Dot gain table and developer voltage prediction for the HP Indigo press

Carl Staelin, Ruth Bergman, Mani Fischer, Darryl Greig, Marie Vans, Gregory Braverman, Shlomo Harush, and Eyal Shelef
HP Laboratories Israel HP Indigo Division
Haifa, Israel Rehovot, Israel

Abstract

Color consistency is crucial for both photo and commercial printing applications. Dot gain tables are currently updated sporadically, and between updates colors can shift due to process drift in the press. The goal of this investigation is to dynamically control the dot gain table and developer voltage to ensure more consistent color control while minimizing waste and calibration measurements.

In this article we approach the elements of this calibration process as a series of machine-learning problems and investigate the efficacy of replacing physical calibration measurements with model-based predictions. The current state of the machine, expressed as sensor measurements, are used to model both the developer voltage, and the subsequent dot gain look up table. We also consider models that make a prediction based on a restricted set of calibration measurements, not necessarily including the full machine state vector. Our initial investigation using a preliminary dataset shows that machine learning methods are suitable for predicting the dot gain table.

Introduction

Color consistency is crucial for both photo and commercial printing applications. Look up tables (LUTs) for estimating dot gain values are currently updated on demand when the operator notices color consistency problems, and between updates colors can shift due to process drift in the press. The goal of the work presented in this article is to dynamically control the dot gain table and developer voltage to ensure more consistent color control while minimizing waste and calibration measurements.

Currently the dot gain table and developer voltage are controlled by printing special calibration test patterns on demand which are measured internally by the press. The calibration process begins by printing one or more test patterns with 100% ink coverage to determine a developer voltage setting for each ink such that the ink thickness at 100% coverage is within specification. Once the developer voltage is set, the actual ink thickness or optical density at 100% coverage is measured. Finally one or more sheets of test patterns with monochromatic swatches of uniform digital dot area are printed to measure the physical dot area for each of the digital dot areas. These measurements are the values required for the dot gain.

There are two distinct phases in this process which may be formulated as machine learning problems: (1) predict the developer voltage and corresponding ink optical density at 100% coverage per ink given the current machine state, and (2) predict the dot gain table values for each digital dot area of interest for each ink given the current machine state, developer voltage, and ink optical density at 100% coverage. A related problem, predicting the dot gain table given one or more measured dot values without any state information, is also examined here. A large number of machine learning regression algorithms are applicable to these problems. We evaluate the accuracy of three common methods: artificial Neural Networks (NN), Support Vector Machines (SVM), and linear regression. Neural networks are a well-known technique for machine learning. Both Bishop [1] and Ripley [2] give an excellent and readable treatment of theory and methods. Support vector machines are a kernel-based approach to machine learning. A good tutorial introduction to SVM was written by Burges [3].

If a method is found to supply sufficiently accurate predictions, we can replace or augment the calibration procedure with a prediction-based process that has much less impact on customer workflow and consumable usage. The minimal requirements for the HP Indigo press are that the absolute difference between the predicted dot area and printed dot area is less than 2 at least 67% of the time.

The dot gain is defined as follows:

$$dot\,gain = \frac{printed\,dot\,area}{digital\,dot\,area} \tag{1}$$

Both the digital dot area and printed dot area are expressed as a percentage of the area that is covered, where 100 means that the whole area is covered with ink. The dot gain table contains the printed dot area value, used in Equation 1, for each digital dot area of interest.

The calibration process uses an inline optical densitometer to read the printed dot areas from a swatch of uniform density in a single color. The current physical constraints of the system allow up to fifteen such swatches on a single sheet. Since the presses can have up to seven separations (inks), this implies that we may measure up to two digital dot areas for each separation in a single sheet.

The densitometer used for LUT calibration approximates the true dot area with a small measurement error. The practical accuracy of any model-based predictions is limited by the accuracy of the densitometer. In some cases, the models suggested here approach this accuracy limit.

As an alternative to the full calibration process, we consider a "fast calibration" process that measures one, two or more points per color separation, and then uses the measured information and the machine state to predict the rest of the dot gain lookup table values.

We analyzed a dataset of dot gain LUT's collected by HP Indigo. Our results for this dataset are promising in that the models give predictions within the required limits. It is important, however, to keep in mind that this dataset is small by machine learning standards - approximately 130 samples for each ink separation and halftone screen.

