"Made (E80-10206) DESIGN SPECIFICATION FOR N80-29785
in the EQUIPROBABLE BLOCKS DENSITY
semina ESTIMATORY/CLASSIFIER/DOT SELECTOR (Lockheed
Progra Electronics Co.) 19 p HC A02/MF A01 Unclas
for any CSCL 05B G3/43 00206

160637

NASA CR-

DESIGN SPECIFICATION

FOR

SECONDARY ERROR SOURCES

EQUIPROBABLE BLOCKS

DENSITY ESTIMATOR/CLASSIFIER/DOT SELECTOR

Job Order 71-695

(TIRF 77-0058)

Prepared By

Lockheed Electronics Company, Inc.

Systems and Services Division

Houston, Texas

Contract NAS 9-15200

For

EARTH OBSERVATIONS DIVISION

SPACE AND LIFE SCIENCES DIRECTORATE

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

December 1977

LEC- 11676

JSC- 13824

DESIGN SPECIFICATION
FOR
EQUIPROBABLE BLOCKS
DENSITY ESTIMATOR/CLASSIFIER/DOT SELECTOR

Job Order 71-695

(TIRF 77-0058)

Prepared By

P.J. Aucoin Jr.

P. J. Aucoin, Jr.
Lockheed Electronics Company

APPROVED BY

Philip L. Krumm

P. L. Krumm, Supervisor
Applications Software Section

B L Carroll

B. L. Carroll, Manager
LACIE Development and Evaluation Dept.

Prepared By

Lockheed Electronics Company, Inc.

For

Earth Observations Division

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

December 1977

LEC- 11676

CONTENTS

1. SCOPE

This document contains a design specification for software to implement an equiprobable blocks program system. This system will, on option, perform the following functions.

a. Conditional probability functions

b. Classification

c. Dot selection

d. Block population comparison

Training fields will be user-provided. Test and sample fields, also user-provided, are used in classification and comparison options.

The equiprobable blocks program will be added as a processor to the EOD-LARSYS pattern recognition system. It will be implemented on the Purdue-LARS computer (IBM 370/148). (Conversion of the EOD-LARSYS system from the UNIVAC EXEC 2 system at JSC to Purdue-LARS is currently under way.)

2. APPLICABLE DOCUMENTS

- Final Design Specification for EOD-LARSYS Procedure 1, Houston, Texas, August 1977, JSC-12742, LEC-10417.
- Job Order 63-1347-1695
- TIRF
- M. P. Gessamon, "A Consistent Non-Parametric Multivariate Density Estimator Based on Statistically Equivalent Blocks", Ann. of Math Stat., Vol. 41, No. 4, pp. 1344-1346, 1970.
- M. P. Gessamon and P. H. Gessamon, "A Comparison of Some Multivariate Discrimination Procedures", J. of Am. Stat. Assoc., Vol. 67, No. 338, pp. 468-472, 1972.

3. SYSTEM DESCRIPTION

3.1 HARDWARE DESCRIPTION

N/A

3.2 SOFTWARE DESCRIPTION

The Equiprobable Blocks processor will proceed computationally along the following path.

Read block parameters, action options, features, spatial grid and parameter D (option). Read training fields for each category (class) defined. Read test fields, sample fields. For each category, build the set of equiprobable blocks. This procedure is described in detail later in this section. On option, read a ground truth file, assign categories to training field pizels.

Depending upon action options, perform one or more of the following:

a. Compute conditional probability density estimate, $f(z|C_j)$, j=1, NC where z belongs to one of the training fields of category C_j and C_j is the jth category. Output these estimates along with training field geometry, features, category names, and so on.

b. Classify the sample fields. This procedure is described in detail later in the section. If there are more than two categories defined, successive NC-1 sets of blocks are combined to obtain the "best" set of blocks for classification. Write a classification map file (MAPTAP).

c. Select dots based on an input spatial grid and maximum count parameter D. The equiprobable block set used will be a combination of blocks of all categories specified. Output the dots selected.

d. Combine equiprobable block sets from blocks from each category to form a combination block set. Compute the proportion of pixels per category per block and output these results.

In all cases where test fields are input, compute the number of pixels per category per block and output. In this way, if training fields for categories are also provided as test fields, a comparison of categories can be achieved.

Certain details of the computation procedure follow.

