Classification of Paper Images to Predict Substrate Parameters Prior to Print

Matthias Scheller Lichtenauer, Safer Mourad, Peter Zolliker, and Klaus Simon

Swiss Federal Laboratories for Materials Testing and Research
Media Technology Lab
Ueberlandstrasse 129, 8600 Duebendorf, Switzerland
{matthias.scheller,klaus.simon,peter.zolliker}@empa.ch,
Safer.Mourad@fisba.ch
http://empamedia.ethz.ch/

Abstract. An accurate characterization of the substrate is a prerequisite of color management in print. The use of standard ICC profiles in prepress leaves it to the printer to match the fixed substrate characteristics contained in these profiles. This triggers the interest in methods to predict, if a given ink, press and paper combination complies with a given characterization. We present an approach to compare physical and optical characteristics of papers in order to achieve such a prediction of compliance by classification methods. For economical and ecological reasons it is preferable to test paper without printing it. We therefore propose non-destructive methods.

Keywords: image classification, color management, dotgain, tone value increase, ISO 12647, non-destructive testing.

1 Introduction

Print by Numbers. With digital workflows in graphical industry, digital data has become the dominant representation of color and customers of the printing industry increasingly expect repeatable and predictable results. In order to meet these expectations, the printing process has in recent years undergone standardization efforts. Paper as a printing substrate however remains a source of variability.

Empirically characterizing a particular ink, press and paper combination in offset print requires production of a set of plates, printing patches at different levels of ink provision, measuring with a spectrophotometer — and in the worst case iteration of the entire process. This means "producing a profile". The costs have to be amortized by print runs with the substrate characterized. To characterize the substrate for a single job can more than double the fixed costs in offset print. But fixed costs are the major drawback of offset relative to its competition subsumed under the term of digital printing technologies. Thus, offset has a high interest to reuse characterization of paper.
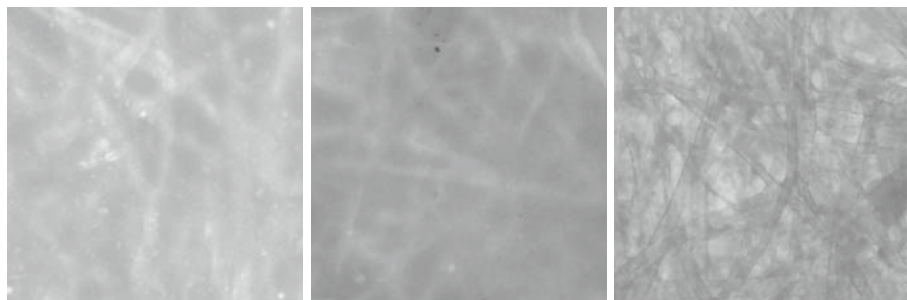
Fig. 1. Microscopic images of papers used in classification. The resolution of the original images was about $0.4\,\mu\text{m}$ per pixel. The area shown is about $0.4 \times 0.4\,\text{mm}$.

Find a Matching Profile. In digital publishing, color is specified in a device independent space. To map these color specifications to a print, substrate dependent information is interpolated from data in a look up table (ICC profile). An ICC profile describes color transformations for a single, fixed substrate and a fixed set of colorants. The users of profiles in prepress are interested to minimize profile handling expense by using a few generic profiles. So, instead of being used as a tool for matching data to a specific press room situation, ICC profiles became normative references printers have to meet. Therefore, the printing industry must find procedures to check, whether a given lot of a paper brand is within an acceptable range of variability relative to data given by an ICC profile. This is a classical categorization problem we try to approach with image classification methods.

Organisation of this Paper. In this paper, we will introduce a few paper properties known to impact color in print, assess their impact and the actions a printer can take. We then discuss our understanding of radiative transfer in paper substrate developed during the last years, in order to explain the choice of features used in classification. And finally, we will present our attempts to classify offset papers to then compare their performance to ISO 12647-2 standard.

2   Prior Knowledge

When passing from analog reproduction by photographic techniques to a digital workflow and digital plate production, the limitations of the process by a substrate are only met in the last stage [1], but still have to be accounted for earlier. When matching the digital specification of a color to a physical representation, the following properties of paper have to be taken into account:

Coating: Some papers consist of a body of mixed fibers and fillers only, while other papers are additionally surface coated in variable thickness and flatness. This influences paper gloss and roughness, the diffusion of ink components in the paper and the drying process of wet ink.

