

AD-A043 802

RICE UNIV HOUSTON TEX
USE OF PATTERN RECOGNITION TECHNIQUES FOR TYPING AND IDENTIFICATION--ETC(U)
APR 77 M L CURTIS

F/G 13/2

DOT-CG-81-75-1383

NL

UNCLASSIFIED

USCG-D-38-77

| OF |
AD
A043 802



END
DATE
FILMED
9-77
DDC

REPORT NO. CG-D-38-77
Task No. 4243.3.1

14

B-5

AD A 043802

USE OF PATTERN RECOGNITION TECHNIQUES
FOR TYPING AND IDENTIFICATION OF OIL SPILLS

Morton L. Curtis
Rice University
Houston, Texas 77001

Under contract to:
U. S. Coast Guard Research and Development Center
Avery Point, Groton, Connecticut 06340



April 1977

Final Report

DDC
RECEIVED
SEP 7 1977
RECEIVED
B

Document is available to the U.S. public through
the National Technical Information Service
Springfield, Virginia 22161

AD No. ~~1~~
DDC FILE COPY

PREPARED FOR
U.S. DEPARTMENT OF TRANSPORTATION
UNITED STATES COAST GUARD
OFFICE OF RESEARCH AND DEVELOPMENT
WASHINGTON, D.C. 20590

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

The contents of this report reflect the views of the Coast Guard Research and Development Center, which is responsible for the facts and accuracy of data presented. This report does not constitute a standard, specification or regulation.

D. L. Birkimer

Donald L. BIRKIMER, Ph.D., P.E.
Technical Director
U.S. Coast Guard
Research and Development Center

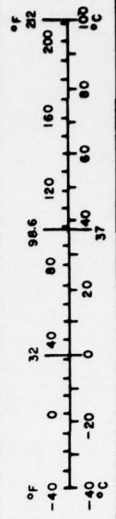
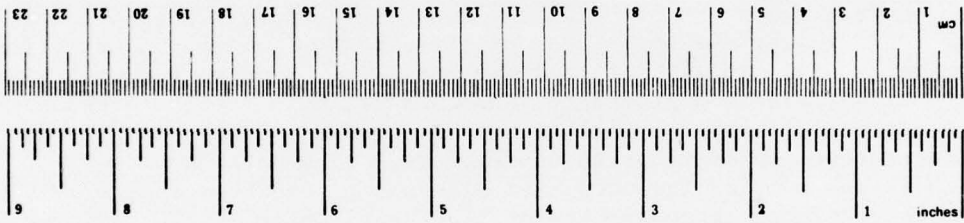
<p>1. Report No. 18) US CG-D-38-77 ✓</p>	<p>2. Government Accession No.</p>	<p>3. Recipient's Catalog No.</p>	
<p>4. Title and Subtitle 6) USE OF PATTERN RECOGNITION TECHNIQUES FOR TYPING AND IDENTIFICATION OF OIL SPILLS</p>		<p>5. Report Date 11) April 1977</p>	<p>6. Performing Organization Code</p>
<p>7. Author(s) 10) Morton L. Curtis</p>	<p>8. Performing Organization Report No. 12) 217.</p>		<p>10. Work Unit No. (TRAIS) 774243.3.1</p>
<p>9. Performing Organization Name and Address Rice University Houston, TX 77001</p>		<p>11. Contract or Grant No. 15) DOT-CG-81-75-1383 NEW</p>	<p>13. Type of Report and Period Covered 9) FINAL REPORT</p>
<p>12. Sponsoring Agency Name and Address Department of Transportation U. S. Coast Guard Office of Research and Development Washington, DC 20590</p>		<p>14. Sponsoring Agency Code</p>	
<p>15. Supplementary Notes The contract under which this report was submitted was under the technical supervision of the Coast Guard Research and Development Center, Groton, Connecticut. R&D Center Report Number 12/77 has been assigned.</p>			
<p>16. Abstract The infrared spectra of oils have many sharp peaks and good identification results from reading just the amplitudes of these peaks. Indeed, about fifteen properly chosen peaks will usually suffice. The Coast Guard contracted with Rice University to set up identification procedures based on ultraviolet fluorescence spectra of oils, and to determine whether or not techniques used for infrared analysis could be applied to interpretation of fluorescence spectra. (The data base used for this purpose is given in the body of this report.) (The basic result is that simply reading peak amplitudes of fluorescence spectra will not suffice to give good identification. The same conclusion holds if one also reads the amplitudes at the valleys in between the peaks. However, by assigning a sharpness factor to valleys and peaks, one can extract sufficient information to result in very reliable identification. An algorithm for doing this has been devised and implemented in the form of a computer program. A second purpose of the contract was to test the independence of the infrared and fluorescence spectra. (The data base used for this is given in the body of the report.) More precisely, if interpretation of infrared spectra gives a probability P_1 that spill and suspect sample are the same, and the interpretation of fluorescence spectra gives a probability of P_2, what is the correct formula for combining P_1 and P_2 to obtain the "combined" probability of a match? Histograms show evidence of a considerable degree of independence but cannot suffice to give the desired formula. Scatter diagrams and correlation coefficients offer further evidence of independence, but a detailed analysis of interspectral distance would be required to determine a specific formula for combining P_1 and P_2. A conservative formula is given based on the limited information currently available.</p>			
<p>17. Key Words pattern recognition, oil spill identification, fluorescence spectroscopy, infrared spectroscopy</p>		<p>18. Distribution Statement Document is available to the U. S. public through the National Technical Information Service, Springfield, Virginia 22161</p>	
<p>19. Security Classif. (of this report) UNCLASSIFIED</p>	<p>20. Security Classif. (of this page) UNCLASSIFIED</p>	<p>21. No. of Pages 21</p>	<p>22. Price</p>

