Proceedings of the 35th
Conference on Decision and Control
Kobe, Japan • December 1996

FP19 4:30

Toward a Control Oriented Model of Xerographic Marking Engines

L. K. Mestha, Y. R. Wang, S. Dianat, E. Jackson, T. Thieret
Control & Diagnostics Laboratory,
Wilson Center for Research & Technology,
Xerox Corporation

P. P. Khargonekar, D. E. Koditschek
EECS Department, College of Engineering,
University of Michigan

Abstract

This paper presents some preliminary results from a recently initiated research collaboration concerning the modeling and control of color xerography. In this first communication of our work, we describe our efforts to develop a model for a monochrome marking engine. We adopt the technique of principal component analysis for choice of output coordinates and demonstrate preliminary experimental evidence suggesting that this procedure yields accuracy in data reconstruction superior to present industry practice. Preliminary analysis of the experimental evidence suggests that the process has a nonlinear component that we will seek to model using a mixture of physical and empirical insight.

1 Introduction

In the last year we have established a joint research collaboration between the Xerox Corporation and the University of Michigan devoted to modeling and control of xerographic processes. The objective of the Xerox-UM collaboration is to address basic control systems problems to obtain significant improvements in stability and performance. Color is typically specified in a three dimensional space – (L^*, a^*, b^*) – wherein each point corresponds to a particular "color." In a digital document representation, the color at every location on the page is specified by a triplet of quantized coordinates in this space. The metric that is normally used to measure the color consistency is ΔE, the Euclidean distance norm, to determine the difference between two colors in the color space being considered [1]. The human visual system can distinguish colors separated by roughly one unit of distance.

One metric of color print quality is the ability of the printing system to reproduce and maintain every color to a specified tolerance. Typically, printers are capable of keeping individual color variations within a sphere of radius of 5-10 in the (L^*, a^*, b^*) space (in some absolute units) for a single print run. For high quality color reproduction, variations within a radius of approximately 3 are required. Currently, no printing technology has demonstrated this level of color stability.

The aim of this paper is to present some of our initial results on modeling of a monochrome xerographic process. Typically, four monochrome colors — cyan, yellow, magenta and black are used in a dry powder color printing process. We need to get every color right every time, to meet the challenges of the color printing market. We first present a brief qualitative description of the xerographic process. This is followed by a discussion of representation, specifically, coordinate transformations in the input and output spaces. Next, we present some preliminary experimental results on input-output modeling of a monochrome xerographic process.

2 Overview of the Xerographic Process

For the purposes of this paper, we find it convenient to break the process into two principal components connected in series to form a xerographic printer, as depicted below.
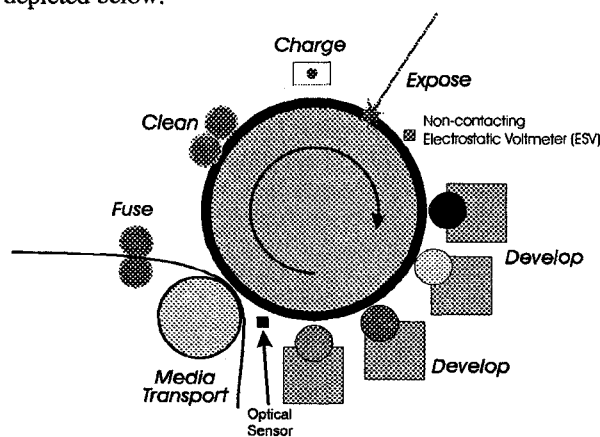
2.1 Principal Subsystems

2.1.1 Plant A — Charge/Expose/Develop: The xerographic process centers around a photoreceptor (P/R), a multi-layer belt or drum which retains charge in the dark but discharges when exposed to light. Plant A begins with an uncharged P/R and finishes with the customer page (in mirror image) expressed as a pattern of small colored particles (toner) on its surface. The first step in the xerographic process is to "charge" the surface of the P/R uniformly by varying the voltage, u_g, on a parallel ion emitting wire grid. Second, the "expose" step of the process involves discharging the charged P/R surface using a laser modulated according to a bit stream representative of the desired image which leads to a dimensional electric charge distribution on the P/R. The magnitude of the voltage reduction is a function of the illumination intensity, u_l, of the laser. In the "development" process, the binary electrostatic image residing on the P/R is developed using a bias voltage u_b into a visible image of toner particles whose mass density distribution mirrors the exposed charge density distribution, and, hence, the original image. The physics underlying many of the development methods used in xerographic engines is discussed at length in the book by Schein [2]. The three control signals are denoted by u_A.

