A Variable Data Digital Offset Workflow[1]

Dhiraj Kacker and Russell Muzzolini
Imaging Science
Shutterfly, Inc.
2800 Bridge Parkway, Suite 101
Redwood City, CA-94065, USA

Abstract

With the increasing availability of high quality digital offset presses, one of the key challenges faced by the industry is to develop new profitable markets for variable data content. In this paper we will provide an overview of Shutterfly's variable data products manufactured on the HP-Indigo line of digital offset presses. Shutterfly produces Greeting Cards, Note Cards, and Calendars on these presses. The end users provide text and digital images to customize these products. A key aspect of this purely variable-data workflow is that the customization is "consumer centric" with minimal job variability on the press. The presses are part of a fully automated printing workflow with no job specific previews, setup, or calibration. The presses are calibrated to be colorimetric devices and color is communicated in standardized color spaces to ensure an accurate color reproduction workflow. Data management is designed to operate the presses in a continuous printing mode; thus maximizing throughput.

1. Introduction

Over the last decade there has been significant penetration of digital offset presses. The presses come in many different price-performance and quality levels and therefore address wide ranging needs of various printing markets. As with any new technology, digital presses have to be evaluated based on their impact on revenues and cost. In digital offset printing, the new revenue opportunities are primarily driven by two factors: The ability to do short-runjobs at little or no incremental cost, and the ability to provide variable data content. The cost implications of digital offset presses are mainly driven by the prospect of digitizing most of the printing workflow and thus making it feasible to automate a lot of the processes.

Most traditional offset companies are presently using digital presses to add capability for doing short-runjobs. However, there is still a lot of experimentation around exploiting the variable data capabilities of the digital offset presses. Some of the more popular opportunities being explored involve mass mailings customized to each individual, and creating personalized products with digital images. In this paper we will describe Shutterfly's use of HP-Indigo digital offset presses to provide customized products to consumers from their digital images. The main focus of this paper will be our use of digital printing to create an automated, color-managed workflow. For very good discussions on color management please see [1, 2, 3].

Shutterfly is an online digital printing company providing consumers with a variety of printed products from their digital images. A typical customer uploads their digital pictures to an online album and orders Silver Halide and Digital Offset based printed products. At present we offer three different products using digital offset presses: Greeting Cards, Calendars, and Note Cards. The Greeting Cards are 5"×7" folded cards where both the front image and the inside greetings are customized by each individual consumer; Calendars are 12 or 18 month 8.5"×11" spiral bound pages where customers personalize each month with their pictures; and Note Cards are 3.5"×5" folded cards where only the front image is customized. In addition, for large volume customers, we offer the ability to personalize the logo printed at the back of these products by replacing the Shutterfly logo with one they provide.

2. Production Workflow

As mentioned earlier, we offer three products printed on the HP-Indigo digital offset presses. Each of these products have their own unique ordering patterns and variable data content and present very different production challenges. From a production efficiency point of view, it is desirable to have the least scalable component of the production pipeline be the bottleneck. Due to the cost of a single printer, the physical speed of a press is the least scalable resource. Therefore the entire data and production workflow needs to be designed to keep the presses printing at their rated speed. This is particularly accentuated by the end-of-year Holiday demand for Greeting Cards, Calendars, and Note Cards. A HP-Indigo 3000 press is rated to print 8000 impression per hour on a 12"×18" sheet and has a native

[1] Aspects of the technologies described in this paper are covered in US patent 6,583,852 and other pending US patent applications.

print resolution of 812 dpi. The final screening operation is performed in hardware on the press just before printing so the press needs to be sent continuous-tone(8 bit/channel) data. To save time on any on-press color conversion from RGB to CMYK, resampling operations, or file format conversions, we supply the press with 32 bit/pixel CMYK data rendered at 812 dpi in HP-Indigo's proprietary compressed format called Indigo Compressed Format (ICF).

2.1. Data Generation and Throughput

After ICF compression a typical 5"×7" Greeting Card needs about 3-5 MB of image and text data. Each 12"×18" sheet can accommodate two 5"×7" Greeting Cards resulting in about 6-10 MB or 48-80 Mbits per sheet. With 4/4 (four impressions per side, duplex) printing at 8000 impressions per hour and 60% duty cycle this translates to an effective throughput of 0.16 sheets per second. Therefore, in order to keep the press running continuously, we have to create and feed image data at about the rate of 8-13 Mbps. A typical customer orders multiple copies of each Greeting Card enabling us to make extensive use of caching by only rendering unique image and text files. This dramatically reduces the required data generation throughput and makes it feasible to easily create data at a rate faster than a press can print. Note Cards have even smaller data throughput requirements: there is no customized text and we require customers to order them in sets of 12. We therefore have no problem keeping up with the press when printing Note Cards.