305 750

mt

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures				Approximate Conversions from Metric Measures			
Symbol	When You Know	Multiply by	To Find	Symbol	When You Know	Multiply by	To Find
LENGTH							
in	inches	2.5	centimeters	mm	millimeters	0.04	inches
ft	feet	30	centimeters	cm	centimeters	0.4	inches
yd	yards	0.9	meters	m	meters	3.3	feet
mi	miles	1.6	kilometers	km	kilometers	0.6	miles
AREA							
in ²	square inches	6.5	square centimeters	cm ²	square centimeters	0.16	square inches
ft ²	square feet	0.09	square meters	m ²	square meters	1.2	square yards
yd ²	square yards	0.8	square meters	km ²	square kilometers	0.4	square miles
mi ²	square miles	2.6	square kilometers	ha	hectares (10,000 m ²)	2.5	acres
MASS (weight)							
oz	ounces	28	grams	g	grams	0.035	ounces
lb	pounds	0.45	kilograms	kg	kilograms	2.2	pounds
	short tons (2000 lb)	0.9	tonnes	t	tonnes (1000 kg)	1.1	short tons
VOLUME							
tsp	teaspoons	5	milliliters	ml	milliliters	0.03	fluid ounces
Tbsp	tablespoons	15	milliliters	l	liters	2.1	pints
fl oz	fluid ounces	30	milliliters	qt	quarts	1.06	gallons
c	cups	0.24	liters	l	liters	0.26	gallons
pt	pints	0.47	liters	m ³	cubic meters	35	cubic feet
qt	quarts	0.95	liters	m ³	cubic meters	1.3	cubic yards
gal	gallons	3.8	liters				
ft ³	cubic feet	0.03	cubic meters				
yd ³	cubic yards	0.76	cubic meters				
TEMPERATURE (exact)							
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature



*1 in = 2.54 (exact). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price \$2.25, SD Catalog No. C13.10.286.

TABLE OF CONTENTS

	<u>Page</u>
1.0 DATA BASES USED	1
2.0 DIGITIZING DATA	1
3.0 CLUSTERING	3
4.0 ASSUMPTIONS FOR PROBABILITY COMPUTATION	3
5.0 THE ALGORITHM	4
6.0 THE COMPUTER PROGRAM	5
7.0 INDEPENDENCE OF DATA	5
7.1 Data Base Used	6
7.2 Experiments Performed to Determine the Degree of Independence	8
REFERENCES	16

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1	Histogram of Distances in Infrared Domain	9
2	Histogram of Distances in Fluorescence Domain	10
3	Scatter Diagram of Distances in Fluorescence and Infrared Domains	11
4	Histogram of Distance Ratio	13
5	Distances Between Means for Various Oil Classes	14

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Oils Used in Study	2
2	Frequencies at Which Samples Were Taken	7

NTIS		White Section	<input checked="" type="checkbox"/>
DDC		Buff Section	<input type="checkbox"/>
UNANNOUNCED			<input type="checkbox"/>
JUSTIFICATION			
BY			
DISTRIBUTION/AVAILABILITY CODES			
Dist. AVAIL. and/or SPECIAL			
A			

1.0 DATA BASES USED

The first phase of this study involved generating an algorithm for giving a probability that two oil samples (the suspect sample and the spill sample) are the same. This was done using fluorescence spectra of some 230 oil samples furnished by the Coast Guard Research and Development Center. These include crude oils, heating oils and lubricating oils.

The second phase of this study involved independence of infrared and fluorescence spectra of oils. For this phase we had infrared and fluorescence spectra of 30 oil samples. The fluorescence spectra were furnished by the Coast Guard R&D Center and the infrared spectra were furnished by J. Mattson. These oils are listed in Table 1.

2.0 DIGITIZING DATA

A spectrum S is the graph of a function defined on some interval on the real line. To do computations we need to approximate the spectrum S by a vector in some Euclidean space R^n ; i.e., with an ordered n -tuple X of real numbers (1). This should be done in a manner dictated by the computations to be made. In general we would like to make n as small as we can without losing so much information that our computational goal cannot be achieved. Thus we want the numbers going into our n -tuple X to contain the kind of information we seek.

