

Spectral-Based Characterization of Inkjet Printers by Neural and Genetic Algorithms

Silvia Zuffi^{°}, Raimondo Schettini^{°§}, Giancarlo Mauri[§]*

[°]Tecnologie Informatiche Multimediali, ITC-CNR, Milano, Italy

[§]DISCO, Università degli Studi di Milano Bicocca, Milano, Italy

Abstract

The spectral-based characterization of inkjet printers is often based on a physical description of the printing process, but our experience tells us that it is not unusual to be unable to fit a mathematical model to a given printer. The objective of our work is to see whether the use of methods such as genetic algorithms and neural networks is an effective strategy to approach the complex problem of printer spectral modeling. In our experiments, we make no assumption concerning the sequence of inks during print and the printers are treated as RGB devices (the printer-driver operations are included in our model). To test our characterization methods, which require only 133 measurements for training the learning algorithms, we consider different inkjet printers and use different kinds of paper and printer drivers. The test set used for model evaluation was consisted of 777 samples, uniformly distributed in the RGB color space.

Introduction

The spectral-based characterization of inkjet printers requires a mathematical description of the printing process, that is a function that converts between digital counts and reflectance spectra: a look-up table approach, indeed, would require an excessive amount of data. Many methods have been proposed for the spectral-based characterization of printers¹⁷, most of them based on the Neugebauer equation. Clearly, the Neugebauer model alone can not foresee with enough accuracy the reflectance spectrum of a printed color, as the effects of interaction of inks with paper and interaction among inks are not accounted for. Many authors have suggested strategies to model mechanical and optical dot-gain^{2,3,4,5}; the majority of them try to understand how the physical placement of each ink determines its contribution to the final reflectance. In these methods, therefore, the printer driver plays a role.

In our work we follow a different approach. We aim to see if the use of methods such as genetic algorithms and neural networks is an effective strategy to meet the requirement of a mathematical solution to the problem of

the spectral modeling of inkjet printers, while avoiding a deep knowledge of the printing process.

In the genetic approach, we propose an analytical model based on the Yule-Nielsen Spectral Neugebauer equation, formulated with a large number of degrees-of-freedom in order to take into account dot-gain, inks interaction and printer driver operations. To estimate the model's parameters we use genetic algorithms. In the neural network approach we train a neural network to compute reflectance spectra from RGB digital counts. For comparison, we consider the same training and test set for both the methods.

Both approaches have been tested on a set of printers using different printer drivers and kinds of paper. In our experiments we make no assumption concerning the sequence of inks during the print and the printers are treated as RGB devices.

Genetic Algorithms

Genetic algorithms are a general method for solving optimization problems, inspired by the mechanisms of evolution in biological systems. In the basic genetic algorithm (GA), every candidate solution for the optimization problem is represented by a sequence of binary, integer, real, or even more complex values, called an individual. A small number n of individuals (with respect to the cardinality of the whole solution space) is randomly generated as an initial population P . Then, the GA iterates a procedure that produces a new population P' from the current P , until a given "STOP" criterion is satisfied. Each iteration consists of the following steps:

- fitness evaluation: for every individual x in P , the value $f(x)$ of a suitable "fitness" function is computed;
- selection: $n/2$ pairs of individuals are randomly selected from the population P ; the probability of selection is higher for individuals of greater fitness;
- crossover: two new individuals (sons) are generated by cutting the two elements of each pair (parents) at a randomly chosen point and interchanging the four parts thus obtained;
- mutation: the value of each position of the elements in P' is changed with a given probability p_m .

*zuffi@itc.cnr.it, <http://www.itim.mi.cnr.it/Linee/Linea1/persone/zuffi.htm>, [schettini, mauri]@disco.unimib.it

The main advantages of using the genetic approach are that it allows to manage simultaneously many parameters, and can deal with irregular training data sets. The disadvantages are that it cannot guarantee an optimal solution, and that it is, in general, also difficult to tune the genetic algorithms' free parameters.