HP Indigo dataset

The various parameters registered in the HP Indigo dataset are best understood by a brief introduction to the printing process of the HP Indigo Press.

The printing process

The process of image production consists of three stages (see Figure 1). The first step is image generation in which a latent image is created on the Photo Imaging Plate (PIP) foil. The second step is image development. During this stage the latent image is developed by ink on the PIP. The last step is image transfer in which the developed image is transferred from the PIP to the Blanket that wraps the Intermediate drum (ITM). At this stage, the developed image is transferred from the Blanket to the substrate.

This process of converting a digital signal to a physical dot on a piece of paper can be affected by any number of system elements and interactions. Many key elements, such as the PIP foil and blanket are regularly replaced and each replacement part has its own characteristics. Thus, it is likely that a full dot gain table measurement will need to be taken after each major part replacement. In addition, during normal operation other parameters, such as temperature, vary continuously.

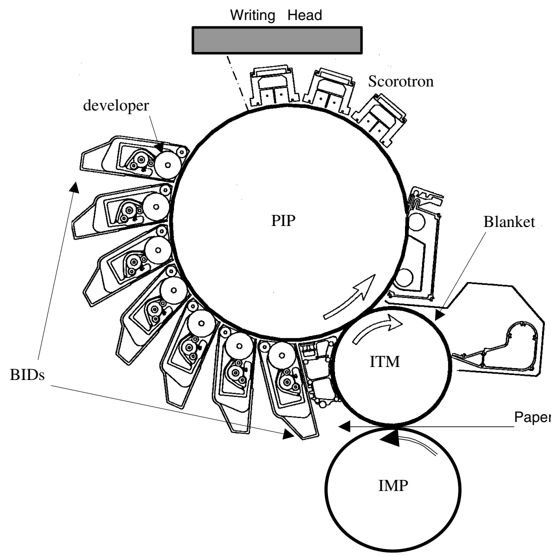

Figure 1: *Indigo Press.*

Dataset information

The HP Indigo dataset includes parameters and dot gain LUTs collected over a one week period from a single Series I Indigo Press by a single operator using the automatic calibration process. Note that Figure 1 actually refers to the more sophisticated Series 2 Press, however the essential printing process remains largely the same between the two models.

The dataset contains 269 dot gain tables each containing fifteen printed dot area values for each of the four separations Black, Cyan, Magenta and Yellow. Associated with each set of printed dot area values (the dot gain LUT) are a set of observed parameters, twelve of which are common across all separations, such as the ITM temperature, and seven others which vary according to the current separation, e.g., ink characteristics and the developer voltage. For a more detailed description of the dataset refer to [4].

Due to space constraints in this document, all the graphs shown are for the 175lpi HDI-175 screen.

Dot gain prediction results

There are a variety of subproblems under the general problem of dot gain LUT prediction, which are appropriate for different usage models. The subproblems we consider here fall into four general categories: (1) dot gain LUT predictions using only machine state measurements (i.e. no consumables are required); (2) parameter ranking / relevance for problem (1) models; (3) dot gain LUT predictions using the measurement of one or more dot gain values (i.e., without measuring any machine state variables).