Denoting by p_A the uncontrolled sources of variation in the charge/expose/develop process (for example, surface feature variations on the P/R), we may now represent Plant A by the equation

$$m_d = \Phi_A(u_A, p_A, m_c) \qquad (1)$$

where m_d is the toner mass density and m_c is a representation of the customer image to be reproduced.

2.1.2 Plant B — Transfer/Fuse: In the "transfer" process, paper transported on a conductive roller is brought in contact with the developed toner image on the P/R. A voltage, u_t, opposite to that of the toner charge is applied to the roller so that the reversed electric field breaks the electrostatic force between the toner particles and the P/R surface, thus transferring the particles to the paper. The "fusing" process fixes the toner to the paper. The most common process is hot roll fusing system [3] which consists of a pair of rollers, one of which is heated internally with a quartz lamp to a setpoint temperature.

The composite xerographic system may be regarded as a series interconnection of Plant A followed by Plant B.

2.2 Representation of Photoreceptor State

The combination of Plants A and B results in an effective spatial distribution of points in (L^*, a^*, b^*) across the printed page. Indeed, ultimately, it is this visual appearance of the marked paper that counts. However, as may be imagined, the full measure of the impact of the various signals on the xerographic process and, in particular, on the optical output, (L^*, a^*, b^*), is imperfectly understood. Industrial experience indicates that a far simpler abstraction of the vast quantity of information contained in the state of the P/R can be used instead to deal with the interrelationships between these signals. This reduced representation of the P/R state is called a reproduction curve which we will now describe.

2.2.1 Halftone Dots and Reproduction Curves: A halftone dot is the printer's unit of area on a page — for high quality printing roughly 150 dots to the inch are needed. It is traditional in the photographic industry to measure gray level in terms of a continuous scale of image density, which we will call contone value, at each dot of the image perceived at normal reading distance. The contone scale is represented by the interval [0,1] where 0 denotes the reflection of the bare substrate and 1 denotes the reflection from a fully developed solid area patch. Intermediate image densities are accomplished through digital halftoning.

At each step along the xerographic process one may define a function, called a reproduction curve (RC), that relates contone value at a dot to the appropriate measure of the image at that step. For example, the "developed reproduction curve" (DRC) is a function, C_d, that maps contone values into the percentage of total possible toner mass at a dot in the toned image on the P/R that comes out of the developer stage. The most important of these functions is, C_c, the customer's "desired tone reproduction curve" (TRC, to follow industry parlance) that maps contone values into the percentage darkness that the user wishes to ascribe each dot on the output fused image in advance of seeing the results of the physical printing process. This plays the role of a reference command in the xerographic process. It is important to note that each reproduction curve maps the interval [0,1] onto itself in a monotonic fashion.

2.2.2 Plant A As a DRC Generator: Having introduced the notion of a reproduction curve, we may now rewrite (1), as

$$C = F_A(u, p) \qquad (2)$$

where $u = \left[u_g, u_l, u_b \right]^T \in R^3$ now replaces u_A in (1), p replaces p_A, C denotes C_d, and we have simply taken the representation of the customer's image, C_c to be the identity reproduction curve.

While C has the physical character of a monotone function from the interval [0,1] into itself, the sensors available to read this function will deliver a discretized version. Hence we consider the developed reproduction curves (DRC) to be $C \in \Gamma := R^n$ where n denotes the discretization of a continuous DRC. This discretization arises from the data representation of each halftone dot density as an 8 bit integer. Typically in commercial printing the process sensors observe only a small fraction of the 256 states available. For the experiments reported in this paper $n = 111$.

3 The Modeling Problem

Roughly speaking, there are three aspects to developing a useful experimental model of a plant (2). One needs to choose coordinates for the output space, coordinates for the input space, and, of course, adopt a strategy for fitting input output models to the experimental input output pairs when expressed in those coordinates. In the present study, we have concentrated on the first of these and merely mention in passing our ongoing efforts regarding the second and third.