The Printer Model

The model of the printer we refer to is based on the well-known Yule-Nielsen Spectral Neugebauer (YNSN) equation. According to the YNSN model, the spectrum of a N -inks halftone print is the weighted sum of 2^N different colors, called Neugebauer primaries, given by all the possible overprints of inks. The weight of each Neugebauer primary is the area it covers in the halftone cell. The YNSN model for a 4-ink halftone print is:

$$R_{print,\lambda} = \left[\sum_{p=0}^{15} a_p R_{p,\lambda}^n \right] \quad \lambda = 1..8 \quad (1)$$

where $R_{print,\lambda}$ is the reflectance of the printed color, n is the Yule-Nielsen factor, $R_{p,\lambda}$ is the reflectance of the p -th Neugebauer primary, and a_p is the primary area coverage. In our work, reflectance spectra are sampled at intervals of 40 nm in the range from 400 nm to 680 nm, producing 8 samples.

The Yule-Nielsen factor is an empirical correction that has been applied to the Neugebauer model to account for the optical dot gain effect due to the scattering of light in the paper.⁶ The area coverage is the percentage of the halftone cell covered by the Neugebauer primary, and can be computed as expressed in Table 1, where $\mathbf{c} = [c, m, y, k]$ represents the concentration of inks for printing a given color.

Table 1. Calculation of the Area Coverage for Each Neugebauer Primary Given the Concentration of Inks.

Index (p)	Neugebauer primary	Area coverage (a_p)
0	W	$(1-c)(1-m)(1-y)(1-k)$
1	K	$(1-c)(1-m)(1-y)k$
2	Y	$(1-c)(1-m)y(1-k)$
3	YK	$(1-c)(1-m)yk$
4	M	$(1-c)m(1-y)(1-k)$
5	MK	$(1-c)m(1-y)k$
6	R	$(1-c)my(1-k)$
7	RK	$(1-c)myk$
8	C	$c(1-m)(1-y)(1-k)$
9	CK	$c(1-m)(1-y)k$
10	G	$c(1-m)y(1-k)$
11	GK	$c(1-m)yk$
12	B	$cm(1-y)(1-k)$
13	BK	$cm(1-y)k$
14	CMY	$cmy(1-k)$
15	CMYK	$cmyk$

We consider each printer as an RGB device, therefore assumptions must be made on how the printer driver computes the CMYK amounts, given an RGB value. Usually the black replaces the "gray component" of the tone. The gray component is a neutral gray that results from removing the smallest of the CMY components together with the amount of the other two colors necessary to form, all together, a neutral gray. This process is named Gray Component Replacement.⁸ We model the driver according to Equation 2, where the gray component is composed of equal amounts of cyan, magenta and yellow.

$$\begin{aligned} c &= c' - U \cdot k, \\ m &= m' - U \cdot k, \\ y &= y' - U \cdot k, \\ k &= R \cdot \min(c', m', y'), \\ 0 &\leq U \leq 1; R = 1 \end{aligned} \quad (2)$$

where

$$c' = 1 - r, m' = 1 - g, y' = 1 - b.$$

The real size of a dot printed on a substrate is larger than its theoretical size, due to the spreading of the ink on paper. The area coverage, therefore, cannot be computed from the theoretical amount of ink. To account for this phenomenon, a non-linear function relating the theoretical concentration of inks with the effective concentration is employed for any possible overprint of inks. With the aim of using the smallest possible number of parameters for the model, we looked for a function that could describe dot gain and be tuned with only one parameter. The function used is:

$$c_\lambda = \frac{c_t}{(1 - C_\lambda) \cdot c_t + C_\lambda} \quad (3)$$

where c_t is the theoretical concentration of ink, computed from RGB values using Equation 2, and C_λ is the function parameter.

Dot gain functions are usually estimated to model the spreading of inks on paper, nevertheless the spread of ink on paper may differ from its spread on a previously deposited ink. Different strategies can be employed to account for this phenomenon^{2,3,5}; we chose to take into account the interaction of the inks, providing a different dot gain function depending on the presence of inks. Moreover, in our model effective concentrations are wavelength-dependent.

Table 2 lists the parameters used to compute the effective concentration of inks, using Equation 3, to obtain the area coverage for any Neugebauer primary according to Table 1.

Table 2. Dot Gain Parameters for Computing Effective Ink Concentration Using Equation 4.