White Point: Paper color usually becomes the white point of the print gamut.

Dimensions: In most mass printing technologies, halftoning is used to achieve a color impression. Areas covered with ink dots are combined with areas left blank at a microscopic scale above 10 μm. This limit corresponds to 2400 dots per inch of a digital plate setter using nonperiodic screen. A paper fibre has a lateral extension of about 10 to 50 μm [2]. The extension of graphical paper or cardboard perpendicular to the printed surface is 50 to 500 μm, and coating has to level out thickness variations of 5 to 15 μm per side. The thickness of an eventual coating may vary in these dimensions. Spatial dimensions of a paper surface have an impact on minimal dot size.

Penetration: Some of the properties even depend on the paper/press/ink combination, for instance the colorimetric prediction of a print result by colorimetric measurements of wet prints in offset, due to varying penetration in the paper body over time.

Trapping: Impact printing processes bring ink on a plate, which is then pressed against the substrate and doing so transfers the ink to it. If another ink has already been printed at a particular place, there may be less ink accepted than on the bare paper next to it, a phenomenon known as trapping. The extent of this effect depends not only on material properties, but also on the sequence and timing of printing the different colors. Trapping also influences gray balance.

Dotgain: Dotgain is also known as tone value increase. Halftoned colors usually appear darker than predicted by the Neugebauer model [3], where tone is proportional to intended area coverage. Some of this accounts to lateral spreading of ink (mechanical or physical dotgain), but the major cause of this effect is lateral scattering of light in the substrate (optical dotgain, see Fig. 2).

In order to account for dotgain and trapping in impact printing, the separation and amplitude modulating halftoning algorithm can compensate for it, letting for instance a 34% tone area appear on the plate where a 50% tone level is intended on paper.

Plate correction, separation and print sequence as well as ink provision are the major factors a printer can influence, when the substrate is given. Therefore, dotgain, trapping and maximal ink provision are the parameters we have to estimate. Other parameters can be measured.

The Approach of ISO 12647. ISO provides normative references for color locations of papers and inks so as to justify the use of a few generic profiles in prepress and proofing. This has led to the standards ISO 12647 and ISO 2846. These norms do not provide models predicting achievable print results for given substrate parameters.

There are first principal models of turbid media based on radiative transfer [5]. Transfer theory models give some insight in the optical part of dotgain and solid ink color. We will discuss such models and our experience with applying them prior to position our classification approach relative to the one of ISO.
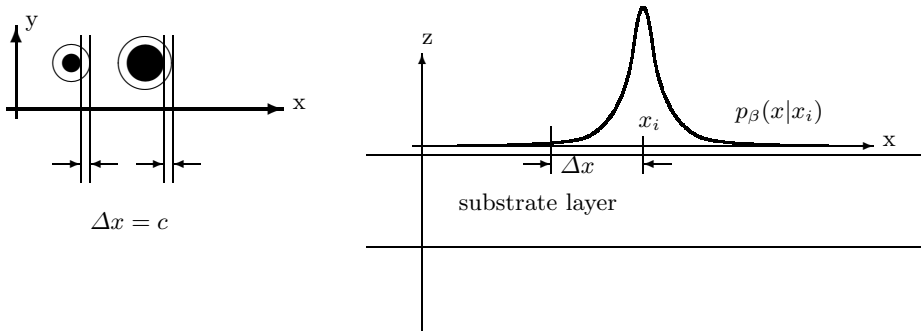
Fig. 2. In a simple model, optical dotgain can be seen as a convolution of a probability density function of remittance $p_\beta(x)$ and $S(\lambda, x)$, the spectral power distribution of incident light at position x in direction $-z$. Hence, light entering the substrate at x_i contributes to the exiting light in a small vicinity Δx of x_i. In the left half, we schematically show influence area of two ink dots.

Paper as a Turbid Medium. In a very thick substrate of infinite lateral extension, all photons will sooner or later get absorbed. Absorbed light might either be *extinct*, meaning that absorbed energy is completely transferred to heat, light may be *scattered*, meaning that a photon of same wavelength as the absorbed one is emitted in a different direction, or a *fluorescent* process takes place, the re-emitted photon is shifted in both, color and direction.

