a neighboring cell may result in a loss of a legitimate match or the gain of an unfounded match. To avoid this problem the comparison algorithm allows for the searching of a match in the neighboring cells differing by one row or one column, or both. This means that for each minutia in a cell of the questioned fingerprint, there are nine cells to be searched in the fingerprint being compared. This relaxation in defining matching positions unavoidably results in many false matches and increases matching scores. As a consequence, the customarily adopted "legal identification," which represents a probability of 10^{-20} (or a matching score of 20), can no longer be considered as an identification in the matching mechanism used in this study. Therefore, the comparison algorithm abandons the idea of making an identification; rather, it simply ranks all matching scores and prints out a specified number of fingerprints that constitute the best matches. It is then the operator's task to identify which of this limited number of fingerprints constitutes an identification. This algorithm also resolves problems caused by missing or miscoding a few minutiae in the coding process.

Associated with the relaxation in allowing a match with the minutia in a neighboring cell,

the characteristic of a cell without a minutia [9] can no longer be used in the matching process. Undoubtedly, one of the nine cells used in the comparison will be empty; therefore, keeping score of this match is meaningless.

With this matching and scoring mechanism, the highest possible score depends on the number and the types of minutiae coded in the fingerprint. Undoubtedly, the highest score will result from matching with itself and represents a one-to-one perfect match. With the 76 fingerprints used in this study, the scores range from 173.185 to 43.431 (Table 4). The scores, number, and types of minutiae in the ten fingerprints (which constitute the five highest and the five lowest matching scores obtained by matching with themselves) are listed in Table 4. Entries in the second column of Table 4 are the highest scores obtained by matching with the other 75 fingerprints. The distinct difference between the entries listed in the first and the second columns of Table 4 indicates the validity of this approach, providing the coding process is accurately performed.

More significantly, this matching mechanism is tested by observing results obtained from an inconsistent minutiae coding process. Two steps were taken to test the variation of matching scores resulting from using a less-than-perfectly-coded fingerprint to compare with those in the data base. First, a fingerprint was coded and then coded three more times by the same person. These three "unknown" fingerprints were then compared to those in the data base. Second, after being instructed for about 40 min on the coding rules, a group of six college students in a sophomore class, who had had no previous knowledge about fingerprints, were each asked to code Fingerprint 8. These "unknown" fingerprints were similarly compared to those in the data base.

Results obtained from these two sets of experiments are listed in Table 5. The high scores (as compared to those listed in Columns 4 through 8) listed in the third column of Table 5 indicate that these unknowns always give the highest comparison scores with the control fingerprints stored in the data base and are therefore easily identified.

Obviously, this computerized matching mechanism is far from perfect at this stage. The following issues require further development and improvement:

1. More universal and more accurate rules are needed to determine the core and the baseline so that partial fingerprints can be easily positioned for coding.
2. A few more pretesting parameters and information, such as pattern and ridge count (class characteristics), are needed so that if a fingerprint in the data base differs from the questioned fingerprint in even one of these aspects, no further comparison will be performed. This prerequisite will certainly reduce computer time.
3. A mechanism should be developed so that the questioned fingerprint can be compared incrementally and successively to small sections of those fingerprints in the data base.
4. A large data base should be established so that more representative occurrence frequencies, probability parameters, and weights of each type of minutia can be calculated.

These improvements will undoubtedly facilitate the comparison of partial fingerprints.

Acknowledgments

The authors thank J. Chang, S. J. Hsiu, M. H. Lin, and C. J. Chen for their assistance in collecting the fingerprints used in this study. C. H. Lin thanks H. T. Lee, president of the Central Police College, Taiwan, for providing financial support during the course of this study.

Appendix: Minutiae Coding Rules and Procedure

I. Definition and General Rules

Lengths used in the definitions below refer to measurements made on the original size of the fingerprint.

TABLE 5—Scores obtained by matching "unknowns" with their corresponding fingerprints and other fingerprints in the data base.