Index (p)	Neu. primary	Dot gain parameters			
1	K	$C_{ck,*}$	$M_{mk,*}$	$Y_{yk,*}$	$K_{k,*}$
2	Y	$C_{cy,*}$	$M_{my,*}$	$Y_{y,*}$	$K_{ky,*}$
3	YK	$C_{cyk,*}$	$M_{myk,*}$	$Y_{yk,*}$	$K_{ky,*}$
4	M	$C_{cm,*}$	$M_{m,*}$	$Y_{ym,*}$	$K_{km,*}$
5	MK	$C_{cmk,*}$	$M_{mk,*}$	$Y_{ymk,*}$	$K_{km,*}$
6	R	$C_{cm,*}$	$M_{my,*}$	$Y_{ym,*}$	$K_{kmy,*}$
7	RK	$C_{cmk,*}$	$M_{myk,*}$	$Y_{ymk,*}$	$K_{kmy,*}$
8	C	$C_{c,*}$	$M_{mc,*}$	$Y_{yc,*}$	$K_{kc,*}$
9	CK	$C_{ck,*}$	$M_{mck,*}$	$Y_{yck,*}$	$K_{kc,*}$
10	G	$C_{cy,*}$	$M_{cm,*}$	$Y_{yc,*}$	$K_{key,*}$
11	GK	$C_{cyk,*}$	$M_{cmk,*}$	$Y_{yck,*}$	$K_{key,*}$
12	B	$C_{cm,*}$	$M_{mc,*}$	$Y_{yc,*}$	$K_{kcm,*}$
13	BK	$C_{cmk,*}$	$M_{mck,*}$	$Y_{yck,*}$	$K_{kcm,*}$
14	CMY	$C_{cm,*}$	$M_{cm,*}$	$Y_{cm,*}$	$K_{cm,*}$
15	CMYK	$C_{cmk,*}$	$M_{cmk,*}$	$Y_{cmk,*}$	$K_{cmk,*}$

Subscripts in the dot gain parameters refer to the inks present in the corresponding Neugebauer primary. There are 32 different dot gain parameters in the model; each one is a vector, the dimension of which is the number of wavelengths considered. The area coverage of paper is computed as the difference between the sum of the area coverage of the inks and their overprints, with the constraint to be positive.

$$a_0 = 1 - \sum_{p=1}^{15} a_p \quad (4)$$

To effectively tune the model, the training set must be customized to employ all the parameters. The training set consists therefore of ramps of eleven patches, ranging from the absence of ink to full ink coverage of cyan, magenta, yellow, red, green, blue, black, cyan with black, magenta with black, yellow with black, red with black, green with black and blue with black, for a total of 143 samples.

The Genetic Algorithm Adopted

A genetic algorithm is used to estimate the printer model parameters described above. The choice of the representation of candidate solutions (individuals) and the fitness function used to evaluate individuals are crucial factors in the effectiveness of the genetic approach.

The genetic material of each individual, called the genome, must consist of the minimum amount of data requested to represent a solution to the problem. In our work, a genome consists of an array of real numbers.

In the printer model, we have introduced:

- the Yule-Nielsen factor (Equation 1),
- U, in the printer driver model (Equation 2),
- and 32 parameters for the dot gain functions (Table 2).

The dot gain parameters are wavelength-dependent, giving us a total of 258 real numbers.

As the genome is an array of real numbers, a range must be specified. We consider that the dot gain functions

do not alter the theoretical value of ink concentration by more than some 30%: consequently the range for real parameters is set at [0.3; 3.0]. The theoretical value of the Yule-Nielsen factor ranges from 1.0, corresponding to the absence of scattering, to 2.0, corresponding to Lambertian or perfectly diffused scattering, with the assumption that the dots are rectangular in cross section. In reality, the dots have soft transitions, and in cases of high frequency rotated screens, or error diffusion, much of the paper is covered by transitory regions. In these cases, if the Yule-Nielsen factor is experimentally computed, it may exceed the theoretical limit of 2¹. In our experiments, we consider a range [1.0; 12.0] for the Yule-Nielsen factor.

The fitness function is computed as:

$$fitness = \frac{1}{S} \cdot \left(\sum_{s=1}^S \frac{1}{\Gamma} \cdot \left(\sum_{\lambda=1}^{\Gamma} (R_{print,\lambda,s} - R_{meas,\lambda,s})^2 \right) \right) \quad (5)$$

where S is the number of elements in the training set (S=143), Γ the number of wavelengths ($\Gamma = 8$), and R_{print} is computed with Equation 1.

