A Human Readable Hardcopy Backup System for Digital Images

Eric Hoarau, Ingeborg Tastl, Nathan Moroney; HP Laboratories; Palo Alto, CA, USA

Abstract

Large amounts of images are captured every day and are stored either on digital or analog storage solutions. Each has advantages and disadvantages when used to preserve one memories. This paper proposes to augment the existing solutions with a modernized, inexpensive hardcopy backup, which can serve as a fall-back solution if everything else fails. In particular, the collection of images is printed in form of an inexpensive photo-album, which is augmented with printed human readable meta-information designed to enable an easy reconstruction of the original images in the future and can be enjoyed meanwhile. The methodology is illustrated using a set of original digital sRGB images and printed simulations of faded versions of those images, which are scanned and reconstructed in an automatic and manual way and are compared to the results of a scanner reconstruction software and to the original images.

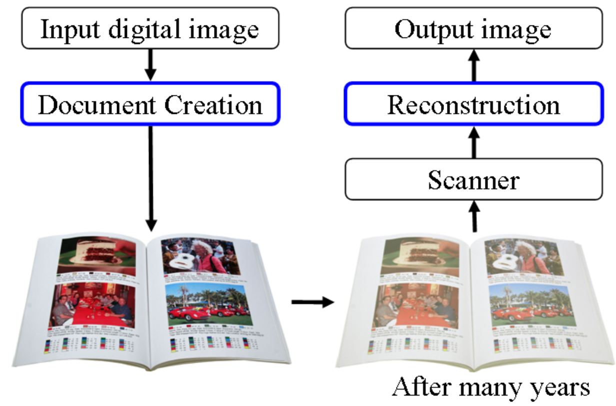

***Figure 1**. Reconstruction Framework*

Introduction

Most users tend to store their images using either digital storage solutions such as hard drives, CDs and DVDs, and online storage services or analog solutions such as single prints or photo albums. The International Imaging Industry Association (I3A) has created a site which details the options and limitations of current technologies [1]. While digital back-ups might offer an exact copy of the original data, they are vulnerable to hardware failure, system obsolescence, or simply user errors. Printed images on the other hand can be viewed and enjoyed even in a damaged state for decades to come, but they are a reduced representation of the original data.

Extensive research has been performed in restoring faded Silver Halide images and old movies [2][3][4]. It is performed professionally by either automatically inverting specific fading models based on estimated parameters, or by skilled operators in a tedious manual fashion. Although there are many publications discussing the issue of color appearance changes of digital prints over time due to illumination, temperature, ozone and pollutants [5], barely anyone talks about restoring those prints in a systematic way that goes beyond a general image enhancement approach. Even fewer modern publications look into anticipating an easy future image reconstruction at the point of printing itself (although that concept might have been used in the past). One exception is a paper by R. Samadani and D. Mukherjee [6]. Their solution proposes to encode information in printed media as side information. The results are promising; however the very nature of encoding the auxiliary information digitally in form of dense barcodes could make them unreadable in 10, 20, or 30 years if not regularly updated to the latest standard encoder/decoder.

We are not aware of any solution that focuses on augmenting the printed image with visual features to track appearance changes and facilitate a reconstruction after a long period of time. In this paper, we describe a robust hardcopy backup and reconstruction system for digital images. Our solution is not based on specialized ink and paper as it is resilient to some degree of image fading. We introduce an automated document creation layout algorithm, manual and automated reconstruction methods and present our experimental setup and results.

Method

We are proposing to establish a simple yet robust hardcopy backup framework of a digital archive. A user can automatically store their favorite images on a printed page or album that also contains human readable metadata and color tracking features. The album can be enjoyed for years to come and in the event of loss of the original digital file used to reconstruct an approximate digital file as shown in Figure 1.

Printing the auxiliary information in a human readable format ensure that it can be used by current and future reconstruction methods. The embedded color patches track the modifications that are happening to the images (print-scan channel perceptual color transformations, appearance changes due to light, temperature, humidity and pollutants). Reconstructing the original images simply means inverting those changes by using the original numeric color values printed next to the patches. This reconstruction approach is different than the ones commonly used for Silver Halide prints where the inversion models are built for the average fading of specific media. The accuracy of these solutions depends on how close to the model the actual fading is and how well the parameters can be estimated. Treating changes as a black box makes our method suitable for any printing technology and it is independent on whether the actual changes fit a particular model or not. Thus it can be considered as more flexible and more robust, which makes it well suitable in real world applications.