Sample	Total No. of Minutiae Coded	Compared with Control	Compared with Five Other Prints Having Highest Scores				
FINGERPRINT 19, CODED BY ONE PERSON FOUR TIMES							
Control	46	56.847	29.690	28.640	27.628	27.181	27.061
Duplicate 1	32	38.775	33.321	32.411	30.295	30.114	29.659
Duplicate 2	31	36.853	30.380	30.107	29.935	29.211	28.264
Duplicate 3	32	37.869	35.714	29.208	28.920	28.264	28.195
FINGERPRINT 8, CODED BY SEVEN PERSONS							
Control	94	137.932	52.942	52.339	48.613	47.865	47.438
Student 1	47	64.783	41.661	38.842	36.828	36.310	36.298
Student 2	28	37.582	27.006	25.072	24.963	24.731	24.167
Student 3	36	58.604	27.909	21.118	26.933	26.259	25.531
Student 4	32	47.550	31.720	28.883	28.584	28.106	27.474
Student 5	26	38.100	32.765	30.335	26.986	26.376	26.143
Student 6	52	77.374	47.047	40.323	39.451	38.978	38.828

1. Ending ridge (or ridge ending): An ending ridge is defined as a minutia that has a distinct break in a ridge, called an ending point. Its location, indicated by the ending point, is coded as *1*. (The edges of inked fingerprints have many incomplete ridges that are not ending ridges; these are coded as *0*.)

2. Short ridge (or stick ridge): A short ridge is defined as a minutia whose length lies between 1 and 5 mm. Its location is indicated at the center of the whole ridge. If the length is 1 mm or less, it is defined as a dot. If the length is more than 5 mm, it is defined as two ending ridges. The short ridge is coded as *2*.

3. Dot (or island, very short ridge): A dot is defined as a minutia whose largest width or length is 1 mm or less. The location of a dot is the center of that dot. A dot may not be attached to a ridge line. There must also be room for at least one pore to be encompassed. If its largest diameter is less than 0.25 mm, it is an incipient ridge and is not counted. Also, a dotted ridge is not a ridge characteristic. A dot is coded as *3*.

4. Bifurcation (or fork): A bifurcation is defined as a ridge line that diverges into two ridges or two ridges that converge into one ridge. Its location is the converging point. It is coded as *4*.

5. Spur (or hook): A spur is defined as a bifurcation in which one leg is 3 mm or less. The end of a spur is not counted as a ridge ending. If the length of the leg is more than 3 mm, it is coded as bifurcation and ending ridge. The location of a spur is the converging point. It is coded as *5*.

6. Double bifurcation (or divergent ridge): A double bifurcation is two bifurcations in which the two converging points are separated by 3 mm or less. It is only counted as one minutia. If the length between two converging points is more than 3 mm, it is coded as two bifurcations. Its location is the first converging point. It is coded as *6* (Figure 4a).

7. Eye (or enclosure, central empty island, lake): An eye is defined as a minutia that looks like an eye. The length between the two converging points is 3 mm or less. If the length between two converging points is more than 3 mm, it is coded as two bifurcations. Its location is indicated at the center of the eye. It is coded as *7*.

8. Broken ridge: A broken ridge is defined as a ridge that is clearly broken on the way; however, its two ends are still on the same imaging line. If the break is vague, it is not counted. If its two ends are not on the same imaging line, it is coded as two ending ridges, *1*. The break distance is 1 mm or less. If the break distance is larger than 1 mm, it is coded as two ending ridges. A scar on the fingerprint is coded as a broken ridge. If there is more than