We use the 'simple' genetic algorithm in the Galib⁹ library. It employs non-overlapping populations: at each generation the algorithm creates an entirely new population of individuals by selecting from the previous population, and then mating to produce the new offspring. The best individual from each generation is also carried over into the next generation (elitism). The probability of crossover and mutation are set at 0.9 and at 0.002, respectively. Selection is based on the "roulette wheel" method. The initial population of 12 individuals is randomly selected; only the initial value for the parameter U in the conversion from RGB to CMYK (Equation 2) is initialized at 1. The stopping criterium is the number of iterations performed: we consider 4000 iterations.

Neural Networks

Neural networks are discrete time computational systems composed of interconnected processing units (called neurons) which are inspired by biological nervous systems.¹⁰ As in nature, the interconnections between elements (weights) determine the network function and are adaptively updated on the basis of examples to model the data.¹¹ More precisely, at each time step every neurons updates its state as a function (e.g., a sigmoid function) of the weighted sum of the inputs it receives from other neurons, and in turn outputs a signal which is function of its state. A neural network is specified by its basic components: the processing units (neurons), their connections, and the learning rule.

The printer model has been approximated by means of a feed-forward neural network trained with backpropagation.

Multiple-layer feed-forward neural networks consist of several distinct layers of neurons. The first, or input layer, serves as a holding site for the values to be processed; the last, or output layer, is the point at which the final state of

the network can be read. Connections can only go from neurons of one layer to the neurons of the next layer. Back-propagation provides the prescription for changing the weights of any feed-forward network so that it can learn to compute a function from a set of input-output data pairs (training set). Standard back-propagation is a gradient descent technique designed to reduce the error between the actual and the desired output of the network.

Of course the interest does not lie in learning a particular training set, but in building networks that can generalize, that is behave correctly in new cases. Properly defined networks with biases, an hidden layer with sigmoid activation function, and a linear output layer are capable of approximating any function with a finite number of discontinuities. Since there are no theoretical results to help in choosing of the architecture (number of hidden layers and neurons per layer), this is usually determined either on the basis of previous experience with a given problem domain, or experimentally by a time-consuming activity of training and testing different architectures. To improve generalization, usually the training data are divided into a training and a validation set. In the training procedure, following each epoch, the performance of the network is evaluated on both the training and validation sets. As long as the network performance improves on both the data sets the learning is continued, when the error on the training set still decreases but the network shows poorer performance on the validation set the learning phase is stopped in order to avoid network overfitting of the training data.

The Neural Network Adopted

We used the Matlab Neural Network Toolbox to implement the neural network. For back-propagation we applied a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization and minimizes a combination of squared errors and weights to produce a network which generalizes well (Bayesian regularization). We consider as data set for the training phase, the same set of colors used for the genetic algorithm solution. The data set of 143 colors is divided into a training and a validation set, simply considering alternatively one color for one set and the successive for the other one. The network is composed by a sigmoid layer of 10 neurons and a linear output layer. The network dimension has been arbitrarily chosen; similar configuration have been however tested and those adopted gave actually better results. In the training procedure a maximum of 100 epochs is admitted.

Experiments and Results

We apply our model to describe four inkjet printers:

- Epson Stylus Color
- HP 2000C
- Epson Stylus C80
- Epson 890

With the Epson Stylus Color and the Epson 890 we adopt a driver that employed Floyd Steinberg dithering, while for the Epson Stylus C80 and the HP 2000C, we adopt the drivers from the printer manufacturers, disabling any color management or color enhancement. Different types of papers are used (plain paper and Epson Photo Quality paper).

The characterization procedure starts with the printing and measurement of the Neugebauer primaries and the training set (Fig. 1). The Neugebauer primaries are obtained by measuring the printed inks at full coverage, and their overprints, by successive prints on the same sheet. The test set consist of 777 samples, uniformly distributed in the RGB color space (Fig. 2). Measurements of the spectra are executed with a Gretag Spectrolino, considering values in the wavelength range from 400 to 680 nm with a step of 40 nm. Reflectance spectra are in the range [0;100].

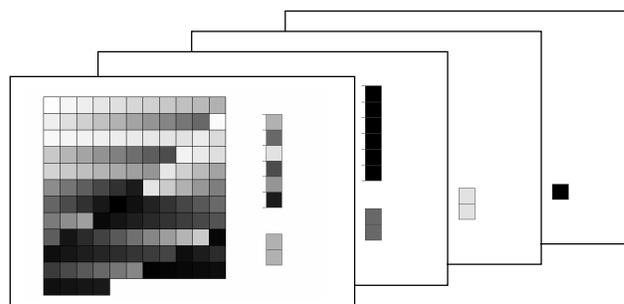


Figure 1. Specimen for the training set and set of images to print in sequence on the same sheet for the Neugebauer primaries measurement (133 patches).