COMPUTATIONAL PROCEDURE: BLOCKS

Read training field pixels per category, sun angle correct on option.

Construct a set of equiprobable blocks for each category j, j=1, NC:

First compute block parameters

Option - compute

$$K_j = \mathrm{FIX}\left(M_j \Big/ \prod_{i=1}^{P} Q_{ij} \right)$$

P = number of selected features

where M_j = number of pixels in training fields for category j

K_j = minimum number of pixels per block

Q_{ij} = number of subdivisions in each co-ord direction i for each category j

Option - compute

$$Q_{ij} = \mathrm{FIX}\left(\left(\frac{M_j}{K_j}\right)^{\frac{1}{P}}\right)$$

Option - compute

$$K_j = \mathrm{FIX}\left(M_j^{\frac{1}{1+P}}\right)$$

$$Q_{ij} = \mathrm{FIX}\left(\left(\frac{M_j}{K_j}\right)^{\frac{1}{P}}\right)$$

where $Q_{ij} = Q_j$ (Q_{ij} independent of i)

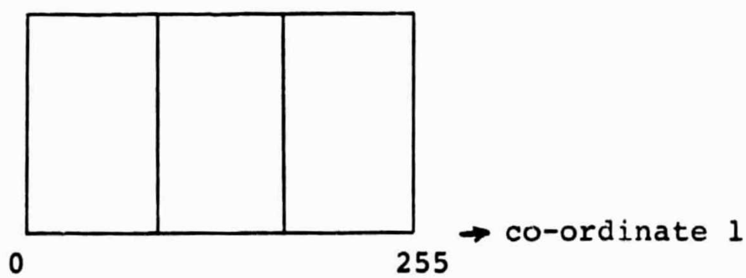
Secondly, determine cuts in each direction i (i.e., block boundaries)

Compute $r_{1j} = \mathrm{FIX}\left(\frac{M_j}{Q_{ij}}\right)$

$$S_{1j} = M_j - r_{1j}Q_{1j}$$

The first S_{1j} block divisions will have $r_{1j} + 1$ pixels, the last $Q_{1j} - S_{1j}$ block divisions will have r_{1j} pixels, where the right-hand boundary pixels are included in the block division to the left.

Example
3 block divisions, (strips) first co-ordinate



$$Q_{1j} = 3,\ M_j = 7,\ r_{1j} = 2,\ S_{1j} = 1$$

In order to know which pixels go into the block divisions, an ordering on the first co-ordinate must be made. This will be the time-consuming part of the computation. The pixels involved will belong to the training fields of category j.

This ordering process also determines the co-ordinate valve of the cut. The cuts must be made so that no two are at the same location. (The original requirements document resolves this.)

i=2: $X = r_{2j}Q_{2j} + S_{2j}$

where $X = r_{1j} + 1$ for the first S_{1j} pixels

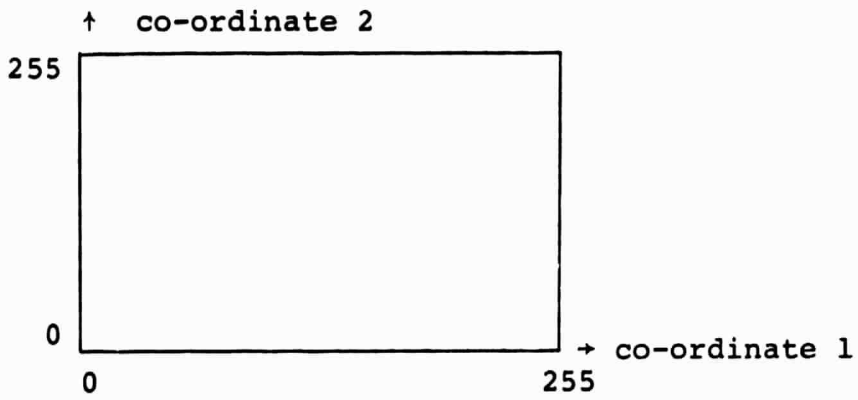
$= r_{1j}$ for the others

Compute $r_{2j} = \text{FIX}\left(\frac{X}{Q_{2j}}\right)$

$S_{2j} = X - r_{2j}Q_{2j}$

The first S_{2j} secondary divisions in each strip will contain $r_{2j} + 1$ pixels, (upper or "right" boundaries included), the rest will contain r_{2j} pixels.

