

SANDIA REPORT

SAND95-2051 • UC-706

Unlimited Release

Printed September 1995

An Optimal MAP Classification Algorithm for Color Printed Pattern Evaluation

P. H. Eichel

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-94AL85000



SF2900Q(8-81)

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof or any of their contractors.

SAND95-2051
Unlimited Release
Printed September 1995

Distribution
Category UC-706

An Optimal MAP Classification Algorithm for Color Printed Pattern Evaluation

P. H. Eichel
Optical Systems and Image Processing Department
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

An algorithm designed to classify print process defects for the on-line color printed pattern task of the Computer Aided Fabric Evaluation (CAFE) project is developed in this paper. The algorithm takes an optimal uniform Bayes cost approach that makes explicit use of ancillary CAD color separation and color palette data to make a probabilistic determination of defect cause. Easily implemented in a real-time processing environment, the technique is appropriate for print patterns utilizing both solid and tonal color designs.

Acknowledgement

The author would like to acknowledge his colleagues, K. T. Stalker, G. Sleaf, and A. Yee, for many helpful discussions on this and related efforts involving machine inspection of color printed patterns.

1.0 Introduction

This paper describes a defect classification algorithm proposed for use in the Color Printed Pattern task of the Computer Aided Fabric Evaluation (CAFE) project. The algorithm is intended to work in conjunction with a defect detection algorithm described in a companion paper [1]. Both are part of an integrated system whose goal is to monitor the printed fabric issuing from a rotary screen print range, detect and log print defects in real time, and suggest to the operator possible causes and remedial actions. The classification step in particular must accept information about a defect that has been flagged by the detection algorithm and determine: 1) the cause of the defect, and 2) the type or classification of the defect. The algorithm described herein is an optimal (Uniform Bayes Cost) approach that makes use of such ancillary information as CAD color separation data and a pattern's color palette to make a probabilistic determination of cause. It is appropriate for patterns utilizing both solid and tonal print screens.

2.0 Classification Requirements

The classification subsystem operates on data forwarded from the defect detection algorithm. This algorithm is derived and described in [1]. The purpose of that algorithm is to detect and flag individual pixels in the real-time camera data that do not agree in color space with the anticipated color at that location in the exemplar, or model, image. The degree to which the colors do not agree in order for a detection to occur is a complex function of operational parameters, noise levels, and temporal persistence. When a pixel has been so detected, the defect detection subsystem records the following information about the occurrence:

- The pixel location registered to the exemplar image.
- The pixel color value.
- The exemplar color value.
- The exemplar color separation number.

The location relative to the exemplar image in the first item above is known because the camera and exemplar images are very precisely registered as part of the detection process. The last item refers to the color separation associated with that location in the pattern, i.e. the print screen number. This information is derived from the CAD data for the pattern.

Having received the above information regarding a detected print defect, the function of the classifier is to: 1) Identify the color separation that is being printed incorrectly,

i.e. determine which print screen is the *cause* of the defect; and 2) Determine the type or *class* of the defect. As we shall see shortly, these two functions are not the same.

The rotary screen print process is fundamentally one of forcing colored dyes through porous print screens onto a cloth substrate. Nearly all of the print defects that can occur in this process can be broadly classified into three categories. First, we may have the situation of “the right color in the wrong places.” That is, we assume that the color extruded from a given screen is correct in coverage and shade but that it is being printed in the wrong places. There are two distinct subcategories. An entire screen may be misaligned with respect to the others, causing its color to be out of registration with the rest of the pattern. This is a fairly common occurrence in screen printing, especially at startup, and is known as misfit. It is characterized by the strong correlation of incorrectly dyed regions with the shape of the corresponding color separation. Alternatively, the incorrectly applied areas may occur randomly due to screen damage. Second, we may have “the wrong color in the right places.” Here we assume that the screen geometries are correct and registered, but that the color being applied is incorrect in some way. The color may have insufficient coverage (scabby), be of an incorrect dye formulation (off shade), be applied too thickly (mottled), or simply not be printed at all (missing color). Again there are two subcategories. The color problem can be global or at least apply to large areas of the print. Scabby, off shade, mottled, and missing color often fall in this subcategory. Conversely, the defect can occur at random, such as stickins. A third category results largely from substrate defects. Included in this category are seam marks, scrimps, print resist, butterflies, glue streaks, crooked cloth, and bias printing. In these cases, both the geometry and color extruded from the screens are assumed correct, the problem arises from the cloth passing beneath the screens. Seam marks, crooked, and bias printing fall into the global subcategory, the others into the local. These broad categories are summarized in the table below:

<i>Class</i>		<i>Defect</i>
I (Geometric)	Global	misfit, print smear
	Local	screen damage
II (Color)	Global	missing, scabby, off shade, mottled
	Local	stickin
III (Substrate)	Global	seam mark, crooked, bias
	Local	butterfly, scrimp, resist, glue

Table 1. A categorization of color print defects.

From this discussion, it should now be clear that knowledge of the position of a defect is not sufficient to determine its cause. While it is true that the CAD file data provides a means to determine the screen number from a location, the defect might not be caused

by that screen. Category I defects from the table above are characterized by screens printing *in the wrong places*. So the location of a defect does not point to the defective screen; rather, its color does. Similarly, the color of a defective region cannot always uniquely identify the cause. The color may be used to point to the cause of a Category I defect, but not a Category II defect. For them, positional information is paramount. Thus it may be seen that for these two Categories, determining the defect *classification* and identifying the *causal screen* requires mutually exclusive data. For Category I defects, classification is based on positional information, but causation is determined from color. For Category II, classification is a function of color, but causation is determined from position. This is the cornerstone of the algorithm advanced in the next Section.

Category III defects cannot be so analyzed. Since they fundamentally result from problems with the substrate and not the screens or dyes, classification and causation are not easily accomplished in general. Certain exceptions exist. Seam marks, for example, exhibit a fairly unique geometry, extending as they do fully across the web in a narrow region in the machine direction. However, the point at which the camera is positioned on the machine in the proposed color print system may preclude detection of crooked, bias, scrimp, and markoff defects. For these reasons, Category III defects will not be treated in this paper.

3.0 The Classification Algorithm

In this Section, we will develop an optimum defect classification algorithm based on maximum a posteriori probability (MAP) detection. Before doing so, however, we must take care of a few preliminaries. The first has to do with the exemplar image model we are hypothesizing for this problem and the second some mathematical concepts.

3.1 A palettized exemplar model

As discussed in [1], the exemplar image model is a *palettized* image. That is, the image consists of a two-dimensional array of indices into a color palette table. We also assume that the CAD separation data is available to the system for building the exemplar. By properly registering the CAD separations with an actual camera image of the pattern, Equations 2-5 in [1] demonstrate how we can build a palettized exemplar image with the geometric precision of the CAD data and the camera specific colors of the camera image. Although not stressed in that paper, this model has an additional benefit for the classification problem. That is, we can easily build a table linking palette color to separation number (or equivalently screen number). For color print patterns consisting only of solid colors, this is trivial since there is a one-to-one correspondance between palette colors and separations. For patterns containing tonal colors, however, this is not the case.

The problem with tonal colors is that various shades of a given color can be printed

from a single screen. Such a pattern can still be palettized. Of course, the number of colors in the palette may be quite a bit larger than the number of screens or separations. In fact, using good quality commercial color segmentation algorithms, it is not unusual to require up to 256 colors in the palette to accommodate a tonal image. Also, and most importantly, there is generally no one-to-one relationship between palette colors and separation numbers. A given palette color can arise in several separations and vice versa. Therefore, obtaining separation number from color involves *conditional probabilities*.