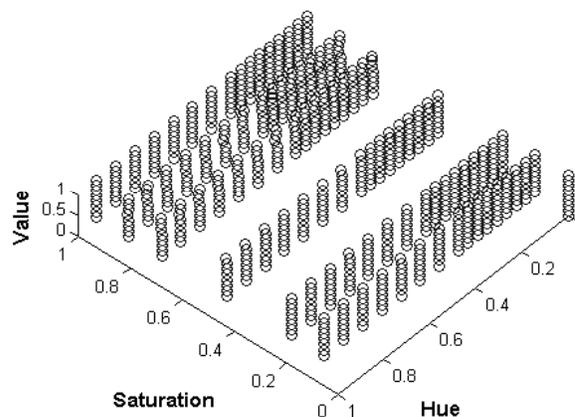


Figure 2. HSV diagram of the 777 colors in test set specimen.

The results are reported in terms of color difference in CIELAB ΔE_{ab}^* and CIELAB ΔE_{94}^* ¹², hue and lightness difference,¹³ and root mean square error. In Table 3a and 3b are reported the error statistics for the genetic algorithm solution (training set and test set respectively). In Table 4a

and 4b are reported the results for the neural network solution. Statistics for the training set refers to the whole set of 143 colors, comprising the training and the validation set.

Table 3a. Genetic algorithm: statistics of color distances and spectra differences for the training set (m = mean, M = maximum, sdv = standard deviation).

	Epson 890 photo quality paper			Epson Stylus Color photo quality paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	2.30	7.63	1.80	1.49	6.47	1.15
ΔH	0.93	4.61	1.02	0.70	3.67	0.74
ΔL	0.96	6.25	1.01	0.52	3.27	0.56
ΔE_{94}^*	1.85	6.95	1.43	1.18	4.85	0.91
RMS	0.86	3.43	0.73	0.65	2.34	0.41
fitness	1.26			0.59		
	HP 2000C plain paper			Epson Stylus C80 plain paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	2.29	8.02	1.67	1.67	5.93	1.21
ΔH	0.87	3.29	0.82	0.78	3.33	0.71
ΔL	0.79	3.56	0.73	0.92	5.24	1.11
ΔE_{94}^*	1.71	5.77	1.25	1.47	5.85	1.18
RMS	1.12	3.23	0.82	0.96	4.69	0.68
fitness	1.92			1.38		

Table 3b. Genetic algorithm: statistics of the color distances and spectra differences for the test set.

	Epson 890 photo quality paper			Epson Stylus Color photo quality paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	3.21	11.0	2.05	1.92	7.99	1.22
ΔH	1.36	7.87	1.26	0.87	6.31	0.97
ΔL	1.70	7.97	1.64	0.9	3.41	0.71
ΔE_{94}^*	2.78	8.54	1.72	1.64	6.03	0.95
RMS	1.42	4.49	0.85	0.90	5.82	0.56
	HP 2000C plain paper			Epson Stylus C80 plain paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	5.4	18.5	3.12	2.74	8.51	1.45
ΔH	2.01	12.2	1.88	0.97	6.00	0.76
ΔL	1.74	6.62	1.31	1.74	7.98	1.56
ΔE_{94}^*	4.09	11.0	1.95	2.51	8.45	1.42
RMS	2.24	12.2	1.34	1.72	5.37	1.15

Conclusion

In this work we experimented two novel approaches for the spectral characterization of inkjet printers. Our objective was to see whether it is feasible to approach the complex problem of printer spectral modeling by using genetic

algorithms and neural networks. If this were so, using our method we would be able to spectrally characterize a generic inkjet printer with only 133 measurements, performed manually by an operator, without any of the equipment needed to measure the large number of samples required for colorimetric printer characterization based on interpolation methods.

Table 4a. Neural network: statistics of color distances and spectra differences for the training set.