3.1 New Coordinates for the DRC Space

Because Γ, being a space of functions, has such high dimension, it is imperative to replace it with a much lower dimensional projection of the DRC space. Otherwise, useless computational effort might be expended in accounting for degrees of freedom far greater in number than can be compensated for by the authority of the available xerographic actuators. Thus, we are led to introduce a projection,

$$D := \psi(\Gamma), \tag{3}$$

where ψ is a suitably chosen transformation onto a low dimensional vector space.

It has been traditional practice to sense DRCs at a fixed small set of sample points, so that the measured output takes the form

$$c := \Pi^T C \tag{4}$$

where each column of Π is a vector of zeros with a single unity value stuck in the slot corresponding to the

quantized value at which the DRC is to be sampled. Thus, this particular "Boolean" projection, common to many machines has long been embraced in general industrial practice.

We will use instead in out input/output experiments an alternative projection that enjoys certain optimality properties. In particular, if we take the notion of optimality to be least squares with respect to a calibration sample, then this results in a choice for ψ dictated by the principal components of the covariance matrix as we will outline below. The reduced output space is thus $D = R^r$, where r is the number of principal components and $d \in D$ takes the form

$$d = \psi(C) := \Psi^T C. \tag{5}$$

In section 4, we will present experimental results comparing the traditional practice with these "optimal" coordinates.

3.1.1 Principal Component Analysis: Denote the matrix resulting from a set of output measurements — N — as

$$\Omega := \begin{bmatrix} C_1 & C_2 & \cdots & C_N \end{bmatrix} \in R^{n \times N},$$

and define its associated (non-central) sample covariance matrix

$$\Sigma := \frac{1}{N} \Omega \Omega^T \tag{6}$$

Since Σ is a positive semidefinite symmetric matrix it admits the singular value decomposition

$$\Sigma = \begin{bmatrix} \Psi & \Phi \end{bmatrix} \begin{bmatrix} \Lambda_r & 0 \\ 0 & \Lambda_{n-r} \end{bmatrix} \begin{bmatrix} \Psi^T \\ \Phi^T \end{bmatrix} \tag{7}$$

where $\Lambda_r \in R^{r \times r}$ is a diagonal array containing the r - largest eigenvalues of Σ in descending order, $0 \approx \Lambda_{n-r} \in R^{(n-r) \times (n-r)}$ contains the smallest $n-r$ eigenvalues, and, hence, we may approximate

$$\Sigma \approx \Psi \Lambda_r \Psi^T \tag{8}$$

where $\Psi \in R^{n \times r}$ is the matrix of r "principal component" vectors that we may take as our basis vectors for D as used in (5).

3.1.2 Data Reconstruction: In order to compare our choice of output coordinates, (i. e., d) against the

standard industrial practice, denoted by c, we will use certain metrics which capture their power to retain information relative to the original data. Thus, if we desire to reconstruct the DRC implied by the D-space coordinates, $d \in R^r$, then the obvious computation is

$$\hat{C} = \Psi^+(d) = \Psi d. \tag{9}$$

However, a particular commercial machine may not afford data in this form. Instead, given any other linear projection π, as in (4), there is a corresponding least squares optimal reconstruction, π^+, with the property of best recovering (again with respect to the set of N sample vectors that form the columns of Ω). That is, if we consider among all linear reconstructions

$$\hat{C} := \pi^+(c) = \pi^+\left(\Pi^T C\right)$$

that one which minimizes the sum of squared reconstruction errors, then the optimum takes the form

$$\pi^+(c) = \Pi^+ c \tag{10}$$

where Π^+ minimizes the Frobenius norm of

$$\Omega - \Pi^+ \Pi^T \Omega.$$

Using standard facts from linear algebra,

$$\Pi^+ = \Sigma \Pi \left(\Pi^T \Sigma \Pi\right)^{-1}. \tag{11}$$

Notice that if Π were chosen to coincide with the first r eigenvectors of Σ then we would have $\Pi^+ = \Psi$, justifying the obvious choice in (9), presuming the availability of sensed data in the form (5). As a matter of fact, the least squares optimal choice for Π itself is precisely the principal components basis given by Ψ [5]. In principle, then, the alternative reconstruction (10), would only be indicated if available sensor data takes the form (4) rather than (5), as in indeed the case in some commercial machines.

The reconstructions described above are least squares optimal — the mathematical preference for (9) over and against (10) only relevant in the face of an appropriately richer sensory system, that might yield outputs of optimal form (5) rather than (4). In contrast, for present standard industrial practice, DRC reconstruction from the sampled values contained in c as sensed by (4) proceeds according to spline fit (that we denote by π (spline) in the sequel). Therefore, as a matter of significant interest independent of input/output models, we will compare experimentally the reconstruction accuracy of these three variant methods in Section 4.2..