Here is a specific example. Fluorescence spectra of oils decay exponentially after some wavelength (e.g., about 380 nm for crudes). Now an exponential curve is completely determined by knowing two points on it. So if we read amplitudes at more than two points in the exponential tail-off region, we are increasing n without getting any more information for identification purposes.

Similarly, if we know that between a peak and an adjoining valley the spectrum is essentially linear, we get no information by reading an amplitude in this linear region.

Now the ordered n -tuple X of real numbers does not have to be simply the amplitudes of S read at certain wavelengths. Any real number assignment which gives good differences for some different samples is a candidate for a component

TABLE 1

OILS USED IN STUDY

OIL NUMBER	COAST GUARD DESIGNATION
1	AM032F02
2	AM035F02
3	AM013F02
4	AR022F02
5	HE005F02
6	GU031F02
7	GU011CVE
8	EX045F04
9	AM029F05
10	EX017F05
11	GU032F05
12	GU058F05
13	AM025F06
14	EX007F06
15	IN001F06
16	CO022F06
17	GU034F06
18	GU033F06
19	UN013D02
20	CO023D02
21	SH038D00
22	AM016D00
23	UN002D00
24	EX044D00
25	GE008D0D
26	EP021DOX
27	PEGU075MOD
28	IN003MOD
29	EX006MDF
30	EX059F04

of X (2). For example, one may assign a real number describing sharpness of a peak or valley. But, of course, the i^{th} coordinate of X must be measuring the same thing for each spectrum.

For fluorescence spectra of oils we have the exponential tail-off and we have essential linearity between a peak and adjacent valley. Also amplitude readings at all peaks and valleys is insufficient for we would have spectra $S_1 \neq S_2$ for different spectra and yet the corresponding vectors X_1, X_2 are almost identical. Thus components for sharpness must be added in.

3.0 CLUSTERING

In the 230 fluorescence spectra, assigning amplitudes for all valleys and peaks and assigning sharpness factors for all valleys and peaks gives for each spectrum S a vector X in 68-dimensional space, $X \in R^{68}$. Since probabilistic calculations involve measuring a large number of distances between vectors, they would be beyond the capabilities of most computers with 230 vectors in R^{68} .

However, examination of these spectra shows that they naturally fall into 33 groups or clusters: the intercluster distances being great compared with intracluster distances. (For a review of cluster analysis, see reference (3).) Now, within a given cluster, some peaks and valleys are completely absent and readings taken there would be wasted. Thus, for a given cluster, we can greatly reduce the dimension of the vector X assigned to a spectrum. Of course we still need sharpness factors, because the problem of $S_1 \neq S_2$ but $X_1 = X_2$ (when not using sharpness) quite naturally occurs for S_1 and S_2 in the same cluster. The dimension of vectors X which give good separation within a cluster was about 12 to 14.

4.0 ASSUMPTIONS FOR PROBABILITY COMPUTATION

1. Reproducibility of spectra is sufficiently accurate that a spectrum may be considered a point in R^n .

2. The library is sufficiently representative so that for any given oil we can unambiguously say that it is "essentially like" exactly one in the library.

3. The clusters have been chosen so that there is essentially zero probability of any two oils being the same if they fall into different clusters. Also we assume that assigning any oil to the cluster with the nearest centroid gives the right assignment.

With these assumptions the algorithm given in the next section gives a combined probability satisfying the standard conditions required of a probability function. These standard conditions are listed in any text on probability theory such as references (4) or (5).

5.0 THE ALGORITHM

The probability that two unknowns X and Y (spill and suspect) are the same is obtained by combining two probability factors. (This is all for fluorescence -- not to be confused with our later combinations of probabilities for fluorescence and infrared.)

Let N be the number of spectra in the library (about 230). Let C_1, C_2, \dots, C_{33} denote the number of elements in the clusters (so that $C_1 + C_2 + \dots + C_{33} = N$). Let C_i be the cluster containing X. If Y is also in C_i we set

$$P_i = \frac{N - C_i}{N}$$

and if Y is not in C_i we use zero. (This slight lack of symmetry in treatment of X and Y is unimportant, for if X and Y are in different clusters, the probability of identity is quite low).

If X and Y are not in the same cluster, we say the total probability of identity is 0. If X and Y are in the same cluster (say the i^{th}) we calculate a second probability factor \hat{P}_2 as follows: We now represent all vectors in the i^{th} cluster as vectors in the space of dimension 12 to 14 used for that cluster. Let X' and Y' now denote the spill and suspect in this space. Let n_x be the number of vectors in the i^{th} cluster within a distance $(d(X', Y'))$ of X'. Define n_y similarly. Set

$$\hat{P}_2 = \frac{C_i - \frac{n_x + n_y}{2}}{C_i}$$

The final probability then that $X = Y$ is given by:

$$P = 1 - (1 - \hat{P}_1)(1 - \hat{P}_2)$$

6.0 THE COMPUTER PROGRAM

The computer program on punched cards has been sent to the Coast Guard R&D Center in Groton. It suffices to say here that the program will: (1) compute the sharpness factors needed to augment the amplitude readings, (2) do linear interpolation and exponential interpolation to generate the 68-dimensional vectors for the entire library, (3) decide (using these 68-dimensional vectors) to which clusters X and Y belong, and (4) carry out the probability computation as described above in the algorithm.