3.2 Discrete memoryless channels and transition probabilities

We will consider a simple mathematical model to describe the input-output relationship between separation (or screen) number and palette color. A *discrete memoryless channel* or DMC is an object that accepts one of r input possibilities and expels one of s output possibilities at a time. It is called *discrete* because there are only a finite number of inputs and outputs and *memoryless* because any given output depends only on its corresponding input and no other [2]. Let the alphabet of inputs be the separation numbers for the print pattern, $[1, 2, \dots, r]$, and the outputs the palette colors, $[1, 2, \dots, s]$. Consider a certain pixel location in the image. According to the CAD model, one of the r screens will print at this location. This is the input to the DMC and will be denoted as: $x \in [1, r]$. After this location has been printed and imaged by the inspection camera, the color at that pixel is compared to the color palette. The closest (in the sense of color space), palette color is found. This is the DMC output and is denoted as: $y \in [1, s]$. What is important in this model is that the output is not a definite function of the input. Just because this certain location was printed by screen $x = 4$, say, does not imply that the closest palette color to the imaged pixel is $y = 37$. Sometimes screen 4 yields color 32 or 46 and sometimes color 37 results from screen 11, to continue the example. This is because the colors printed by individual screens span a region in color space and the regions spanned by different screens can overlap. In addition there are process variables and sensor noise sources that exasperate the problem.

In the DMC model, the output is governed by a set of *transition probabilities*, $P(y|x)$. Knowing the screen number $x = t$ does not tell us precisely the resulting color, $y = c$. However we can estimate the *probability* that color $y = c$ is the result. This may be accomplished, as with exemplar formation, by registering the CAD separations with a camera image of the pattern. First, the camera image is palettized using a color segmentation algorithm. Then for every separation, $x = t \in [1, r]$, and every palette color, $y = c \in [1, s]$, the ratio of the number of pixels of palette color $y = c$ in separation $x = t$ to the total number of pixels in separation $x = t$ estimates the probability $P(y = c|x = t)$.

To see this, let us adopt the mathematical notation developed in [1]. A camera image is a three valued function (red, green, blue) for every position (i, j) :

$$\mathbf{f}(i, j) = [f_r, f_g, f_b] \quad , (i, j) \in \mathcal{I}^2 \quad (1)$$

and the CAD separation files are simply binary valued images:

$$\text{sep}_t(i, j) = \begin{cases} 1 & \text{if separation } t \text{ is on at } (i, j) \\ 0 & \text{if separation } t \text{ is off at } (i, j) \end{cases}, t \in [1, r] \quad (2)$$

We define an indicator function $I_c(i, j)$ as:

$$I_c(i, j) = \begin{cases} 1 & \text{if } \mathbf{f}_p(i, j) = c \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where the subscript p of the image function \mathbf{f} indicates the closest palette color to the pixel value, $\mathbf{f}(i, j)$. An estimate of the transition probability $P(y = c|x = t)$ is then:

$$P(y = c|x = t) \simeq \frac{\sum_{i,j} I_c(i, j) * \text{sep}_t(i, j)}{\sum_{i,j} \text{sep}_t(i, j)}, t \in [1, r], c \in [1, s] \quad (4)$$

With the transition probabilities determined as above, the DMC provides us with a model of the relationship between pattern separations (or screens) and palette colors. In the next section, we will use this model to construct an optimum defect classifier for the color print problem.

3.3 The MAP Classifier

In Section 2.0 we learned that there are two distinct defect categories: Type I defects that are characterized by screens printing proper colors but in the wrong places of the pattern, and Type II defects where the print geometries are correct but the colors are wrong. The classifier has the task of receiving information from the defect detector algorithm identifying defect pixels, their location, and their color, and determining which screen is the cause of the defect. In addition, it should make a determination of the class of defect for process logging.

Of the two, Type II defects are the easier to identify the problematical screen. The assumption here is that all the screens are printing in their correct locations. Since the camera imagery has been carefully registered to the exemplar model in the detection process, and since the CAD data has furnished the exemplar with knowledge about where each separation prints in the pattern, the location of the defect points directly to the separation number in question. The only difficulty arises in determining the actual classification of the defect. Defect size may help distinguish between stickins and the more global defects. Missing and scabby color might be distinguished by the level of saturation of the color, but off shades could represent a discrepancy in any direction in color space. Ultimately, a certain amount of judgement is involved in this determination.