Example



i=3, . . . P

This process should be continued for i=3, . . .P. At each stage, the pixels must be ordered in the ith co-ordinate (channel) for block division inclusion determination.

There will be $\prod_{i=1}^{P} Q_{ij}$ blocks at the conclusion.

The volume of each block must be computed and stored. These are computed from block boundaries which have been stored.

Once the blocks have been formed for each category, the options are exercised.

Option to compute conditional probability density estimates:

$$f(Z|c_j) \sim \frac{1}{(M_j+1)\,VOL_j}$$

where: Z belongs to the sample fields,

M_j = number of pixels in category j training fields,

VOL_j = volume of the block of category j containing the pixel Z.

Output these estimates.

Option to classify the test fields:

The "best" set of blocks must first be determined. In the case where only 2 categories are specified, the best set is determined as follows.

Using the training fields of category 2 as test fields, classify these fields using the blocks of category 1. Compute the proportion of pixels misclassified. Next, take the training fields of category 1 as test fields and classify using the blocks of category 2. Compute the proportion misclassified. Select the block set yielding the smaller proportion of misclassification for use in classification of sample fields.

In the case of more than 2 categories (NOCAT>2), successive combinations of NOCAT-1 block sets are used to classify the remaining category training fields and the proportion of misclassification is computed. The block combination of blocks from NOCAT-1 block sets yielding the smallest proportion of misclassification is used in classification.

The process of classification is as follows. Given a set of blocks and sample fields, for each pixel Z in the sample fields, determine which block contains it. Next, for each category, determine the proportion of pixels from the training fields in that block. The category having the largest proportion will claim the pixel Z. In case of ties, pixels will be allocated among the categories involved in the ties. These classification results will be outputted under the EOD-LARSYS MAPTAP format for subsequent display.

Option to select dots:

Here the combined set of blocks over all categories is utilized. Normally, only one category will have been defined under this option. (The equiprobable blocks will be built from all the training fields.) The sample field provided will be overlaid with the spatial grid provided. The D (or fewer) grid intersections contained in each block will be selected. If there are more than D intersections in a particular block, the D dots closest (using L_2 distance) to the center of the block will be chosen. If no grid is provided, the D pixels closest to the center of each block will be selected.

The spatial co-ordinates of the set of selected dots will be outputted.

Option to combine category training sets:

All training fields will be combined to form an equiprobable block set. The proportion of pixels in each block from each training category will be determined and outputted. Also, the proportion for test field categories will be determined and outputted.

3.2.1 LINKAGES

The Equiprobable Blocks processor will use the FORTRAN IV-G compiler, IBM system routines, EOD-LARSYS utility routines, and EOD-LARSYS common blocks GLOBAL, ISOLNK and INFORM. A new common block EQUPRB will be created for other transfers of data among subprograms.

3.2.2 INTERFACES

The Equiprobable Blocks processor will require as an input file an MSS data tape (ERIPS-type unload tape). It will be invoked from the MONTOR routine of EOD-LARSYS upon encountering a control input card of the form $EQUPRB. Back-to-back runs with the DISPLAY processor of EOD-LARSYS will be possible.

3.2.3 INPUTS

a. Processor card

Keyword

$EQUPRB

b. New Control Cards

Keyword	Parameter	Function
CHANNEL	DATA=$C_1, C_2, \ldots C_{NOFET2}$ (Default: 1,2,...,16)	Channel numbers to be used. These must be a subset of channels available on the MSS data tape.
NCPASS	N (Default: N=4)	Number of channels per pass. Used to group channels into passes (acquisitions).
SUNANG	TAPE or $S_1, \ldots S_{NA}$	Sunangles for sun angle correction. If TAPE parameter specified, sunangles are extracted from the MSS Universally formatted data tape.
DATAFI	INPUT/UNIT=N,FILE=M (Default: N=3, M=1)	Unit and file numbers for an MSS data tape.
FORMAT	U or F	Universal or LARSYS III
MAPTAP	OUTPUT/UNIT=N,FILE=M (Default: N=2, M=1) (Card ignored if classification option not specified)	Unit number and beginning file number for the optional classification map file group.
DSTTAP	OUTPUT/UNIT=N,FILE=M (Default: None) (Card ignored if DST output option not invoked)	Unit number and beginning file number for the optional output of conditional probability density estimates.