We are interested in the local distribution of remission $p_\beta(x|x_i)$ of a layer — also known as *Point Spread Function* — and the total fraction of incident light being remitted (Fig. 2). If we assume that $p_\beta(x|x_i)$ is similar across the whole printable surface, this principally explains sensitivity of dotgain to dot pattern and resolution.

Would $p_\beta(x, y|x_i, y_i)$ be known, the remittance $\beta(\lambda)$ of a halftone pattern could be approximated as integrated convolution $\int p_\beta \circ S_i(\lambda) \, dx dy$ with the spectrum of the incident light $S_i(\lambda)$. Incident and reemitted power spectra themselves can be convoluted with functions representing the halftone patterns. Should we also account for backside reflectance, multiple layers or fluorescence, the model will become far more complex.

The extinction p_α can be expressed as a product of probability density of length of way in a layer and constant extinction probability per unit length. If the incident beam is collimated, this will result in different $\int p_\alpha dx dy dz$ as with a diffuse incident light, since a photon's expected length of way in a substrate layer will differ [6]. It can be derived from simulations [7] that multiple scattering in a paper substrate is very likely to happen. These results coincide with experience in printing industry that coated papers exhibit lower dotgain compared to uncoated papers of the same mass per area. We attribute this to higher scattering in coating, resulting in a sharper peak of p_β. In solid patches on uncoated papers, some paper spots will shine through. This effect influences the solid ink color

and hence limits reproducible colors. These relationships of dotgain and solid ink saturation to resolution and coating are the principal insight we could gain. We now have to discuss the limitations we encountered when synthesizing these ideas to a radiative transfer theory model.

Transfer Theory Models. A family of such models dates back to 1931 [4], when Kubelka and Munk presented a model to determine the efficiency of paint in covering the background. They used probabilistic assumptions about the directional distribution of light to predict the effect of adding more paint on opacity.

It is not simple to apply a Kubelka-Munk-type model for offset halftone print, since coated papers can not be approximated as uniform bodies. Additionally, the penetration depth and UV-absorptance of inks vary, incident light is not perfectly diffuse and fluorescence may not be neglected. Furthermore, the model has to be extended in order to deal with halftone patterns. Nevertheless, there has been an attempt to do so in our group. The parameters of this model had to be determined by fitting. Mourad demonstrated its performance predicting single-color wedge dotgain from white and solid patch measurements [8]. Unfortunately, the numerical fitting approach showed tendency to overfitting in absence of calibration measurements of individual solid ink patches. Our model based on radiative transfer was inappropriate in a prediction *prior* to print.

3   A Non-destructive Classification Method for Paper

Most properties of a paper can be tested with non-destructive methods [9], but dotgain and trapping have yet to be determined by printing sheets of a paper with particular colorants on a particular press. This is expensive for a paper batch in quality control and inapplicable for a facsimile reprint. These are the reasons for the interest in non-destructive methods to predict dotgain, trapping, ink provision needed and color of solid ink. Transfer theory models don't predict these parameters. We therefore turned to develop a learning system, taking all prior knowledge into account and extrapolating it to a new, unknown sheet. Considering the relation between coating and light scattering discussed above, we postulated that thickness of the coating would be a distinguishing feature for dotgain.

In this section, we will present the experimental approach and the feature extraction methods. We printed a sample of commercially available offset papers under comparable conditions and empirically tested the correlations between substrate parameters measured, image analysis features extracted and colorimetric results achieved.