Type I defects present a more complicated problem. Here we will use the mathematical machinery presented in 3.1 and 3.2. For Type I defects, the defect position provides no clue as to which screen is printing in error. However, the assumption that the screens are all printing correct colors does. We need to make our best determination of the separation in error based on the color of the defective pixels. We develop an optimum classifier as follows.

Assume that the closest palette color of a defective pixel is y . The specific problem here is to determine which input x *most likely* resulted in this color. Let us further assume that the classifier makes the choice: $x = t$, $t \in [1, r]$. The probability that a correct choice was made can be written as the sum:

$$P(C) = \sum_y P(C|y)P(y) \quad , y \in [1, s] \quad (5)$$

It is clear from this expression that we can maximize the probability of making a correct choice by maximizing $P(C|y)$ for every y . The conditional probability of a correct decision can be written:

$$P(C|y) = P(x = t|y) \quad (6)$$

But from Bayes rule, the a posteriori probabilities $P(x|y)$ are given by:

$$P(x|y) = \frac{P(x)P(y|x)}{P(y)} \quad (7)$$

The denominator on the right hand side of Eq. 7 is independent of the choice of x . Therefore the optimum classifier in the sense of maximizing the probability of a correct decision when an output color $y = c$ is observed, is to choose the input $x = t$ that maximizes the function:

$$\max_{t \in [1, r]} P(x = t)P(y = c|x = t) \quad (8)$$

That is, we choose $x = t$ such that Eq. 8 is maximized. To do this, of course, requires knowledge of the transition probabilities $P(y = c|x = t)$ from Section 3.2.

The decision rule just derived is known as a maximum a posteriori probability detector, abbreviated MAP [3]. Note that it weights the decision regions by the a priori probabilities $P(x)$. A detector that determines x by maximizing only the conditional $P(y|x)$ without regard to the a priori probabilities is called a *maximum likelihood* (ML)

detector. An ML detector yields the minimum probability of error when the a priori are all equally likely.

This MAP rule applies to decisions made on individual pixels. Its performance on individual pixels may not be sufficiently good to provide a high confidence level in the case of overlapping conditional density functions. However, to have just a single detected defect pixel would be a highly degenerate case. Much more typical is to have hundreds or thousands of detections occur, especially in the case of global defects such as misfit or missing color. The generalization of the above to the joint classification problem is straightforward. Recall that the DMC is a memoryless model (where memory refers to *spatial* not *temporal* correlation here). Thus, the mapping from input color separation to output palette color is independent from pixel to pixel (although the inputs themselves are obviously highly correlated, spatially). This is known as *conditional independence*:

$$P(y_1, y_2|x) = P(y_1|x)P(y_2|x) \quad (9)$$

Accordingly, the *joint* correct decision, given N defect detections $[y_1, \dots, y_N]$, is to choose the input $x = t$ that maximizes:

$$\max_{t \in [1, r]} P(x = t) \prod_{i=1}^N P(y_i|x = t) \quad (10)$$

or, in the case that the a priori probabilities are either unknown or assumed equally likely,

$$\max_{t \in [1, r]} \prod_{i=1}^N P(y_i|x = t) \quad (11)$$

These two equations represent our classifier for the Type I category of defects. Recall that the Type II category defects were classified by a straightforward mapping from position to color separation, based on the CAD data. A final question remains: How do we determine which classification category we have present? This is not easy to answer. Both the position and the color of the defective pixels can tell us something. If the defects all occur on or near a boundary between two color separations, then a misfit is more likely. Also, if the defect color is far from any palette color in color space, then a Type II defect is more likely. But neither test can conclusively determine the category. A better approach is to perform the classification test twice, once under the hypothesis that the defect is Type I, and once under the hypothesis that it is Type II. Both answers may

be advanced, perhaps weighting their probabilities by the above contextual information.