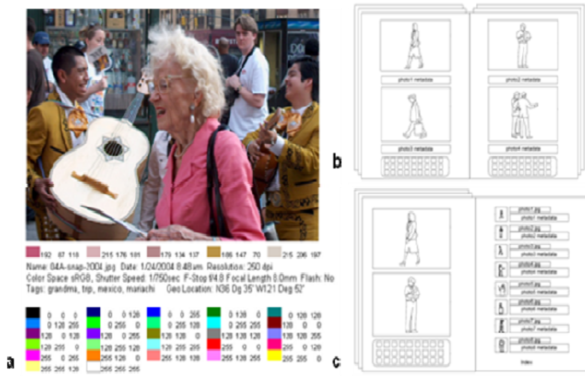

***Figure 2**. a) image with meta-information, b,c) sample layouts*

The method, which will be described in more detail in the rest of the paper, can either be used to reconstruct the original digital information (e.g. sRGB values of the images), or alternatively it could be used to reconstruct the appearance of the printed image immediately after printing.

Document Creation

In order to embed tracking features and key metadata information alongside the images, we developed an algorithm to automatically layout the images from an archive into a PDF formatted document. That digital file can be printed and collected into finishing formats such as folders, binders, and bound documents. The software was developed in C++ coding language and utilizes various imaging libraries. The software GUI also enables the user to optionally modify or add to the metadata included in the final document.

Two possible layouts are shown in Figure 2b. In the first layout, the images are printed using a 4x6" format with white space around them for both image protection and easy scanning. Such an image size can generate a ~2MP scanned digital image. Larger images could be printed if desired, but our first version use standard paper and form factors to maximize the number of images and minimize production cost. There is a wealth of information in the EXIF file of an image that is generally discarded during the printing stage. In our augmented layout shown in Figure 2a, the image metadata information is printed below the image using a standard font such that it can easily be interpreted by OCR algorithms. It includes image-specific information such as file name, date, tags and GPS location, or camera information such as shutter speed, f-number, focal length, flash, and color space. Such information might not seem valuable at this time, but in the future, might prove to be invaluable for automatic image recognition and understanding.

We are using a color target of 27 color patches evenly sampling the sRGB space as our main tracking tool (placing 3 samples along the Red, Green and Blue axis in a cube results in 3^3=27 color patches). The corresponding sRGB values have to be printed for reference and are used as goal values for the reconstruction process.

An optional swatch of five image dependent color patches can also be added below the image to provide even more granularity in the reconstruction transform and to help in the manual reconstruction steps. They can be especially useful to accurately reproducing skin tones and memory colors. Several solutions have been proposed to extract image specific colors [7]. We are using a similar procedure, however basing the weighting on memory colors and skin tones. Moreover, we are segmenting the image using our research on lexical image processing to ensure fast processing time when handling numerous images [8].

We have also developed an alternate layout where some of the image metadata is included in an index section at the end of the book to keep the main pages clean as shown in Figure 2c. This also provides an opportunity to include additional information into that section supporting the preferred method of reconstruction and a larger color target of for example 125 patches. A larger color target can result in more accurate reconstruction as show in the result section of this paper.

Reconstruction of Scanned Images

In the event that a digital reconstruction of the hardcopy images is needed, the user can either scan the prints and manually adjust the images with the goal of bringing the color patches back to their reference values, or preferably he can scan the prints and automatically reconstruct the images.

Automatic Reconstruction

Any algorithm that can define and apply a color transformation using the scanner values of the faded color patches and their original recorded values can be implemented to reconstruct the images. One possible implementation would be a 3D scattered data interpolation [9].

We were inspired by the method used to generate ICC printer profiles [10]. We started with 27 color patches for which we had the original sRGB values and the corresponding Lab values obtained from a modified HP scanner pipeline. The 27 points of the sRGB color space were tessellated (a structure of tetrahedrons was built) and then the same structure was used in Lab space. Building the ICC reconstruction profile simply meant that for each point of a regular grid in Lab color space an enclosing tetrahedron had to be found and a tetrahedral interpolation had to be performed [11].

Color values that were outside of the structure were first mapped to the gamut boundary and then interpolated. Applying an ICC reconstruction profile to an image simply means that the color of every pixel is interpolated using the pre-calculated table of the ICC profile. As this is a regular grid, the transformations can be done very fast.

A printed, faded and scanned image can simply be reconstructed by applying an ICC reconstruction profile and then assigning a normal sRGB profile to the image. This process can be performed in Photoshop, inside a Matlab or C program or within a scanner software.