Keyword	Parameter	Function
GTTAP	INPUT/UNIT=N,FILE=M	Unit and file numbers of ground truth tape
GTVOTE	N	N=1 designator for strategy (to be provided) for assigning pixel category number from subpixel category numbers (used if numbers are not all the same)
GTNAM	N_1, $CATNAM_1$,... $N_{N\phi CAT}$, $CATNAM_{N\phi CAT}$	Category names corresponding to ground truth numbers

Keyword	Parameter	Function
CATNAM	Category $name_1, \ldots,$ category $name_{NOCAT}$	Category names. These names should match those specified in CLASSNAME field description cards appearing after control card input.
BLOCK	$Q/\{Q_{ij}\}$ i=1,NOFET2, j=1,NOCAT	Q block option.
	or	
	$K/\{K_j\}$ j=1,NOCAT	K block option.
	If card omitted or parameter field blank, default block option is used.	
DOTGRD	$X/x_1, \ldots, x_{DOTX}$ $Y/y_1, \ldots, y_{DOTY}$ (Default: Pick D pixels closest to the center each block).	Spatial dot grid. If dot option not invoked, this card(s) is ignored.
DOTMAX	D (Default: D=1)	Maximum number of dots per block.
OPTION	CNDPRB	Compute conditional probabilities.
	CLASS	Compute conditional probabilities and classify test pixel fields.
	DOTS	Select dots using spatial grid overlay on test field.
	GT	Use ground truth information to label training field pixels.

Keyword	Parameter	Function
	COMBIN	Combine all category training fields to make a combined block set. Compute proportion of pixels per category per block for training fields, test fields.
	PASS	Perform all operations separately by pass, create output files by pass. Used for providing input to the multi-temporal Bayes classifier. The number of acquisitions is computed from CHANNEL and NCPASS control card information.
	TEST	Create blocks for each category. Do not compute conditional probability density estimates. Compute proportion of pixels in test fields in each block of each category. Also, include pixel category numbers in appropriate positions in each two cut subdivisions.
PRINT	PROJ/$\{(c_i, c_j)\}$ (Default: Print all 2-D cuts)	Two-dimensional cuts of blocks to be printed.
	CNDPRB	Print conditional probability density estimates.

Keyword	Parameter	Function
	CLSMAP	Print classification map. NOTE: All results from DOTS, COMBIN, and/or TEST options are printed.
END		Terminates control card input.

c. Field Cards

Field cards will be handled exactly as for EOD-LARSYS, except that DESIGN header cards will include TRAIN, SAMPLE, and TEST designators, as well as OTHER and UNIDENTIFIABLE.

3.2.4 OUTPUTS

Files MAPTAP and/or DSTTAP are output where appropriate.

3.2.5 STORAGE REQUIREMENTS

TBD

3.2.6 DESCRIPTION

The set of subprograms outlined below will provide the processor building blocks.

Subprogram	Function
EQUPRB	Main driver.
SET15	Reads control cards, set data and options.
RDDAT	Reads training/test/sample fields, establish field vertices and enclosing rectangles.
	Reads MSS data tape, extracts radiance valves for selected channels, perform sun angle correction.

Subprogram	Function
BLOCK	Builds blocks for specified category.
COMBIN	Combines blocks from specified categories.
CLSTST	Classifies sample pixel fields.
SLTDOT	Performs dot selection.
PIXPRO	Computes proportion of pixels from each training category and for test fields in each block of specified block set.
PRPROJ	Print 2-D cuts of specified block set.
PRDOTS	Print spatial co-ordinates of selected dots.
PRMAP	Prints classification map.
PRBLK	Prints specified block set information.
PRCND	Prints conditional probability density estimates.
WRTCND	Writes file of conditional probability density estimates.
WRTMAP	Writes classification map.
GTASG	Read ground truth tape, assign category number and hence name to training field pixels, thereby breaking up the training fields by category. For any set of six subpixels of ground truth corresponding to a training field pixel, if the six subpixels do not all have the same category number, a strategy will be selected to choose a number to label the training field pixel.

Common block /GLOBAL/ may be expanded to include the unit and file number of DSTTAP. Common block /ISOLNK/ will pass the sun angle information. A new common block /EQUPRB/ will pass block, data, and option information among subprograms.

The overall architecture will follow that of EOD-LARSYS as closely as possible.