4.0 Summary

In this paper we have derived an optimal defect classifier based on a minimum probability of error criteria. Intended for use in the CAFE color printed pattern evaluation system, the classifier is appropriate for use on patterns with both solid and tonal color screens. The classifier can make a maximum a posteriori (MAP) probability estimate of the screen causing a given defect flagged by a companion defect detection algorithm. The algorithm makes specific use of CAD color separation data to build a transition probability matrix that is the heart of the MAP classifier. Execution in a real time environment is facilitated by the fact that implementation may be accomplished through the use of lookup tables and r product accumulators, where r is equal to the number of color separations for the pattern.

5.0 References

- 1.) P. Eichel, "A sequential detection algorithm for color printed pattern defects," Rep. SAND95-2052 (Sandia National Laboratories, Albuquerque, NM, 1995).
- 2.) R. Gallager, **Information Theory and Reliable Communication**, Wiley, 1968.
- 3.) H. Van Trees, **Detection, Estimation, and Modulation Theory**, Wiley, 1968.

Distribution:

- 1 AMTEX Program Office
Pacific Northwest Laboratory
Attn: Doug Lemon
P.O. Box 999, MS K7-80
Richland, WA 99352

- 1 Burlington Industries Inc./Klopman Fabrics
Department of Quality Assurance
Attn: Andy Russell
P.O. Box 2000
Hurt, VA 24863

- 1 Cone Mills Corporation
CAD Department
Attn: Pamela Richard
P.O. Box 8
Carlisle, SC 29015

- 1 E.I. du Pont de Nemours and Company
Central Research and Development Department
Attn: Arun Prakash
Experimental Station
P.O. Box 80357
Wilmington, DE 19880-0357

- 1 Fruit of the Loom
Department of Product Development
Attn: David Allen
Route 1, Box 1
Rabun Gap, GA 30568

- 1 Institute of Textile Technology
CAFE Department
Attn: Dan McCreight
2551 Ivy Road
Charlottesville, VA 22903

- 1 Lawrence Berkeley Laboratory
Department of Engineering
Attn: Craig Fong
1 Cyclotron Road, Building 90/2148
Berkeley, CA 94720

- 1 Lawrence Livermore National Laboratory
Engineering Research Division
Attn: Jose Hernandez
P.O. Box 808, L-156
Livermore, CA 94551
- 1 Milliken Research Corporation
Live Oak Department
Attn: Morton Reed
300 Lukkin Ind. Drive
LaGrange, GA 30240
- 1 Milliken Research Corporation
Attn: Louis Adams
P.O. Box 1927
920 Milliken Road
Spartanburg, SC 29304-1927
- 1 Oak Ridge National Laboratory
Attn: Glenn Allgood
Bethel Valley Road
Oak Ridge, TN 37831
- 1 Oak Ridge National Laboratory
Instrumentation and Controls Division
Attn: Bryan Allen
P.O. Box 2009
Oak Ridge, TN 37831-8066
- 1 Oak Ridge National Laboratory
Instrumentation and Controls Division
Attn: David Sitter
Building 3546, MS 6011
Oak Ridge, TN 37831
- 1 Oak Ridge National Laboratory
Instrumentation and Controls Division
Attn: Ken Tobin
Bethel Valley Road
Oak Ridge, TN 37831-6011

1 Spartan Mills
Cherokee Finishing Company/Color Shop
Attn: Benji Mills
210 Chandler Drive
Gaffney, SC 29341

1 Spartan Mills
R&D Department, Information Services
Attn: Patrick Chong
805 Spartan Boulevard
Spartanburg, SC 29301

1 Thomaston Mills, Inc.
Finishing Department
Attn: Joe Perdue
P.O. Box 311
Thomaston, GA 30286

1 MS0842 Carolyne Hart, 9100
1 MS0842 Larry Hostetler, 9133
1 MS0843 Terry Stalker, 9136
20 MS0843 Paul Eichel, 9136
5 MS0843 Gerard Sleaf, 9136
1 MS0844 Amy Yee, 9133
1 MS9018 Central Technical Files, 8523-2
5 MS0899 Technical Library, 13414
1 MS0619 Print Media, 12615
2 MS0100 Document Processing, 7813-2, for DOE/OSTI