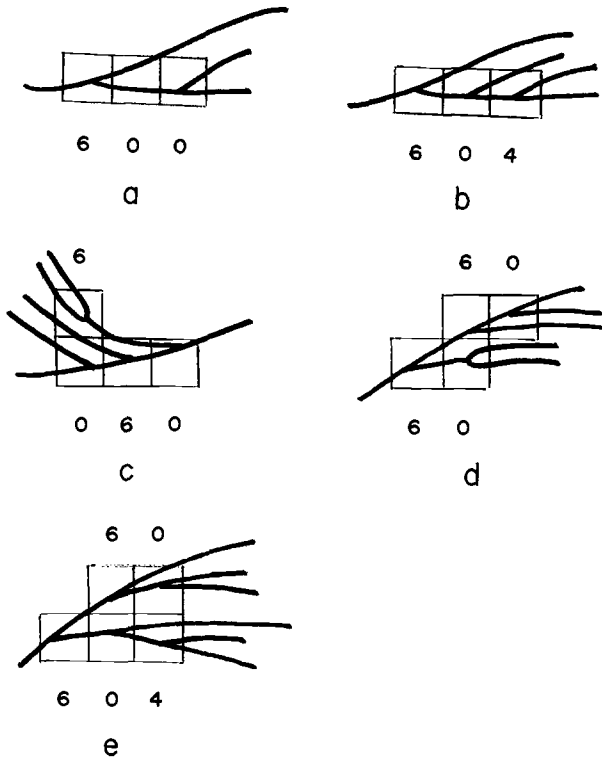


FIG. 4—Conventions for coding minutiae.

one broken ridge in a cell, they are still defined as one broken ridge. Its location is indicated at the center of the break. It is coded as 8. (A crease in a fingerprint does not mean a broken ridge and is not counted.)

9. Angle ridge: An angle ridge is defined as two ridges that converge and end at one point. The angle contained lies between 90° and 170° , inclusively. One of the legs is 3 mm or less in length. If one of the legs is more than 3 mm, it is coded as an ending ridge. An angle ridge is coded as 9.

10. Only ridge lines: Grid cells that are within the inked fingerprint area but without minutiae are coded as 0. Vague and smudged areas of fingerprint are also coded as 0.

11. Blank: Grid cells that are outside the fingerprint area are coded as 88. The ridge lines beyond the third phalange are also coded as 88.

II. Conventions

1. Vague or smudged ridge lines: If there is uncertainty in deciding which kind of minutia exits in a cell, it is coded as 1; if there is uncertainty in deciding whether there is a minutia in a cell, it is coded as 0.

2. Priority sequence: If a minutia is located on a border position of two or more grid cells, it is coded in the (upper) left cell. If this cell has already been coded with a minutia, the subject minutia is coded in the (upper) right cell. The next two cells are (lower) left and (lower) right cells.

3. Many minutiae in one cell: If there is more than one minutia in a cell, the minutia

nearest to the core is coded in that cell. The remaining minutiae, if any, are coded in the nearest cells. If the nearest cell is occupied, the next nearest cell is coded.

4. Complexity of minutiae: In cases where the formation of ridge characteristics can be interpreted as more than one type of minutia, the most complex one will take priority and be coded. The remainder of the ridge characteristics are coded accordingly. For the purpose of this study the complexity of minutiae is defined in the following order: double bifurcation, eye, spur, bifurcation, broken ridge, angle ridge, short ridge, ending ridge. For example, minutiae which include three converging points in three cells can be coded as one 6, one 0, and one 4, or three 4s. But, according to this convention, the former is selected (Fig. 4b). Likewise, minutiae that include four converging points in four cells are coded as two 6s and two or three 0s (Fig. 4c, d). Minutiae which include five converging points in five cells are coded as two 6s, one 4, and two 0s (Fig. 4e).

5. Scar: There are six kinds of scar patterns: pinpoint, multidotted, sharp linear, area furrow, island, and wrinkled. Pinpoint, multidotted, sharp linear, and area furrow scars are coded as a broken ridge, 8. Island and wrinkled scars are coded as dots, 3, or considered as ridge lines, 0, whichever is appropriate.

III. Coding Procedure

1. A core and a baseline are established within each fingerprint.
2. A grid transparency of 26 by 30 cells is superimposed on the fingerprint. The central point of the cell grid is positioned at the core of the fingerprint. The baseline on the fingerprint is set parallel to the horizontal lines of the cell grid with the following exceptions. For resemblance-to-loop arch, ulnar loop, radial loop, and central pocket loop, the baseline of a loop with a positive slope is set parallel to the positive slope diagonal lines of the cells; similarly, the baseline of a loop with a negative slope is set parallel to the negative slope diagonal lines of the cells.
3. Proceed with the coding process cell by cell from left to right and from upper to lower.
4. When the coding process is completed, review the coding sheet to make sure that all minutiae have been coded, especially those that represent the core, delta, and scar(s).