On several unknowns this program proved completely successful. But for some different library of spectra, it should not be used without some refinements which were not needed for our library of fluorescence spectra.

7.0 INDEPENDENCE OF DATA

The purpose of this study was very limited and specific. Our objective was to investigate the degree of independence or, conversely, dependence between data obtained from the infrared and fluorescence spectra of oils. Other work has been done on this problem by Killeen and Chien (6).

The motivation behind such a study is quite simple. Various procedures exist for determining the probability of a match, P_1 , between a spill sample and a source sample using data obtained from the fluorescence spectra of oils (7). In addition, procedures have been developed for determining the probability of a match, P_2 , using data obtained from the infrared spectra of oils (8). What is needed is a method for combining the individual probabilities of a match into an overall probability of a match, P_m . Obviously, such a probability would be a more accurate and reliable measure of the probability of a match than either of the individual probabilities, for it incorporates more information about the problem at hand.

In order to develop the means for combining the individual probabilities, an investigation of the independence of the two types of data, that obtained

from the infrared spectra and that obtained from the fluorescence spectra, must be made. This paper describes the work which has been conducted at Rice University to accomplish this end.

7.1 Data Base Used

The data base used in this investigation came from 30 different types of oils and was supplied by the U. S. Coast Guard R&D Center in Groton, Connecticut. These oils are listed in Table 1. The oil types broke down as follows: six No. 2 fuel oils, two No. 4 fuel oils, four No. 5 fuel oils, six No. 6 fuel oils, two No. 2 diesel oils, nine diesel oils (unspecified), and one crude oil from Venezuela.

The infrared spectra were obtained using a Perkin-Elmer Model 180 infrared spectrophotometer (with cells of pathlengths 0.12 ± 0.03 mm) which used electronic ratioing (rather than usual optical null). The spectrophotometer had an ordinate resolution (in %T) of one part in 12,000 and an abscissa resolution (in wavenumbers) of 0.01 cm^{-1} . It was interfaced with a Data General NOVA 1200 minicomputer.

The fluorescence spectra were obtained from oil samples diluted to 100 ppm in cyclohexane (wt/wt) run on an uncorrected Perkin-Elmer MPF-3 spectrofluorometer with excitation at 254 nm and bandwidths: excitation, 34 nm and emission, 1.5 nm.

For each of the oils in the data base, samples of the amplitudes of the infrared spectra at various frequencies were determined. In addition, measurements of the amplitudes of the fluorescence spectra at selected frequencies were supplied. The frequencies used for the infrared study are given in Table 2.

By digitizing or sampling each spectra at a number of frequencies, one can get a series of real numbers representative of each spectra. In the present study, 20 measurements of the infrared and 20 measurements of the fluorescence spectra of each oil were made. These measurements can be listed in vector form. We shall let the vectors $V_i = (v_{i1}, v_{i2}, \dots, v_{i20})$ and $W = (w_{i1}, w_{i2}, \dots, w_{i20})$ represent the infrared and fluorescence spectra of the i -th oil of the data base. In the following section, the experiments conducted to determine the degree of independence between the two types of data are discussed.

TABLE 2

FREQUENCIES AT WHICH SAMPLES WERE TAKEN

VARIABLE NUMBER	FREQUENCY USED FOR INFRARED	FREQUENCY USED FOR FLUORESCENCE
1	1629 cm^{-1}	290 nm
2	1603	295
3	1304	300
4	1166	305
5	1154	310
6	1032	315
7	963	325
8	918	335
9	888	345
10	870	355
11	846	365
12	832	375
13	809	385
14	793	395
15	781	405
16	765	415
17	741	425
18	722	435
19	697	445
20	673	455

7.2 Experiments Performed to Determine the Degree of Independence

Since both types of spectra are given rise to by the basic physical properties of the oil, we expect there to be some degree of dependence. However, we also expect that since the infrared and fluorescence spectra are actually measurements taken at different domains of the physical spectrum, there should be some independence between these two types of spectra. To begin with, the Euclidean distance between each pair of infrared vectors was calculated. We represent the distance between vectors V_i and V_j as d_{ij} . Similarly, the pairwise distance between vectors in the fluorescence domain were calculated and are represented by d_{ij}^* . The distances are important in the sense that in estimating the probability density structure of the data, the distance between vectors representative of the underlying random process is important (9). In addition, most of the algorithms which exist for determining a match between a spill and a source are based on minimum distance or nearest neighbor arguments.