3.2 New Coordinates for Actuator Space

In recent years there has emerged in industrial practice the suspicion that U should be more properly transformed

$$V = S(U) \tag{12}$$

by a nonlinear change of coordinates, S, into what are called the soft actuators, V.

This view of the xerographic process holds that the physical actuators u, affect the DRC through an intermediate set of *soft actuator* values, $v = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T \in R^3$, in a manner that depends upon a nonlinear change of coordinates involving ratios of the grid voltage and bias voltage as well as the so-called PIDC — photo-induced discharge curves — which capture the relationship between the surface potential of the photoreceptor and the light exposure [4].

Unfortunately, there is no first principles based reasoning on which to develop a model for the PIDC curve. Since the soft actuator values, v, corresponding to a particular hard actuator setting, u, are easily experimentally available in the laboratory we are preparing a separate study wherein we will attempt to fit a model for S.

3.3 Parametrization of F_A

We anticipate that the input coordinate study may have a significant impact on the choice of models for the plant, (2). In the meantime, however, we have adopted the obvious framework of Taylor series expansion, and will consider various models that entail increasingly high order expansions. The results of these experimental studies are presented in Section 4.3.

4 Experimental Results and Analyses

4.1 The Experiments

4.1.1 Experimental Vehicle: A Xerox 4890 highlight color commercial printer, was modified for experimental work. It was instrumented with a state-of-the art data acquisition and measurement system to obtain the DRCs each time a custom built image capable of sampling the area coverages between 0 to 1 is developed on the photoreceptor. DRCs were measured optically on the photoreceptor using a method closely resembling that covered in [6]. Various combinations of input settings of

the hard actuators were developed by externally disabling the internal machine control functions.

4.1.2 Experimental Design:
We have undertaken a full factorial design with five levels, so that prints were made at each of the 125 combinations of the five input settings for the three actuators. Of these 125 experiments, the results of four had to be discarded because of experimental failures, leaving a total of $N = 121$ input-output pairs $\{(u_i, C_i)\}_{i=1}^{N}$, where $u_i \in R^3$ and $C_i \in R^n$, $n = 111$, as related by (2).

4.2 Output Coordinates

From a practical standpoint, implementing our proposed new coordinates (3), along with the accompanying optimal reconstruction (9), might necessitate a new sensor (5) for certain machines. Even given the standard sensor (4), implementing its accompanying optimal reconstruction (10) would necessitate a calibration experiment (to obtain Ω, from which the optimal reconstruction, π^+, might be derived) that is not presently required in standard practice reconstruction, π (spline). Thus, to assess the potential improvement offered by our proposed methods relative to their increased implementation costs, we present a comparative cross validation analysis of the three reconstruction techniques applied to our experimental output data.

4.2.1 Cross Validation Studies:
We proceed as follows: We partition the original data set of $N_o = 121$ DRC curves into two disjoint populations — N_f, a subpopulation "fitting data" that will be used to obtain a "calibration", Σ_f— and N_p, a sub-population that will be used as test data. Using Σ_f we compute ψ_f^+ and π_f^+ according to (10) and (11). Then the resulting ψ_f^+ and π_f^+ in π (spline) are applied to N_p to compare the quality of representation.

We must take care to insure that N_f is large enough to ensure a good fit, but not so large that the prediction test population, N_p is uselessly small. Our criterion of adequate fit will be based upon the ratios of mean square errors as follows. Given a population, N, of DRC data, define the mean total sum of squares to be $MST := SST / N$ where the total sum of squares is defined by

$$SST := \sum_{i=1}^{N} \left\| C_i - \overline{C} \right\|^2 \tag{13}$$

and $\overline{C}$ defines the sample mean.

Further given a reconstruction procedure and a set of DRC curves, define the mean error sum of squares to be $MSE := SSE / N$ where the error sum of squares is defined

$$SSE := \sum_{i=1}^{N} \left\| \hat{C}_i - C_i \right\|^2, \tag{14}$$

and $\hat{C}_i$ is computed by one of the reconstruction method. Thus, MSE_o will denote the MSE resulting from using $N = N_o$, MSE_f will denote the MSE resulting from applying the reconstruction procedure to the fitting population, N_f, and MSE_p to be the MSE resulting from applying the reconstruction procedure to the test data, N_p. We will make sure that the fitting population percentage, N_f / N_o is sufficiently large that MSE_f / MSE_o is within 10 percent of unity.