Manual Reconstruction

Reconstructing the image manually is not easily accomplished due to the non linearity of the fading process, and because current off-the-shelf imaging software packages provide only basic color adjustment. We devised the manual reconstruction into a two steps approach: 1) A global reconstruction method consisting of adjusting the black and white point using the corresponding patches, and 2) a more advanced reconstruction method where the user can locally color correct the image by segmenting the image and using the 5 custom color patches as reference.

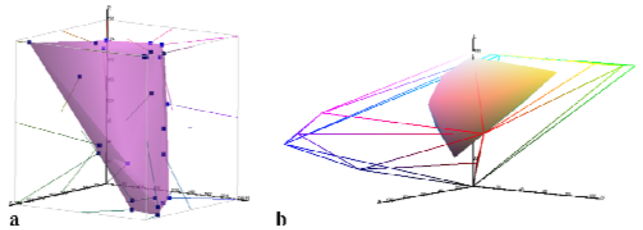

***Figure 3**. a) 27 patches from sRGB to (scanner)RGB (blue squares) in sRGB space, b) Gamuts formed by the 27 original (wire frame) and scanned patches in Lab space.*

Testing

In order to test the feasibility of our approach, we assembled a set of test images, augmented them with 27 (125) regular sampled sRGB color patches, simulated the effects of using 3rd party replacement inks for HP cartridges and 30 years of light exposure, printed the simulated faded image, scanned the image using an HP scanner, retrieved a scanned image (in both scannerRGB and Lab color space) and restored the image first by using the "restore faded image" functionality of the scanner itself, second by using our new automatic method and third by using a manual method in Photoshop. In the following, we will discuss the individual steps.

To simulate the appearance of an image after a certain number of years of light exposure, we applied ICC fade simulation profiles to selected images [10]. They are built on measurements acquired from test charts, whose fading process was accelerated in a fadometer. Tests showed a quite good correlation between images faded in a fadometer and printed images simulating fading over the same time frame.

We printed and scanned the images with the augmented test patches using an HP scanner. The color changes include the perceptual color transformation of the printer, the fading of the print, and finally the re-rendering in the scanner. The effect of that overall transformation is visualized in Figure 3a for 27 regular spaced color patches in sRGB space. The vectors ending in blue squares visualize the changes that happen to the 27 original sRGB colors. The overall gamut shrinks from the RGB cube (gray wire frame) to the transparent magenta gamut in the middle. The strongest changes occur in the dark Red, pure Red, dark Magenta and pure Green. Figure 3b shows the relationship of the 27 original sRGB patches (mesh) in relationship with the (scanner)Lab values in Lab color space. Any attempted image reconstruction has to invert exactly those changes. The reconstruction is non-linear and it has to perform a strong gamut expansion of scanned data, which is commonly noisy. All of that creates an additional challenge for the color transformation implementing the reconstruction.

Results and Discussion

The results for each solution are shown in Figure 4. We selected that image because it contains highly saturated content (robe of the monk), skin tones, neutrals and natural objects. This image also contains colors (reds) that we know change quite strongly for the fading example that we are using. We compared visually the results of the automatic and manual reconstruction against a reconstruction method implemented in an HP scanner and against the original digital image. We opted against a simple

***Figure 4**. "Monk" Image a) Original sRGB image, b) Scanned image, c) Manually restored image, d) Image restored by software from scanner, e) Automated restored image*

pixel-by-pixel numerical comparison of the original and reconstructed image as we think a complex human judgment was more appropriate. Any metric would need to consider that an image of a particular resolution has been halftoned, printed, might have faded quite dramatically over time and has been scanned through which noise has been added and finally restored as good as possible. Any metric would need to consider spatial aspect as well as color differences and could for example be based on S-CIELAB [12].

Method Deployed in HP Scanners

Looking at the results of the algorithm implemented in the HP scanner, we can definitely see an enhancement of the scanned image. However, the image has a bluish cast and is still quite pale in comparison to the original image. We do note that this algorithm does not use any meta-information and has been developed for the restoration of faded Silver Halide prints.

Automatic Method

The automatic reconstruction brings the scanned faded images much closer to the original images, especially compared to the other two reconstruction methods. However, a closer examination reveals opportunities for further improvements (e.g. the floor section has a slight magenta cast.) Those nuances are related to the small set of reference points used to model a non-linear color transformation and because we are using a piecewise linear interpolation method. For a printer profile we typically use 1000 measurement points and then build 17x17x17 tables that are included into an ICC profile. The results are good considering the few color patches used in our layout. However, to achieve higher quality results without any artifacts, we might have to increase the number of samples or include color patches that are image dependent. The larger number of patches could be included at the end of the Photo book with the image dependent ones placed in close proximity to the individual image. Although we are currently using the denoising algorithm of the scanner, the presence of noise could also be considered in the design of the color reconstruction method. It is an interesting topic for future research, which also has applications in other areas.