Printing Experiment on Paper Classification. When standardizing papers, the result should be similar when printed under similar conditions. We experimentally tested the assumption, that paper grades would show comparable results when printed with the same, uncorrected plate in one run with about the same amount of ink. Coherent results under such a condition would
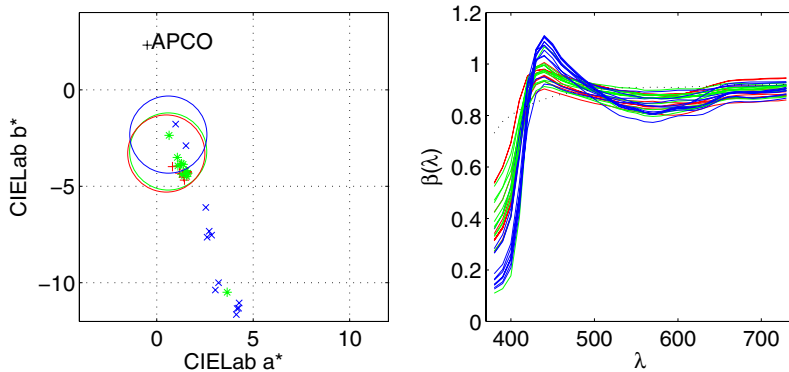
Fig. 3. Relative spectral power distribution of remission. We measured the remission spectra of 32 commercially available offset paper grades around 150 g/m^2 with GretagMacbeth Spectrolino ($45°/0°$, D50, $2°$) on white backing. Each curve shows the average of 20 measured sheets of one brand. The dotted line shows the spectrum of a yellowish, unbrightened APCO paper. The circles denote the location of paper white of the ISO 12647 standard (L* values of all papers are in the norm around 95).

justify the use of the same ICC profile for two papers. We intentionally used the same amount of ink for all papers, adjusted on a coated paper grade the printer was experienced with. Since uncoated offset papers would demand far more ink, colors on them were not as saturated as achievable. The good news is, that uncoated grades were coherently off the norm in solid ink tones. White points of most uncoated papers are off the norm by more than $5\,\Delta E_{ab}$ (Fig. 3).

To summarize, some of the commercially available paper grades we tested are not covered by the norm ISO 12647-2, but are comparable with regard to paper white, dotgain and ink acceptance. These observations suggest a need to generically classify papers with regard to white point, ink acceptance and dotgain characteristics.

Analysis of Paper Image. Since we assume a relationship of dotgain and ink acceptance with coating, we based our classification approach on transmission image analysis combined with physical and optical dimensions of the bare substrate. We made transmission images of all papers in the printing experiment with Leica laboratory microscope equipment, (Fig. 1).

Human Performance. We manually classified the images with regard to visibility of the paper fibres. The number of groups was a priori fixed to three. This allowed humans to achieve a categorization that correlates with dotgain (Fig. 4).

Automatisation. In order to achieve such a classification algorithmically, we investigated the correlation with the following features:
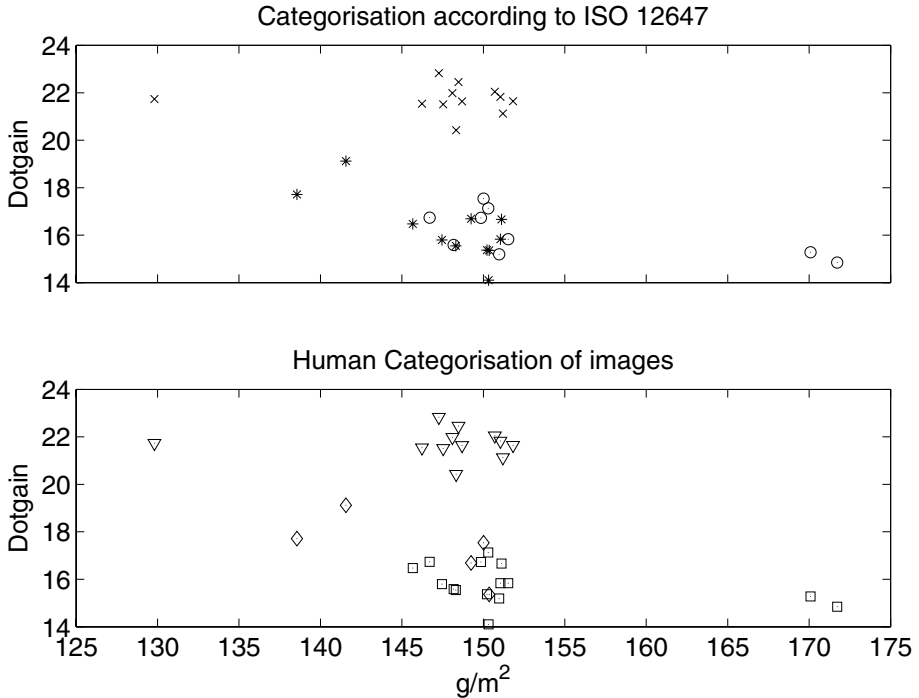
Categorisation according to ISO 12647

Human Categorisation of images

Fig. 4. Classification of paper with regard to dotgain. We tested 32 paper brands around $150 \, \mathrm{g/m^2}$. Each mark stands for the mean dotgain value of a gray patch printed on 20 sheets of a particular paper brand. The symbol of the mark denotes the category this brand was assigned to. For better readability, we plotted values against mean grammage.

