The Challenge of Color for Production Scanning Operations

White Paper
June 2006

Perfect Scans the First Time, Every Time

KOFAX®
A DICOM GROUP COMPANY

Table of Contents

Introduction

Remember black-and-white TV? Not many people do. Color TV presented such a compelling alternative that viewers migrated en masse to what they viewed, literally and figuratively, as the superior medium.

A similar migration is getting underway in the document imaging market. People see the world in color, and expect the documents they work with to reflect that reality. And, increasingly, they do. The democratization of color photography, driven by ubiquitous digital cameras and camera-equipped cell phones, enables al most anyone to incorporate color photographs into documents easily and inexpensively. And lower cost color printers and scanners encourage workers to use color photos and graphics in a wider variety of business documents.

Just as you would have a difficult time today convincing the average viewer of the merits of black-and-white TV, so too is it a problem to persuade the average knowledge worker that the bitonal representation of a photograph or graphic on their screen or print-out is an acceptable alternative to the color original. The problem will grow increasingly acute as document images move beyond the bounds of an organization's internal network and onto the wider realm of the Internet. There, illegible bitonal representations are likely to receive an even harsher critical response.

Consider, for example, auto insurance claim forms or home appraisal reports. Color photographs are an integral part of each. While words can detail the characteristics of a damaged car or home for sale, it is far easier to convey this information with a photograph. And, in each case, a color photograph tells much more than one in black-and-white.

While most would agree that the use of color in documents is moving from a luxury to an expectation, and that color needs to be preserved in the digital representation of the document, the document imaging industry has yet to reach consensus on how best to do so. Many production scanning operations still are not prepared to capture color images contained in documents in an efficient manner. Instead, an exception process interrupts the scanning workflow or, more typically, the color photograph simply is captured as a bitonal image, leaving future viewers only to guess at what additional level of detail color might have provided.

What options are available if an organization elects to capture and preserve the integral color elements of every document it scans? In this paper we will review three alternatives and examine the viability of each.

Scan and capture 100% of images in color

In this approach, all documents are processed as color documents, no matter how much—if any—color they contain. At first glance, with the declining price of color-capable production scanners, this option seems viable. However, a look at expenses beyond the scanner alone reveals just how difficult it is to construct a supporting economic case for this approach.

For example, let's look at storage and retrieval costs. Based on a rate of .001 cent per KB, a typical 20KB bitonal image would cost $0.02 to store/retrieve, while a typical 512KB color image would cost $0.51. At an added cost of 49 cents per image, only the wealthiest organizations could afford to make a blanket conversion to color.

Additionally, the speed of the scanning operation and the overall efficiency of the capture process would suffer. Most scanners run slower in color mode, so the number of documents an organization could process each day would decline. The only way to stay at previous throughput levels would be to purchase more color-capable scanners, and hire more workers to operate them.

Finally, in ad hoc or distributed scan operations, a 100% switch to color could have a negative impact on all network users. The introduction of all-color images, 25X the size of bitonal images, could create a network traffic jam that the IT department may not have the capacity or budget to alleviate.

The higher cost of storage, the loss in productivity and the detrimental impact on efficiency make it difficult to justify a 100% conversion to color scanning, particularly when many, if not most, of the documents do not include color elements. VARs or service bureaus presenting such an alternative to potential clients are likely to lose to competitors who offer less expensive alternatives.

Pre-sort documents before scanning

This solution requires a scan operator or document preparer to review each document and separate those which require color processing from those which need only bitonal processing. Depending on the type of scanner in use, this may require a manual sort into two piles, or the insertion of document separator sheets before and after color images.

In either case, both the scan application and the scanner must be capable of accommodating a mixed process. The application must allow color images to be processed with the rest of the documents and batch, and provide a method for switching between image file types. The application also must control the switching of scanner modes and outputs, and the scanner must be capable of understanding and reacting to these commands.

In a test of three scan applications, the switch between color and bitonal modes required multiple steps, and delayed the scanning process by an average of 30 seconds. Consider the economic fallout such a delay would have on labor costs alone in a batch or production scanning environment. If this operation scans 5000 pages per day (250 batches), 5 days per week, 52 weeks a year, with the scan operator receiving $15 per hour in salary and benefits, and a conservative estimate of two color images per batch, the additional cost would total $16,263 per year.

While a financial hurdle of this magnitude is difficult to overcome, equally challenging is the assignment of this level of responsibility to knowledge workers in an ad hoc or distributed capture environment. It may be unrealistic to expect all but the most experienced workers to switch scanner settings or modes correctly, and to accurately sort or segment those documents which need color scanning.

A VAR or service bureau recommending this type of solution may not properly assess its complete costs and training requirements, creating a potentially unprofitable partnership with its customer.

Automatic color image processing

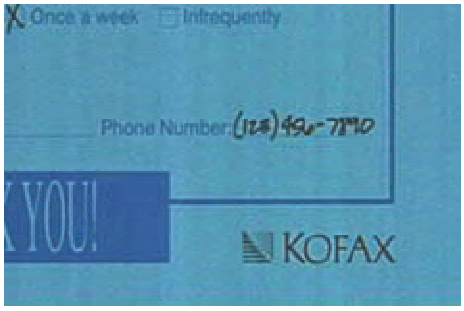
This approach utilizes hardware and/or software that can intelligently detect color within an image and determine its relevance. If it judges color as integral to the understanding of the document, it can automatically switch the scanner to the correct mode, and capture the document using the correct file type. This allows the scanner to operate at optimal speed, with no operator intervention or support.