Manual Method

In our first set of experiments, we focused on developing an easy reconstruction method as it is most likely the first correction step a user would undertake. Our investigation of the automatic reconstruction showed that the different colors in an image fade at different rate. Because of this nonlinearity, it is difficult to achieve the same level of results as our new automated reconstruction based in only a few steps. We are still investigating, but some of our approaches consist of re-setting the image black and white point using the corresponding patches. We also explored using curves to compensate for the flatness of the resulting image. So far, the best results were produced by manually cropping the individual RGB levels.

Conclusion

We presented a complete system for generating a hardcopy backup for digital images augmented with human-readable tracking features, which facilitate a reconstruction of the digital files in the future if necessary. This solution is inexpensive, does not require a specific printing technology and can be enjoyed while fulfilling its purpose as a fall-back solution. The automated reconstruction shows significant improvement over existing solutions. Moreover, since our solution is based on standard color and imaging tools, the user will have the most flexibility in his alternatives for reconstruction. He can use the methods highlighted in this paper or any 3rd party solution available at the time.

References

[1] I3A, International Imaging Industry Association http://www.savemymemories.org/

[2] F.S. Frey, R. Gschwind and L. Rosenthaler, Electronic imaging, a tool for the reconstruction of faded color photographs and motion pictures, IS&T/SID Color Imaging Conference. (1996).

[3] Raymond S. Lee, Innovative Technologies to solve imaging problems, Proc.12th Intern. Symp. Photofinishing Techn.

[4] Yujiro Ando, Aran Handsuesai, and Kanmanus Khantong: "Digital restoration of Faded Color Images by Subjective Method", Jour. Imaging Sci. Technol., Vol. 41, Nr. 3, May/June 1997.

[5] www.wilhelm-research.com

[6] R. Samadani and D. Mukherjee, "PhotoPlus: Auxiliary information for printed images based on distributed source coding", Proc. SPIE, Visual Comm. And Image Processing, San Jose, Jan. 2008.

[7] B. Morse, D Thornton, Q. Xia, J. Uibel: "Image-Based Color Schemes" ICIP 2007.4379355

[8] N. Moroney, P. Obrador, G. Beretta: " Lexical Image Processing" Proceedings of the 16th. IS&T/SID Color Imaging Conference.

[9] I. Amidror, "Scattered data interpolation methods for electronic imaging: a survey", Jour. Elect. Imaging, Vol. 11, pg.157, (2002).

[10] I. Tastl, K. Koh, D. Rossing, D. Berfanger, N. Moroney, "ICC Profile Based Defect Simulation", Proc. IS&T/SID 2006, pp. 230-233

[11] G. Sharma, Digital Color Imaging Handbook, CRC Press, 2003.

[12] X.M. Zhang and B.A. Wandell, "Color image fidelity metrics evaluating using image distortion maps", Signal Processing, Vol. 70, No 3, pp201-214, Nov. 1998.

Author Biography

Eric Hoarau is currently a senior researcher at Hewlett-Packard Laboratories in Palo Alto, CA. He received his B.S. from the University of California at Berkeley (1999) and his M.S. in Mechanical Engineering from the Massachusetts Institute of Technology (2001). His research interests span several fields: Mechatronic systems, imaging systems, color imaging algorithms, distributed computing and system dynamics.

Ingeborg Tastl is a principal scientist at Hewlett-Packard Laboratories in Palo Alto, California. Her research interests are in the areas of color reproduction and color management related to different printing technologies. She represents HP at the International Color Consortium (ICC). Before joining HP she worked at Sony's US Research Lab in San Jose, California. Ingeborg received her Master's degree in Computer Science from the Vienna University of Technology (1990), and a Ph.D. degree in Color Science from the same University (1995).

Nathan Moroney is a principal scientist at Hewlett-Packard Laboratories in Palo Alto, California. Previously, he worked for the Barcelona division of Hewlett-Packard and at the RIT Research Corporation. He has a Masters Degree in Color Science from the Munsell Color Science Laboratory of RIT and a Bachelors degree in color science from the Philadelphia University. His research interests span various color imaging technologies and machine learning techniques.