- specific volume (vol/g)
- thickness orthogonal to printed surface (μm)
- grammage $(\mathrm{g/m^2})$
- CIE Y histogram analysis, see Fig. 6
- coefficients after discrete cosinus transformation of the image, see Fig. 7
- distribution of segment size after image segmentation
- waterlevel transformations combined with fractal dimension analysis
- waterlevel transformations combined with binary pattern statistics

The physical dimensions of the paper were included assuming that the Beer-Lambert law holds -at least approximatively- for paper. Physical dimensions are also good features on their own, due to differences in specific weight between fibres and coating (Fig. 5).

4　Observations and Results

The goal of our experimental feasability study was to predict similarity of a new, unprinted paper with existing papers, for which parameters after print
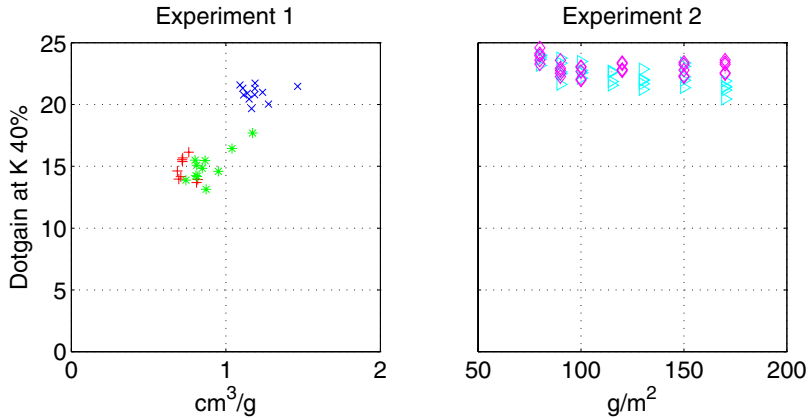
Fig. 5. Relation of dotgain to specific volume and grammage. These plots show data from two printing experiments, on the left with offset print on papers from Fig. 3 as a cross section trough paper grades at equal level of ink provision. On the right side, two paper grades printed with a laser printer to explore dotgain along the grammage axis from 90 to 170 g/m^2.
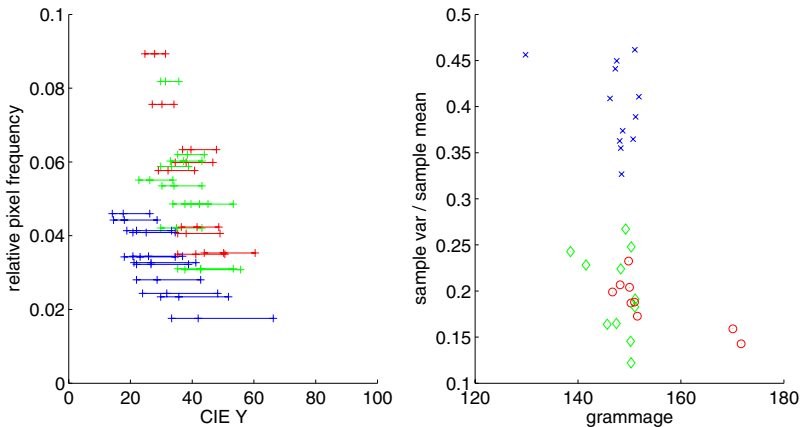

Fig. 6. Histogram analysis. The left figure shows histogram percentiles of 32 transmission images (Fig. 1) transformed from sRGB into a CIE Y grayscale image (horizontally 5%, 50%, 95% percentiles, vertically at the location of peak frequency). On the right, the ratio of sample mean to sample variance against grammage, where the symbol stands for classification according to ISO 12647-2.