Printing Calendars is quite a bit more challenging. On every 12"×18" sheet we print two calendar pages; each page has two images on it: the customer's chosen personal picture and a month image. The coverage on a typical calendar sheet is much larger than that of a Greeting Card resulting in about 15-20 MB of data per sheet. Assuming the same press throughput as above, this translates to a data rate of 20-25Mbps. It is important to note that this data rate needs to be the rate of slowest piece in a pipelined architecture that includes off-press pixel computation, data transfer, and on-press job creation.

2.2. Architectural Choices

In this section we will describe the architectural choices made to achieve the aforementioned data rates. Before getting into the details it is important to know that there are only some pieces of the pipeline that we can control. In particular, we control: generation of images and text files, networking, and storage and retrieval of images. We don't however control the computational and input/output resources of the manufacturer supplied computer that controls the press and have to be cognizant of the limitations of this piece in the pipeline. This on-board computer is typically used to create the final printable job from ICF's of individual image and text files. The networking and hardware are easier choices once the desired physical throughput and cost parameters are decided. We will therefore focus our attention just on image data generation.

The most critical architectural choice we made is to break the image processing into non-realtime and realtime processing. The block diagram of this architecture is shown in Fig. 1 (for a detailed overview of the printing architecture that includes Silver Halide processing, please see [4]). Customers typically add borders and text to their images; in addition we have to resample images to the final printing resolution and do automated color enhancements to make the pictures look pleasing. The corresponding compositing, text rendering, resampling, and image enhancement operations are computationally very expensive and are best performed in the non-realtimestep. In addition, this processing is done in a device-independent space with the results stored in some standard color space [5]. The realtime image processing is done just before printing when the order is scheduled to a given printer. There are three main operations performed in this step: calibrating images to the color reproduction state of the printer, converting the image to the native color space of the output device, and writing the image to the final file format of the output device. Calibrating images to the color reproduction state of the press is a topic in itself and will be addressed in the next section. In the case of the HP-Indigo digital presses, we create 32 bit/pixel CMYK image data from the device independent color space and convert the image to ICF and write it in the HP-Indigo's JLayout file format. Conversion to CMYK/ICF helps optimally utilize resources on the on-board press computer as this is the native format of the press and needs the least amount of computation.

Non-realtimeand realtime processing are both performed on a bank of low-cost image rendering servers functioning in parallel. Generating image data at a speed faster than the press can print entails choosing the right number of image servers working in parallel such that the effective data throughput of the whole system is faster than the press can print.

2.3. Color Reproduction

In addition to the challenges of producing image data at the native printing speed of the offset press, there are also color reproduction challenges driven by the nature of our product offering and a fully-automated fulfillment pipeline. A customer will typically order a single customized card or Calendar from their digital image(s) to get a proof. If they