Given an adequate fit relative to the original population, we may now assess the predictive capability of the reconstruction by considering the various mean squared errors — that of the predictive population being the most important. We expect (except under peculiar arrangements of data within the populations) that MSE_f should be smaller than MSE_p. However, a much smaller MSE_f casts suspicion on the predictive adequacy of the proposed reconstruction.

4.2.2 Comparitive Study of Reconstruction Accuracy:
The accompanying tables summarize the results of six different randomized cross validation studies comparing the three methods of reconstruction introduced in Section 3.1. For each of the six pairs of rows a different fitting population of size, $N_f = 85$ was chosen at random from the original.

Study	Coordinates	MSE_o	MSE_f	MSE_p
1	ψ	0.2961	0.3013	0.2890
1	π	0.5136	0.5637	0.4122
1	π (spline)	0.7235	0.7963	0.5516

| 2 | ψ | 0.2961 | 0.2810 | 0.3479 |
| 2 | π | 0.5136 | 0.5381 | 0.5064 |

2	π (spline)	0.7235	0.6877	0.8080

3	ψ	0.2961	0.2441	0.4326
3	π	0.5136	0.4241	0.7710
3	π (spline)	0.7235	0.6672	0.8564

4	ψ	0.2961	0.3163	0.2516
4	π	0.5136	0.5679	0.4079
4	π (spline)	0.7235	0.7926	0.5603

5	ψ	0.2961	0.2911	0.3233
5	π	0.5136	0.5112	0.5336
5	π (spline)	0.7235	0.7323	0.7026

6	ψ	0.2961	0.3035	0.2817
6	π	0.5136	0.5705	0.4023
6	π (spline)	0.7235	0.8060	0.5287

We see from the MSE_f values that the optimal coordinatization/reconstruction, Ψ, yields roughly double the accuracy of the best possible reconstruction π^+ for the traditional method of sampling the DRC at 3 points. Also, the optimal reconstruction π^+ is approximately 30% better than the spline fit. Moreover, we see from the last column that the optimal coordinatization/reconstruction has predictive value roughly one third better than that of the optimal reconstruction π^+ which in turn is roughly one fourth better than a spline fit.

4.3 Input Output Modeling

As mentioned above, we have carried out a full factorial experimental design and collected input-output data on Plant A (2). Having derived the optimal output coordinates, $d = \psi(C)$, given by Ψ_o, from above, we have fit successively higher degree polynomial models to the projected data, $\{u_i, d_i\}_{i=1}^{N}, u_i \in R^3, d_i \in R^3$ with results that we now discuss.

4.3.1 Fitting Polynomial Models to the Data: We considered five different polynomial models for (2) with degree as summarized in the accompanying table

Model	Terms	Parameters
1	linear	9
2	affine, linear	12
3	pure quadratic	18
4	affine, linear, quadratic	30
5	affine, linear, quadratic, cubic	54

Since each of these models is linear in parameters, standard techniques of linear regression apply, and we obtain the least squares optimal fit in each case by application of the appropriately formed pseudo-inverse matrix function of the input data to the matrix of output data, Ω. To test the relative merit of the five models, we have again performed six sets of comparative cross validation studies. In presenting the results of these studies, we will again take recourse to the data spread statistic, MST, and model residual statistic, MSE from above. We also consider the traditional "adjusted" R^2 statistic

$$R^2 := 1 - \frac{SSE}{SST} \frac{(N-1)}{(N-p)} \tag{15}$$

where N is the population size and p is the number of parameters in the model. Note that the adjusted R^2 statistic approaches unity as the spread of the residuals decreases relative to the spread of the experimental outputs. However, note as well that models with a greater number of parameters are "taxed" more by this measure.

4.3.2 Comparison of the Five IO Models: In each of these five studies we take a different randomly selected sub-population of $N_f = 85$ and compute the best model of each polynomial degree for that population. We then compute the prediction mean sum of squares, MSE_p by applying the fitted model to the new data, $N_p = N_o - N_f = 36$ and adding the norm squared of the residual errors. The results are summarized in the accompanying tables.