	Epson 890 photo quality paper			Epson Stylus Color photo quality paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	2.08	7.57	1.39	2.13	6.87	1.42
ΔH	0.97	4.45	0.91	0.95	6.63	1.02
ΔL	0.81	4.82	0.85	0.72	3.26	0.68
ΔE_{94}^*	1.65	5.30	1.15	1.66	5.62	1.15
RMS	0.90	4.78	0.63	0.95	2.78	0.44

	HP 2000C plain paper			Epson Stylus C80 plain paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	3.70	15.2	2.70	2.48	8.11	1.73
ΔH	2.02	7.62	1.98	1.14	5.50	1.14
ΔL	1.02	6.28	1.12	0.76	4.27	0.77
ΔE_{94}^*	2.89	12.3	2.20	2.01	7.30	1.52
RMS	1.63	3.95	0.70	1.09	2.71	0.50

Table 4b. Neural network: statistics of the color distances and spectra differences for the test set.

	Epson 890 photo quality paper			Epson Stylus Color photo quality paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	3.26	12.2	2.17	2.56	14.3	1.89
ΔH	1.46	9.70	1.60	1.20	6.67	1.04
ΔL	1.55	6.87	1.29	1.12	7.67	1.20
ΔE_{94}^*	2.80	9.25	1.74	2.18	9.58	1.48
RMS	1.43	4.47	0.84	1.38	7.39	1.07
	HP 2000C plain paper			Epson Stylus C80 plain paper		
	m	M	sdv	m	M	sdv
ΔE_{ab}^*	6.53	24.0	3.77	3.66	10.1	2.24
ΔH	2.93	23.7	3.25	1.87	8.87	1.71
ΔL	2.54	8.53	1.81	1.65	7.45	1.38
ΔE_{94}^*	5.12	17.7	2.48	3.25	9.52	1.98
RMS	3.09	12.7	1.79	1.61	5.66	0.97

In the genetic approach we introduced a set of parameters into the mathematical framework of the Yule Nielsen Spectral Neugebauer equation. The results reported

indicate that a characterization is possible, with mean ΔE_{94}^* ranging from 2.18 to 5.12, and a maximum hue difference of 12.2.

In the neural network based approach, we trained a multiple-layer feed-forward network with back-propagation on the same training set of the genetic solution and obtained, on the same test set, mean ΔE_{94}^* ranging from 1.64 to 4.09, and a maximum hue difference of 23.7.

This means that, in the absence of a satisfactory model, the networks could constitute an efficient way of realizing the spectral-based printers characterization.

References

1. D. R. Wyble, R. S. Berns, A Critical Review of Spectral Models Applied to Binary Color Printing, *Color Research and Application*, **25**, 1 (2000).
2. E. J. Stollnitz, V. Ostromoukhov, D. H. Salesin, Reproducing Color Images Using Custom Inks, *Proc. of SIGGRAPH'98*, in *Computer Graphics Proceedings*, Annual Conference Series, pg. 267-274, (1998).
3. K. Iino and R. S. Berns, A Spectral Based Model of Color Printing that Compensates for Optical Interactions of Multiple Inks, AIC Color 97, *Proc. 8th Congress International Colour Association*, (1997).
4. D. Tzeng, R. S. Berns, Spectral-Based Six-Color Separation Minimizing Metamerism, *The Eighth IS&T/SID Color Imaging Conference*, Scottsdale, Arizona, USA, November 7-10, (2000).
5. P. Emmel, R. D. Hersh, Modeling Ink Spreading for Color Prediction, *Journal of Imaging Science and Technology*, **46**, 3, (2002).
6. J. A. Stephen Viggiano, Modeling the Color of Multi-Colored Halftones, *Proc. TAGA*, pg. 44-62, (1990).
7. R. Balasubramanian, Optimization of the spectral Neugebauer model for printer characterization, *Journal of Electronic Imaging*, **8**, 2, pg.156-166, (1999).
8. G. G. Field, *Color and Its Reproduction*, Second Edition, GAFT Press, Pittsburgh, 1999.
9. Galib: A C++ Library of Genetic Algorithm Components, <http://lancet.mit.edu/ga/>
10. W.S. McCulloch, W.H. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.* 5: 115-133, (1943).
11. B. Muller, J. Reinhardt, M. T. Strickland, *Neural networks: an introduction*, Springer, Berlin, 1995.
12. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley, Reading, Massachusetts, 1998.
13. G. Wyszecki, W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, Second Edition, Wiley and Sons, 1982.