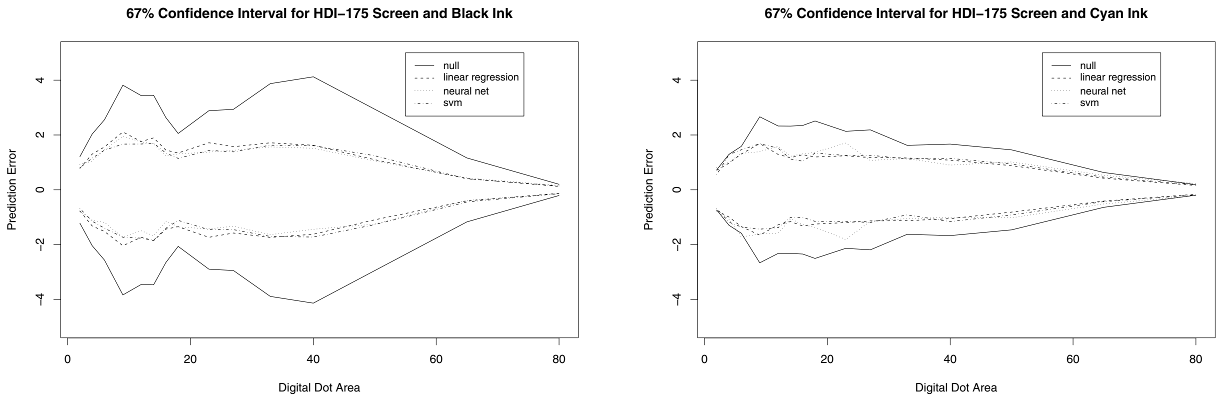

Figure 2: *Prediction Error 67% Confidence Interval.*

For the prediction problems, the prediction errors are the difference between the true printed dot area and the predicted print dot area. The prediction errors were analyzed using a Chi-squared goodness of fit test and found that they are approximately normally distributed. Therefore, we can use the normal distribution multipliers for computing confidence intervals. The results are presented on graphs with the x-axis giving digital dot area, and the y-axis the difference of means. For a given model, a confidence "envelope" is plotted on this axis. That is, points corresponding to the upper limits of the confidence intervals for each digital dot area are joined to form a line, and the lower limits form a second line. This is simply for readability, since we often wish to compare multiple models on a single graph axis.

The graphs in Figure 2 show the 67% prediction error confidence intervals for each of the three machine learning methods: linear regression, neural networks, and support vector machines as a function of the digital dot area. Note that each separation behaves slightly differently. The null model prediction for each separation is also included for comparison.

From the results in Figure 2, it is apparent that the behavior of all the machine learning methods is similar. This means that the prediction of "hard" points is invariant of the learning method. Since linear regression performs comparably to the more complex non-linear methods, all further analysis was done using linear regression. Note that the null model does not supply acceptable predictions.

Parameter ranking & selection

Some of the parameters used in the dot gain prediction models may be redundant. That is, we may find smaller models that fit the data equally well by removing some parameters. This has a two fold advantage - smaller models are more efficient computationally, and they are less prone to overfitting. Furthermore, by analyzing the input parameters, we can gain some insight into the machine operation that may help the manufacturer identify other issues with parameter control etc.

The importance of a parameter may be measured by the effect of removing that parameter from the model. If the predictive power of the model is unaffected (or even improved) we may conclude that the parameter is not significant. On the other hand, if the model has a significant degradation in performance without a particular parameter, we may conclude that this parameter is significant and should be retained.

To implement this method we proceeded in the following manner. First model predictions were obtained for the entire dataset using the whole set of predictors. The predictions were obtained using the 10-fold cross validation technique.

The sum of squared errors (SSE) of these predictions was computed on each of the digital dot values of the LUTs, where both screens and all separations were included in this sum. Then similar predictions and SSE computations were made for models fitted *excluding* each of the input parameters in order.

We repeated this experiment 20 times and from this we were able to estimate the mean and standard deviation of the SSE value for each of the table entries. Using these estimates we can generate confidence intervals for the differences of means between the original (full) model and each of the depleted models, for each DDA value.

Figure 3 summarizes these results. The error bars give the 95% confidence interval for the mean of the relevant depleted model SSE minus the mean of the full model SSE. If the confidence interval does not include the zero line in a particular case, then we can conclude (at the 0.05 level) that the parameter under consideration is relevant for the