As a first step of analysis, the histogram of the distances for each type of spectra was prepared. These histograms are listed in Figure 1 and 2. Comparison of the two histograms is interesting. The infrared histogram peaks for a value of distance of 0.2 and then decays in a generally exponential fashion. In contrast, the fluorescence histogram consists of a number of local minima and maxima. It also contains more of its area for higher values of distance than the infrared histogram. Due to the discrepancy in the shape of these two histograms, one is led to believe there is some degree of independence between the two types of data considered.

In order to gain further insight into the structure of the data, a scatter diagram of the distance data was prepared and is shown in Figure 3. This diagram plots the value of distance in the fluorescence domain vertically versus distance in the infrared domain horizontally. If a true linear relationship existed between the values of d_{ij} and d_{ij}^* , we would see a bunching of the scatter diagram points along the regression line of d_{ij} on d_{ij}^* . If a simple nonlinear relationship existed, we would expect to see this displayed as bunching along a curve on the scatter diagram. However, we see neither of the above in the figure. Instead, we see a general smearing of the data points with some

Relative Frequency (Percent)

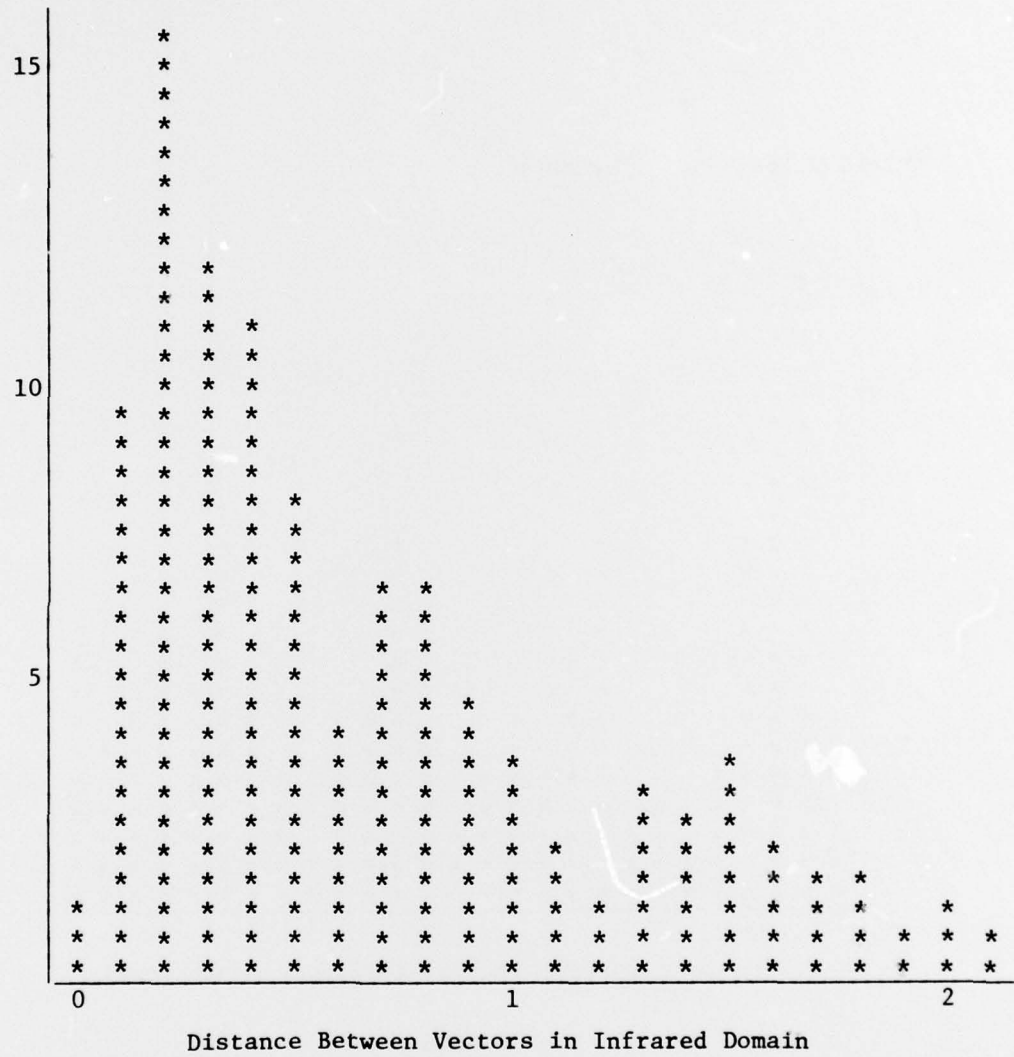


FIGURE 1

HISTOGRAM OF DISTANCES IN INFRARED DOMAIN

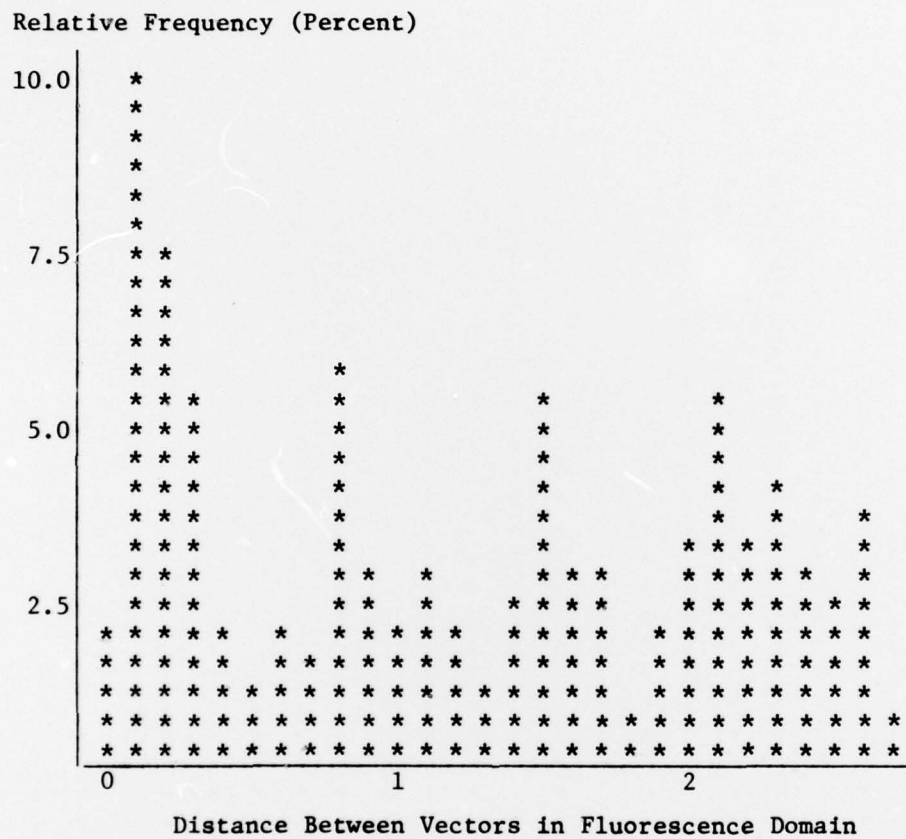


FIGURE 2

HISTOGRAM OF DISTANCES IN FLUORESCENCE DOMAIN

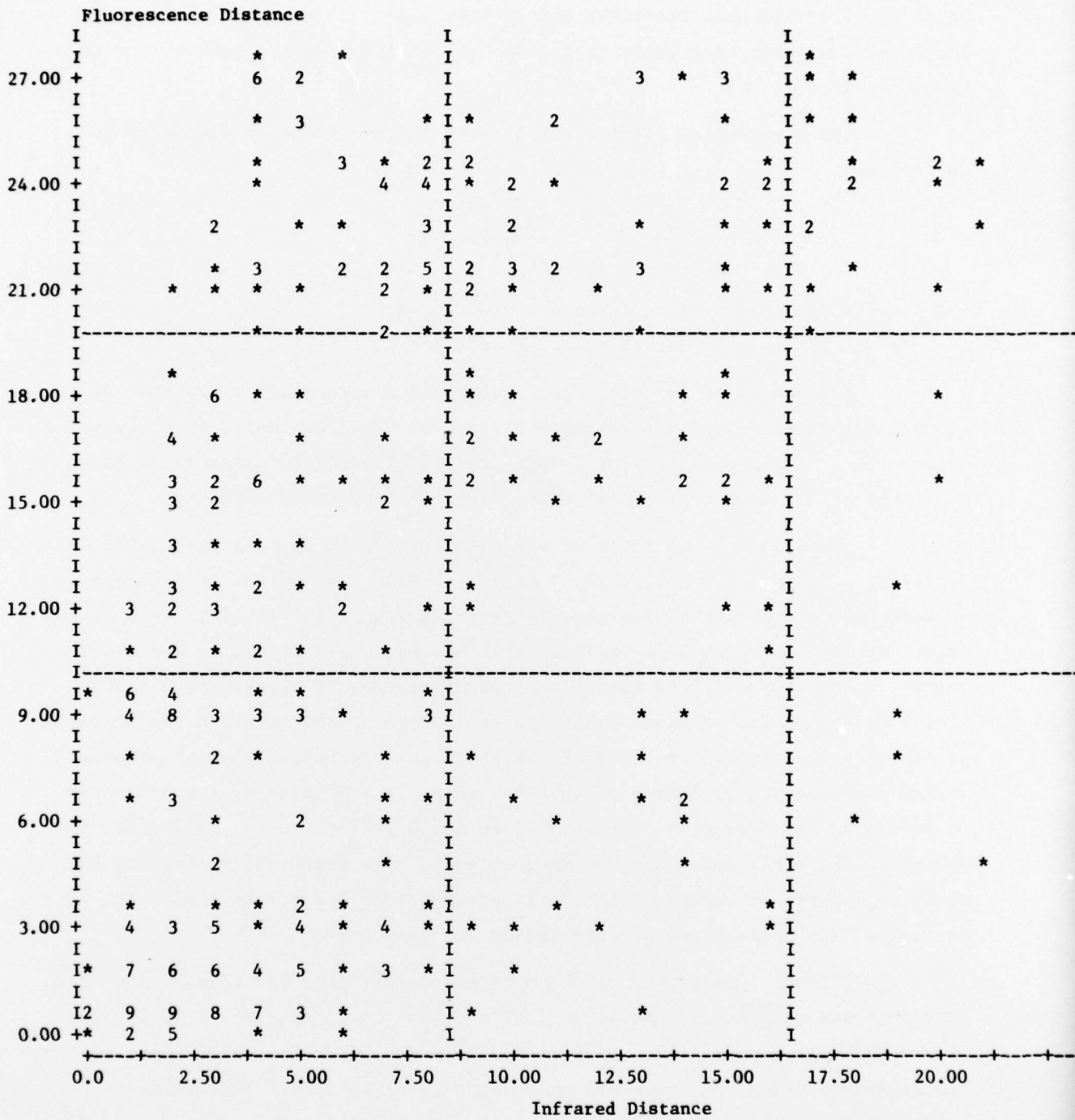


FIGURE 3

SCATTER DIAGRAM OF DISTANCES IN FLUORESCENCE AND INFRARED DOMAINS

bunching of points near the lower end of each axis. This general type of structure indicates that there is a good deal of independence between the distance calculated.

The correlation coefficient between the two types of distances is defined as:

$$\rho = \frac{\sum_{i=1}^{29} \sum_{j=i+1}^{30} (d_{ij} - \bar{d})(d_{ij}^* - \bar{d}^*)}{(\text{var } d)^{1/2} (\text{var } d^*)^{1/2}}$$

where \bar{d} is the mean of the infrared distances and \bar{d}^* is the mean of the fluorescence distances.

For values of $|\rho|$ close to 1, there is a strong linear relationship between the infrared and fluorescence distances. For $|\rho|$ small, there is not a strong linear dependence present. For the data considered, ρ was found to be 0.47, which indicates that there is not a strong linear relationship.

The value of the ratio of the distance d_{ij}^* to d_{ij} as calculated is represented by r_{ij} ($i = 1, 2, \dots, 29$; $j = i+1, \dots, 30$). The values of r_{ij} were histogrammed and the resulting plot is given in Figure 4. If there were a high degree of dependence between the two variables, we would expect to see a high degree of structure in this histogram. For instance, if there were a highly linear relationship, we would expect to see a high narrow spike in the histogram of the distance ratio. In contrast, if the data were completely independent, we would expect a high degree of uniformity in the amplitude of the histogram. By examining the histogram of Figure 4, we see a good degree of uniformity for values of the ratio less than 2. We then see a high degree of uniformity but of a lower level of values of the ratio greater than 2 but less than 7.5. This indicates that these types of data are fairly independent.

Further information can be obtained by dividing the oils of the data base into six categories: No. 2 fuel oils, No. 4 fuel oils, No. 5 fuel oils, No. 6 fuel oils, crude oils, and diesel oils. The mean of each of these classes was estimated using the fluorescence and the infrared data. The distance between the means for each pair of classes was calculated for each type of data. The results are presented in Figure 5. It is readily seen that the classes with

Relative Frequency (Percent)