To save on network traffic and storage requirements, the software also should be capable of optimizing the image file size. Ideally, the software will work with most scanners, and require only a limited implementation or integration effort.

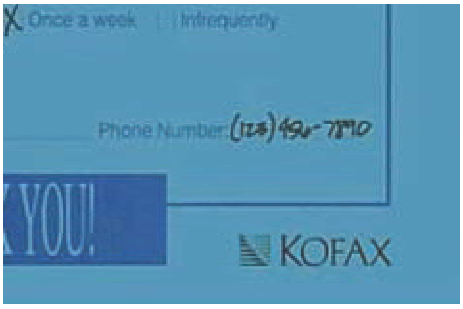
Determining the relevance of color in a document is key. Capturing in color every document which includes some color element is likely to lead to an abundance of documents which could have been processed more quickly and cheaply as bitonal images. For example, a page on which a company logo is the only color element most likely would not need to retain the color in its digital representation in order to preserve a later viewer's understanding of the document. Electing to capture in color every document which includes any color element would not be as costly as the 100% color option discussed earlier, but it also would not provide the full economic benefits more capable automatic color image processing solutions could deliver.

At the same time, just because a document includes only a small color element should not automatically relegate it to the bitonal processing queue. The color may notify the viewer of an important addition to the original document—a date stamp, say, or a highlighted word—and its retention could be critical for future viewers to fully understand the document. Ideally, an automatic color imaging solution should give users the flexibility to set the size or amount of color a document needs to contain to trigger its handling as a color document.

Additionally, when multiple colors appear within a document, the software should be capable of determining which color (or colors) is there simply as a background color. The software could smooth the color automatically to a tone which makes the content stand out more clearly, or drop the color completely. Either approach improves legibility and creates a smaller file size.



Non Color Saturation Image (265Kbytes)



VRS Color Saturated Image (171Kbytes

Finally, when a production scanning operation utilizes multiple input devices built by multiple manufacturers, a single image processing software solution can act as an important unifier. Scanner operators need to master just one program, and through a common interface can control all the capture devices on site. IT departments like that they have only a single program to install, administer and support.

Conclusion

In the document imaging market, even though the use of color today is still in its nascent stage, an all-color future is unmistakably clear. Like the TV market before it, at some point black-and-white only document imaging will be viewed as an anachronism, a mere stepping stone on the path from crude bitonal representations of reality to true color depictions of the world in which we live.

But we're still years away from when the cost differential between color and bitonal scanning will shrink to a negligible amount. Until then, the data imaging industry must determine how best to handle the minority of documents which incorporate color. As we've seen, an all-color solution today is not economical, and manual intervention is costly and inefficient. This leaves automatic color image processing as the best bridge between a bitonal past and all-color future, one capable of delivering real benefits today while laying the groundwork for what's to come.

VirtualReScan (VRS): Kofax's automatic color image processing solution

With VRS, anyone who can load a document feeder and press the Start button can achieve professional scanning results. VRS includes numerous features to correct and adjust images, so that a document's digital representation will be as legible as possible both to the human and mechanical eye. VRS is available as a hardware-assisted technology for high volume scanning operations, and as a software-assisted technology for mid and low volume centers, as well as distributed scan sites.

VRS automates professional-level scanning by removing human intervention from the process as much as possible. Its Automatic Color Detection feature eliminates the need for anyone to pre-sort documents, or insert separator pages. Instead, if VRS detects no color on the page, it knows it is safe to process as a bitonal image. If VRS detects color, it relies on a sensitivity meter users can adjust to determine whether color or bitonal processing is preferred. Users can test documents similar to those they plan to scan and adjust VRS's color sensitivity to achieve the processing outcome they desire.

VRS also includes a clever feature, Detect Small Color Area, that helps guard against missing small, but crucial, color elements in an image. This is particularly useful if the user has set Automatic Color Detection sensitivity meter to a low value, meaning a document must include a significant amount of color to be processed in color. At this setting, VRS may ignore a color time stamp, an "ACCEPT" or "REJECT" stamp, a few words highlighted with a color marker, or some other small color element that holds the key to a later viewer fully understanding the document. But with Detect Small Color Area enabled, VRS will identify and preserve these color elements, even if they are the only color elements on the page, by processing the document in color. As with Automatic Color Detection, the Detect Small Color Area feature has a user adjustable sensitivity meter.

With Detect Small Color Area enabled

To reduce file size and the number of color "false positives," VRS also has the ability to identify a document's background colors and either suppress them entirely, or convert them to an "averaged" color. So, for example, if the only color in a document is the paper stock it is printed on, VRS can convert this to white, and process the document as a bitonal image. Or, if the document has a busy color background, VRS can average it to a single color, making text easier to read. In either case, the resultant file size is smaller, making it faster to transmit and cheaper to store.

The latest release of VRS also includes a Scanner Configuration Utility, making it easy to set up and use most commercial scanners. Users select their scanner from a list of compatible scanners, and VRS automatically applies the correct default settings. If a site employs multiple scanners, the common interface VRS provides makes it easy to switch among them, even if the scanners are from different manufacturers.

To learn more about VRS, visit: http://www.kofax.com/vrs.

16245 Laguna Canyon Road
Irvine, CA 92618-3603, U.S.A.
PHONE 949-727-1733 FAX 949-727-3144 www.kofax.com

Koning Albert 1 Laan 64-66
B-1780 Wemmel, Belgium
PHONE +32-2-456-1720 FAX +32-2-456-1721