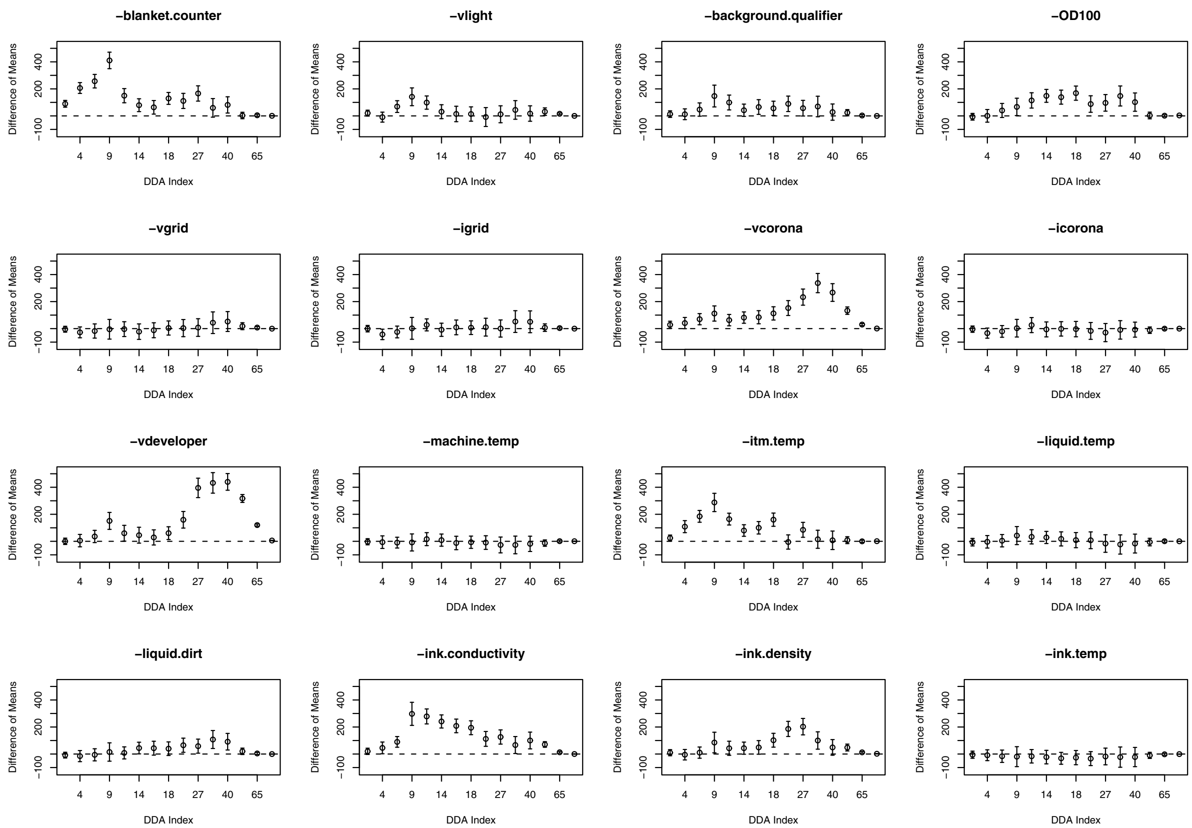

Figure 3: *Difference of means 95% confidence intervals for depleted versus full linear regression models.*

dot gain LUT prediction. From the graphs in Figure 3, we can determine which parameters are significant to the prediction and which parameters are not significant. Note also that the significant parameters are effective at different DDA values. For example, the *blanket.counter* variable is effective mostly at the lower DDA values, while *vdeveloper* has the biggest effect at the high range of DDA values. We can also use this technique to obtain a ranking of the importance of each input parameter (see [4]).

Prediction with one or more measured points

The results in the previous section suggest that we can predict the dot gain LUT to within specification requirements by measuring the machine state. An alternative to eliminating the calibration process is reducing the waste due to calibration with a "fast calibration" where some digital dot area patches are printed and the printed dot area measured. In this section we quantify the prediction quality when using both one and two measured points.

This prediction problem immediately raises the issue of which points to add. From prior figures, such as Figure 2 we can see that the error distribution tends to be bi-modal, with the smaller points, e.g., 2-16 covering one regime, and the middle points, e.g., 27-50 covering another regime.

Figure 4 shows the results of fitting models using a single measured LUT value, and predicting the remainder of the LUT based on this single value. For this case we used neural network models rather than linear regression models to add some non-linearity to the predictions. The measured points in Figure 4 are those that gave the best prediction results. These results suggest that such models can meet the accuracy requirement almost all the time.

Figure 5 shows the results of fitting models from two measured LUT values. Two sets of points are shown {27 & 50}, which gave the best single point prediction, and {16 & 40} which appear to give the best two point prediction. As expected, two point prediction gives better prediction results than one point and compares well to the prediction results using the machine state.