were known. The similarity was expressed in the dimensions of paper white and solid black color location as well as dotgain of a 40% single color patch printed with an elliptic dot at 60 lines per cm. The categorisation of transmission
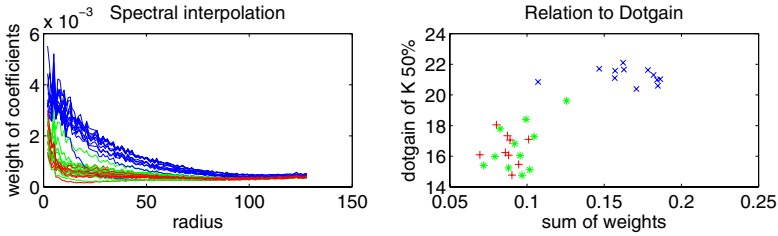
Fig. 7. Spectral interpolation. We transformed 32 transmission images (Fig. 1) to CIE Y grayscale and took 32 random $i_{\max} \times j_{\max} = 256 \times 256$ px samples out of a 1024×1024 px image. We interpolated with discrete cosinus transformation using Matlab `dct2`. The absolute values of coefficients $c \in \{c_{ij} \mid \sqrt{i^2 + j^2} < \frac{1}{2}i_{\max}\}$ were summed as weights. This sum of weights was normalized with c_0 (sample mean) and plotted against dotgain of a grey patch (right).

images by human experts coincides with ISO 12647 in classifying papers with regard to coating. A finer differentiation with regard to dotgain seems achievable (Fig. 4). Classification by grammage, specific volume, histogram and spectral interpolation alone allowed a classification with regard to dotgain of a single color wedge within 4% tone value (Fig. 5, 6, 7).

The b* value of paper white does not correlate with coating, dotgain or solid black color location. The initial differences in b* as shown in Fig. 3 do not completely vanish along Cyan and Magenta ramps. Most uncoated paper grades we tested do not comply with the ISO norm in b* value, but show coherent results with regard to dotgain and density.

Statistical analysis showed correlations > 0.8 of the presented features with dotgain and density, with the exception of grammage, since we arbitrarily included three paper grades with higher respectively lower grammage.

5   Conclusions

We presented an attempt to predict similarity of paper with regard to dotgain and solid tone characteristics by analysis of images and data acquired with non-destructive methods. These data are measurable prior to print with equipment used in the printing industry. Altough methods presented are non-destructive, they do require prints on a training set of paper in order to find a prediction for a new paper. Fortunately enough, training sets is what printing companies produce when controlling quality of their print jobs.

While classification with regard to dotgain of primary color wedges is achievable with the methods presented, prediction of similarity with regard to solid ink color should yet be improved. This will be subject of further study.

References

1. Hunt, R.: The Reproduction of Colour, 5th edn., pp. 684–689. Fountain Press, Kingston-upon-Thames (1995)
2. Downey, A., Hengemihle, F.: The TAPPI Standard Paper Materials Collection of the Library of Congress, Library of Congress Preservation and Conservation Microscopy Laboratory (2006)
3. Neugebauer, H.: Die theoretischen Grundlagen des Mehrfarbendruckes. Z. Wiss. Photogr. 36, 73–89 (1937)
4. Kubelka, P., Munk, F.: Ein Beitrag zur Optik der Farbanstriche. In: Zeitschrift für technische Physik, Leipzig, pp. 593–601 (1931)
5. Chandrasekhar, S.: Radiative Transfer. Dover Publications Inc., New York (1960)
6. Yang, L., Miklavcic, S.: Revised Kubelka Munk theory III. A general theory of light propagation in scattering and absorptive media. JOSA 22(9), 1866–1872 (2005)
7. Jenny, P., Mourad, S., Stamm, T., Vöge, M., Simon, K.: Computing Light Statistics in Heterogeneous Media Based on a Mass Weighted Probability Density Function (PDF) method. JOSA 24(8), 2206–2219 (2007)
8. Mourad, S.: Improved Calibration of Optical Characteristics of paper by an Adapted Paper-MTF Model. Journal of Imaging Science and Technology 51, 283–291 (2007)
9. Možina, M., Černic, M., Demšar, A.: Non-destructive methods for chemical, optical, colorimetric and typographic characterisation of a reprint. Journal of Cultural Heritage 8, 339–349 (2007)