References

- [1] Federal Bureau of Investigation, *The Science of Fingerprints*, U.S. Government Printing Office, Washington, DC, 1977.
- [2] Gaffney, J. E., Jr., Falk, S. T., and Reppard D. L., "Fingerprint Encoding and Matching Procedures for Automatic Recognition Systems," in *Law Enforcement Science and Technology, Vol. 2, Proceedings of the Second National Symposium on Law Enforcement Science and Technology*, S. I. Cohn, Ed., IIT Research Institute, Chicago, 1968, pp. 401-407.
- [3] Wegstein, J. H. and Rafferty, J. F., "Machine Oriented Fingerprint Classification System," in *Law Enforcement Science and Technology, Vol. 1, Proceedings of the First National Symposium on Law Enforcement Science and Technology*, Thompson Book Co./Academy Press, New York, 1967, pp. 459-465.
- [4] Emden, B. M., "Advanced Computer Based Fingerprint Automatic Classification Technique, (FACT)," in *Law Enforcement Science and Technology, Vol. 1, Proceedings of the First National Symposium on Law Enforcement Science and Technology*, Thompson Book Co./Academic Press, New York, 1967, pp. 493-505.
- [5] Bradley, G. O., "Automatic Fingerprint Search Techniques with Videofile," in *Law Enforcement and Technology, Vol. 2, Proceedings of the Second National Symposium on Law Enforcement Science and Technology*, S. I., Cohn, Ed., IIT Research Institute, Chicago, 1968, pp. 391-399.
- [6] Lindh, T. K., "Computerized Fingerprint Searching Progress and Problems," in *Proceeding of the Third International Search Symposium on Criminal Justice Information and Statistics Systems, Philadelphia, Penn.*, J. Laucher, and M. Casey, Eds., Search Group, Inc., Sacramento, CA, 1976, pp. 115-122.
- [7] Banner, C. S., "The State of Development of the FBI's Automatic Fingerprint Identification System," in *Proceedings of the International Symposium on Criminal Justice Information and*

- Statistics Systems*, G. Cooper, Ed., California Crime Technological Research Foundation, New Orleans, 1972, pp. 407-422.
- [8] Kingston, C. R., *Probabilistic Analysis of Partial Fingerprint Patterns*, University Microfilms, Inc., Ann Arbor, MI, 1964.
- [9] Osterburg, J. W., Parthasarathy, T., Raghavan, T. E. S., and Sclove, S. L., "Development of a Mathematical Formula for the Calculation of Fingerprint Probabilities Based on Individual Characteristics," *Journal of the American Statistical Association*, Vol. 72, No. 360, 1977, pp. 772-778.
- [10] Sclove, S. L., "The Occurrence of Fingerprint Characteristics As a Two-Dimensional Process," *Journal of the American Statistical Association*, Vol. 74, No. 367, 1979, pp. 588-595.
- [11] Lin, C. H., Liu, J. H., Osterburg, J. W., and Nicol, J. D., "Fingerprint Comparison. I: Similarity of Fingerprints," *Journal of Forensic Sciences*, Vol. 27, No. 2, April 1982, pp. 290-304.
- [12] Lin, C. H., *The Science of Fingerprints*, Central Police College, Taipei, Taiwan, 1978, pp. 71-100.
- [13] Osterburg, J. W., "An Inquiry into the Nature of Proof: The Identity of Fingerprints," *Journal of Forensic Sciences*, Vol. 9, No. 4, Oct. 1964, pp. 413-427.
- [14] Steinwender, E., "Dactyloscopic Identification," *Fingerprint and Identification*, Vol. 41, No. 10, April 1960, pp. 4-5.

Address requests for reprints or additional information to
 Juei H. Liu, Ph.D.
 Department of Criminal Justice
 University of Illinois at Chicago Circle
 Box 4348
 Chicago, IL 60680