For study number one, we find

Model	MSE_o	MSE_f	MSE_p	R_f^2
1	2.6879	2.6787	2.8675	0.3501
2	1.2119	1.2456	1.2343	0.7266
3	1.7344	1.6865	2.0193	0.5670
4	0.6704	0.6716	0.8070	0.8515
5	0.5742	0.5819	0.7674	0.7742

for study number two,

Model	MSE_o	MSE_f	MSE_p	R_f^2
1	2.6879	2.3753	3.4217	0.4365
2	1.2119	1.2185	1.3223	0.7389
3	1.7344	1.5726	1.9344	0.6070
4	0.6704	0.6420	0.8048	0.8688
5	0.5742	0.5446	0.7610	0.8136

for study number three,

Model	MSE_o	MSE_f	MSE_p	R_f^2
1	2.6879	2.8076	3.0574	0.3236
2	1.2119	1.1296	1.6141	0.7680
3	1.7344	1.8293	2.1345	0.5327
4	0.6704	0.6482	0.8223	0.8719
5	0.5742	0.5449	0.8067	0.8241

for study number four,

Model	MSE_o	MSE_f	MSE_p	R_f^2
1	2.6879	2.4804	3.2176	0.4064
2	1.2119	1.2113	1.3305	0.7391
3	1.7344	1.6957	1.7514	0.5673
4	0.6704	0.6831	0.7605	0.8511
5	0.5742	0.5449	0.8724	0.8088

for study number five,

Model	MSE_o	MSE_f	MSE_p	R_f^2
1	2.6879	2.8413	2.5849	0.3150
2	1.2119	1.2021	1.3393	0.7482
3	1.7344	1.8897	1.7144	0.5148
4	0.6704	0.7406	0.6554	0.8381
5	0.5742	0.6394	0.6286	0.7562

As can be seen from these tables, model 4 (a Taylor series out to second degree) produces the best fit in all studies, both with respect to initial fitting capability, R^2, as well as in its predictive power, MSE_p. Model 2 is often better than models 1 and 3 as might be expected given the absence of the affine term in the latter. Model 5 does reduce the fitting errors in consequence of its much larger number of free parameters. However, the adjusted R^2 statistic shows that adding the cubic terms does not necessarily lead to a "better enough" model.

5 Conclusion

In this paper, we have described the first steps in a much longer program of collaborative research between the Xerox Wilson Research Center and the University of Michigan Controls Laboratory. Based on our experience so far, we believe that the modeling and control of color marking processes is an important and fruitful area of research in control applications.

We have introduced a new coordinate system for dry ink tone reproduction within the framework of principal component analysis and have demonstrated its superior power of data reconstruction relative to present industry practice. This demonstration represents the principal contribution of the present paper.

We have explored as well the validity of simple polynomial models of the xerographic dry marking process. While much more work is necessary in order to draw firm conclusions, the relative fitting and predictive power of the higher order polynomial models — even when they are "taxed" for their reliance on more parameters — suggests the presence of a strong process nonlinearity. We suspect, however, that a more structured nonlinear-in-parameters model arising from physical considerations — a nonlinear change of coordinates in the input space that has emerged from recent industrial research — might yield a more parsimonious and even more accurate I/O model.

In this presentation, we have all but ignored the very real and important effects of uncontrolled perturbations, p in (2), on the experimental data and the limitations in modeling and control that they imply. Developing a theoretical framework and experimental methodology to account for the presence of noise in xerography is the focus of much of our present work in progress.

6 Acknowledgments

7 References

[1] F. W. Billmeyer and M. Saltzman, *Principles of Color Technology,* John Wiley & Sons, 1981.
[2] L. B. Schein, *Electrophotography and Development Physics,* Springer-Verlag, New York, 1988.
[3] D. M. Pai and B. E. Springett, "Physics of Electrophotography," *Rev. Mod. Phys.,* vol. 65., no. 1, Jan 1993.
[4] T.E. Thieret, T.A. Henderson and M.A. Butler, "Method and Control System Architecture for Controlling Tone Reproduction in a Printing Device", United States Patent# 5,471,313, 1995.
[5] R. A. Horn and C. R. Johnson, *Matrix Analysis,* Cambridge University Press, 1985.
[6] L.K. Mestha, "Method for Measurement of Tone Reproduction Curve Using a Single Structured Patch", United States Patent# 5,543,896, 1996.