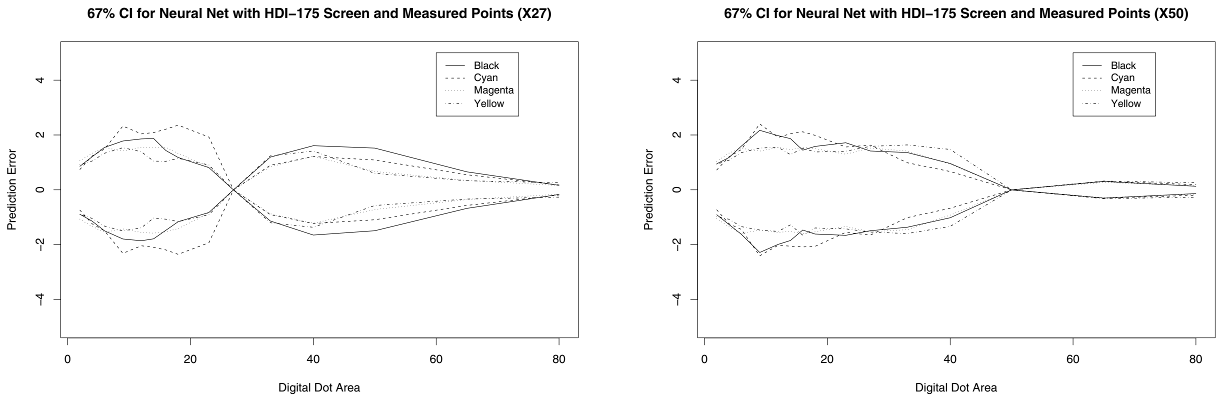

Figure 4: *Prediction Error Confidence Intervals for Neural Network Predictions using One Measured Point and No Machine State.*

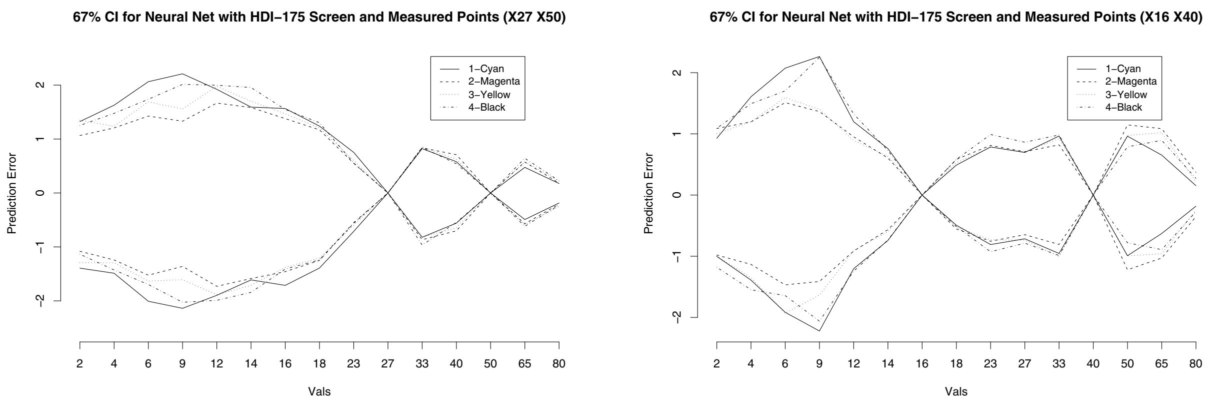

Figure 5: *Prediction Error Confidence Intervals for Neural Network Predictions using Two Measured Point and No Machine State.*

Conclusions

From the initial dataset it appears that given the measurable parameters from Table 1 we can predict the various dot gain values with acceptable accuracy using linear regression. This should allow HP Indigo to greatly improve the color consistency for their presses, while reducing both the consumable waste and work-flow disruption.

We suggested a method for assessing parameter importance. With this method we were able to conclude the some parameters do not significantly affect the model performance, and may, therefore, be eliminated. We also ranked the importance of the input parameters with respect to their effect on dot gain.

The results from predictions based only on one or two measured points, suggest that there may also be opportunity to reduce consumable usage in some circumstances by printing a reduced LUT set ("fast calibration"), or relying on existing values. It does appear that there is a strong enough relationship between point in a dot gain LUT to be exploited by these simpler models.

References

1. Christopher M. Bishop, Neural networks for pattern recognition, Oxford University Press, Oxford, UK (1995).
2. Brian D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press (1996).
3. Christopher J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2(2), 121, http://svm.first.gmd.de/papers/Burges98.ps.gz (1998).
4. Carl Staelin, Ruth Bergman, Mani Fischer, Darryl Greig, Marie Vans, Gregory Braverman, Shlomo Harush and Eyal Shelef, Dot gain table prediction for the hp indigo press, Tech. Rep. HPL-2002-355 (2002).