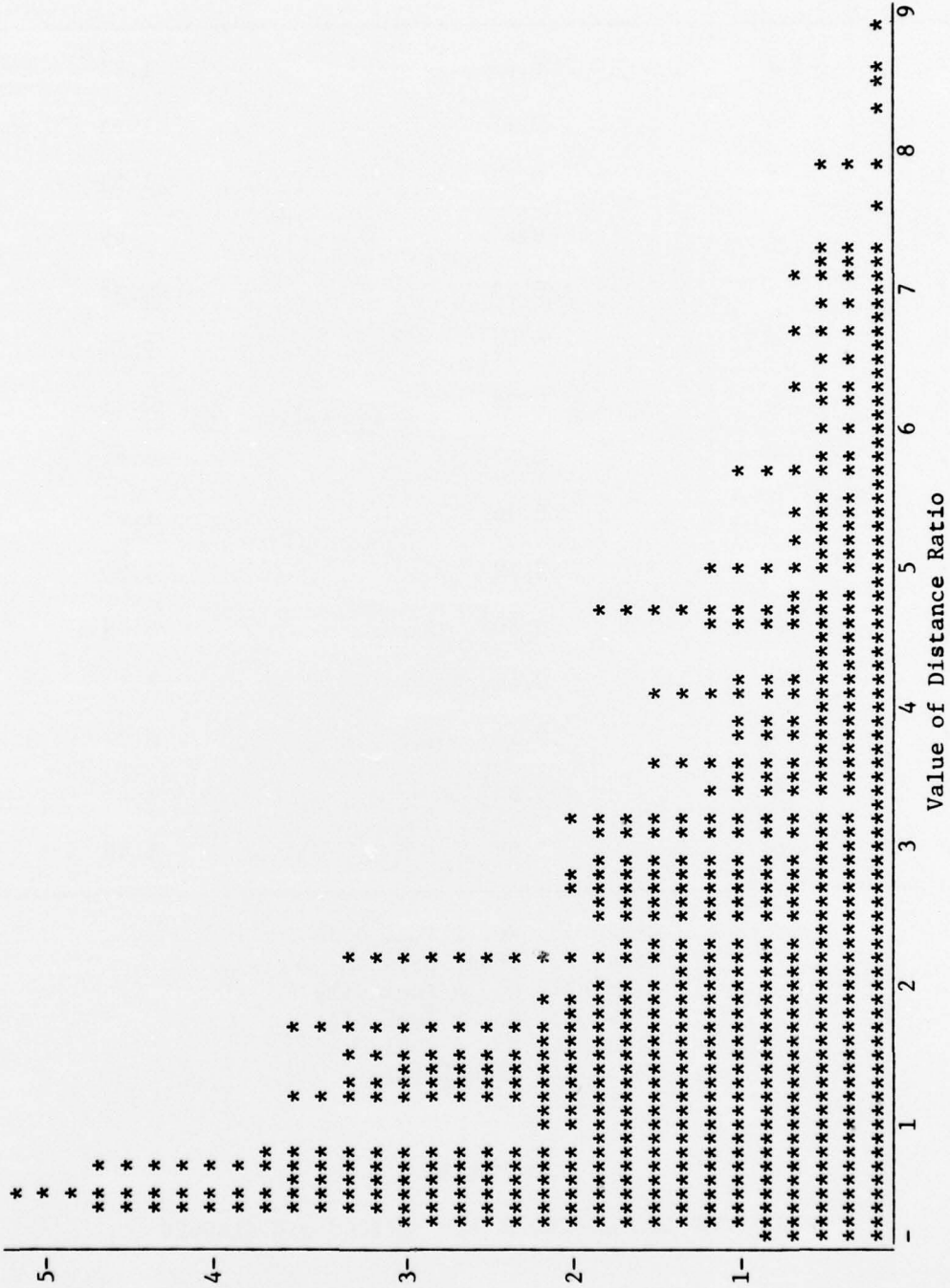


FIGURE 4
HISTOGRAM OF DISTANCE RATIO

CLASS	CLASS	DISTANCE (INFRARED)	DISTANCE (FLUORESCENCE)
1	2	0.43	1.43
1	3	0.27	1.53
1	4	0.37	1.53
1	5	0.48	1.49
1	6	0.23	0.21
2	3	0.42	0.20
2	4	0.58	0.25
2	5	0.67	0.21
2	6	0.36	1.52
3	4	0.19	0.20
3	5	0.30	0.19
3	6	0.48	1.63
4	5	0.14	0.27
4	6	0.57	1.61
5	6	0.69	1.49

Class 1: No. 2 Fuel Oils
 Class 2: Crude Oils
 Class 3: No. 4 Fuel Oils
 Class 4: No. 5 Fuel Oils
 Class 5: No. 6 Fuel Oils
 Class 6: Diesel Oils

FIGURE 5
 DISTANCES BETWEEN MEANS FOR VARIOUS OIL CLASSES

means relatively close under fluorescence data are not necessarily close under infrared and vice versa. This indicates a fairly high degree of data independence.

Based on this data we feel that the formula

$$P_m = 1 - 2(1-P_1)(1-P_2)$$

for combining P_1 and P_2 is quite conservative; i.e., it will not give too high a value top. However, it will require much more computation before we would feel justified in lowering the "2" in the formula.

REFERENCES

- (1) Patrick, E. A., Fundamentals of Pattern Recognition, Englewood Cliffs, New Jersey, Prentice Hall, 1972.
- (2) Meisel, W. S., Computer-Oriented Approaches to Pattern Recognition, New York, Academic Press, 1972.
- (3) Hartigan, J. A., Clustering Algorithms, New York, Wiley, 1975.
- (4) Ash, R., Basic Probability Theory, New York, Wiley, 1970.
- (5) Papoulis, A., Probability, Random Variables, and Stochastic Processes, New York, McGraw Hill, 1965.
- (6) Killeen, T. J. and Chien, Y. T., "A Probability Model for Matching Suspects With Oil Spills or Did the Real Spiller Get Away?", presented at the Workshop on Pattern Recognition Applied to Oil Identification, Coronado, California, November 1976.
- (7) Starks, S. A. and Curtis, M. L., "Use of Ultraviolet Spectroscopy in Oil Spill Identification," presented at the Twenty-Seventh Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy, March 1976.
- (8) Mattson, J. S., Mattson, C. S., Spencer, M. J., and Starks, S. A., "Multivariate Statistical Approach to Fingerprinting of Oils By Infrared Spectroscopy," Anal. Chem., 49, 297-302, 1972.
- (9) Duda, R. O. and Hart, P. E., Pattern Classification and Scene Analysis, New York, Wiley, 1973.

☆U.S. GOVERNMENT PRINTING OFFICE: 1977-701-393/100