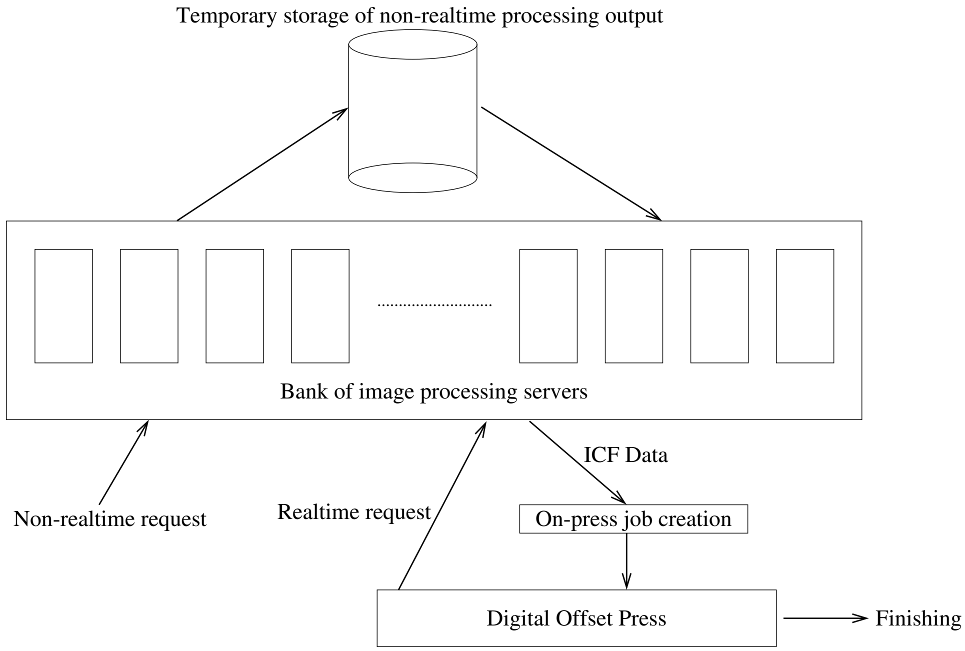

Figure 1: Image Generation Architecture

like the result, at some time in the future they will order multiple copies of the same piece. At no point during the printing do we have hardcopy proofs of what the customer likes but the customer still has an expectation of receiving identical prints between multiple orders of the same image. In addition, we have multiple presses as part of the fully automated workflow. There is, therefore, no prior knowledge of where a particular customers' order might be printed. In addition, a single order can be broken up into multiple pieces and then printed on different presses at different times. This requires that all presses produce almost identical colors at all times. Thus ensuring that customers always get the same print from the same original, and that we don't have to make any complicated scheduling decision to ensure color reproduction quality. In order to achieve this, we have our own proprietary color management workflow called ColorSure^{TM} which is a combination of both proprietary algorithms and production processes. The most important aspect of ColorSure is that all algorithms and processes are designed to keep the presses running continuously by constantly adapting to the changing color reproduction behavior of the press. In order to achieve this, color monitoring patches are automatically scheduled at fixed time intervals and scanned on a spectrophotometer for measuring the accuracy of the color reproduction of that press. In addition, special color calibration targets are also scheduled at fixed intervals and continuously scanned. This data is then used to calibrate images to the state of the press just before printing. In this manner we can maintain all presses as accurate color reproduction devices with inter-press and daily color variation limited to within 2.0-3.0 ΔE of pre-set target values.

3. Conclusion

In this paper we have presented the data workflow around Shutterfly's use of HP-Indigo digital offset presses. There are five key aspects of this workflow that make automation and color management possible:

1. Separating the image processing into non-realtime and realtime components.

2. Creating only unique image/text data and making extensive use of caching.

3. Using a bank of low-cost image processing servers scalable with changing production needs.

4. Writing image data in the native image format of the presses.

5. Continuously calibrating the presses with algorithms and processes that do not require any downtime.

All these factors combine to create a workflow where we can keep our presses printing continuously without compromising accuracy of color reproduction. We are thus able to maximize the throughput of our machines and provide digital offset products cost-effectively.

From our point of view, digital offset presses present new revenue opportunities through the use of variable data capabilities. They also open up the prospect of automating many production processes in traditional offset printing. Accurate communication of color is a key enabling technology that makes it feasible to operate the presses in this mode. Taking a systems perspective and designing the entire workflow such that all pieces of the pipeline scale around the printing speed of the press ensures getting the most productivity from these presses.

References

[1] Jack Holm, "Integrating New Color Image Processing Techniques with Color Management," in *The Seventh Color Imaging Conference: Color Science, Systems, and Applications*, 1999, pp. 80-86.

[2] Ann Lerene McCarthy, "ColorFidelity Across Open Distributed Systems," in *CGIV 2002: The First European Conference on Colour Graphics, Imaging, and Vision*, 2002, pp. 639-644.

[3] Edward J. Giorgianni and Thomas E. Madden, *Digital Color Management: Encoding Solutions*, Addison Wesley Longman, Inc., 1998.

[4] Dhiraj Kacker and Russell Muzzolini, "Shutterfly's Printing Infrastructure," in *IS&T's NIP18: International Conference on Digital Printing Technologies, San Diego, CA*, 2002, pp. 599-602.

[5] Sabine Süsstrunk, Robert Buckley, and Steve Swen, "StandardRGB Color Spaces," in *The Seventh Color Imaging Conference: Color Science, Systems, and Applications*, 1999, pp. 127-134.

Biography

Dhiraj Kacker received his B.Tech. from the Indian Institute of Technology, Bombay in 1993, MS from Washington State University in 1996 and Ph.D. from Purdue University in 2000, all in Electrical Engineering. Since 2000, he has been working as an Imaging Researcher with the Image Science group of Shutterfly, Inc. His present research interests include hardcopy image reproduction, image quality perception, gamut mapping, watermarking, and statistical signal processing. Dhiraj is a member of IEEE and IS&T.

Russell Muzzolini received his M.Sc. in 1991 and Ph.D. in 1996 from the University of Saskatchewan, both in Computer Science. In 1996 he joined AliasÑW avefront to work on an advanced 2D compositing and 3D graphics applications for film and video visual effects in the Entertainment Industry. Since 1999 he has been with Shutterfly working as an Imaging Researcher with the Image Science group. Russell's current research interests include statistical signal processing, hardcopy image reproduction, image segmentation and motion tracking.