

Improving Quality of Digital Images of Art in Museum Collections

Alejandro Jaimes

Columbia University, New York, New York

Fred Mintzer

IBM TJ Watson Research Ctr., Yorktown Heights, New York

Abstract

In this paper, we present a study in which we analyze digital image quality corrections performed by an expert operator at the *National Gallery of Art (N.G.A.)* in Washington, D.C., and we propose a framework to semi-automatically improve the quality of digital images in museum collections. The work presented has two goals: (1) to explore ways to facilitate the color image correction process, and (2) to gain a better understanding of it. We analyze the expert's correction process (i.e., operations and workflow), and compare changes in contrast and luminance for *original* and *corrected* images selected from two different collections (Impressionist and Dutch/Flemish paintings). Results of the study suggest that, although corrections depend on each individual image, it is possible to find patterns in the way that similar images are corrected. Therefore, the proposed framework is based on the assumption that images can be placed in categories (images within a category are more visually similar than images across categories), and that correction patterns can be learned and applied semi-automatically (i.e., under the supervision of an expert operator) for different categories.

Introduction

The proliferation of technology to create, store, and share digital images has inspired the creation of large digital image collections. Although image quality is always a concern, the desired level of quality is highly application dependent. In museums, digital photographs of works of art are generated for many purposes (e.g., archive for inventory, publish on the Internet^{7,8}, etc.). Due to the nature of the collections, however, the goal is often to produce digital images that look, as much as possible, like the original art.

In working towards the goal of producing very high quality digital images, the Image Library Applications group at IBM TJ Watson Research Center has developed a high quality digital scanner. The Pro/3000 Digital Imaging System, which has been installed in several cultural institutions⁷ can create digital images with accurate color at high resolution (3,000 pixels by 4,000 pixels).

In spite of scanner's quality, its digital images must be retouched, using a manual correction process, to meet the most demanding institutional quality standards. The *National Gallery of Art* in Washington D.C., has been actively digitizing its permanent collection for a number of years. An expert operator scans photographic transparencies of paintings and corrects the digital images so that when viewed on a computer monitor (for publication on the Internet, www.nga.gov) they look, as much as possible, like the original works of art.

At the *N.G.A.*, and similar institutions, manual image correction involves demanding and time-consuming tasks, and these tasks create bottlenecks in the creation of large image collections. Furthermore, the process is subjective, and requires experience that is difficult to acquire since corrections depend not only on the output format (e.g., screen vs. printer), but also on the collection and on content of each image. Automating the process, or facilitating it in some way, is important in addressing the bottleneck problem because it reduces the time required to prepare each image. It may also provide a greater level of consistency throughout the digital collection.

Full automation, however, is so difficult that it may be unattainable, especially for applications where the desire is to create images that "look like the original art," as much as possible. Such measures are based on human perception, making them highly subjective. Even partially automating the correction process is very difficult because corrections made by an expert may vary significantly depending on the individual image and collection (e.g., Impressionist vs. Dutch paintings).

In this paper, we study the operations performed by an expert operator to improve the quality of digital images of paintings at the *National Gallery of Art* and we propose a framework to facilitate the image correction process (by learning corrections and applying them semi-automatically). Our goal is to gain a better understanding of the manual correction process and to learn how to facilitate the image correction process for museum collections.

Related Work

Although many image quality models² and experiments have been performed in the past, to the best of the authors'

knowledge, this is the first experiment of its type (i.e., expert input, cultural application, etc.). Similarly, the learning framework proposed is different from previous systems that concentrate on a single aspect (e.g., color appearance models¹, color constancy³, color mapping), that are not based on expert input⁵, or that use a standard-observer approach to automatic quality corrections⁴.

Capture and Correction Process

Because of logistical considerations, paintings and other works of art are first photographed (by professionals in specialized studios) using standard photographic equipment and film. Then, the film transparencies are scanned using the IBM Pro scanner, and corrected by an expert operator (for publication on the Internet) so that on a computer screen they look, as much as possible, like the original works of art (0).

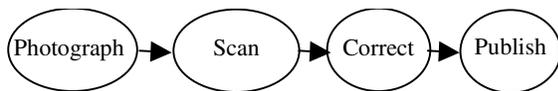


Figure 1. General overview of the digitization and publication process at the N.G.A..

The expert spends perhaps an hour working on each image, making corrections (e.g., adjustments to contrast, color balance, etc.) and viewing the original painting, until he is satisfied with the appearance of the image on a CRT. The correction process can be characterized by a sequence of operations (0) that depends on the *format and purpose of image output* (in this case publication on the Internet), the *image content*, and the *expert's subjective evaluation*. Although objective quality measures are important, each decision is made by the expert based on his experience (in this case several years at the N.G.A.).

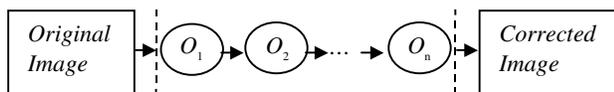


Figure 2. Image correction operations.

A Learning Framework

We propose a framework whose goal is to facilitate the series of steps that are usually carried out by the operator in the correction process. The system we propose consists of two modules, one for *training* and one for *automatic correction* (0). During the training phase, the goal is for the system to learn to make corrections, from input provided by an expert. It is desirable for the system to identify patterns in the correction process, particularly in what we will call *correction factors*: (1) *what* types of corrections are needed (i.e., which operations to use), (2) *where* (i.e., to which images or types of images they should be applied), (3) *when* (i.e., in what order), and (4) *how* they should be applied

(i.e., specific operation parameters). Our hypothesis is that, for a given application (e.g., publishing images on the Internet at the N.G.A.), decisions regarding the *correction factors* are highly dependent on *specific visual content categories* within the image collection (e.g., Impressionist paintings vs. Dutch/Flemish paintings). As a consequence, the first step in the training process consists of separating images into categories, where images within a category are more similar visually than images across categories. Patterns in the correction process, then, are learned independently for each category, and applied to future images depending on the image's category.

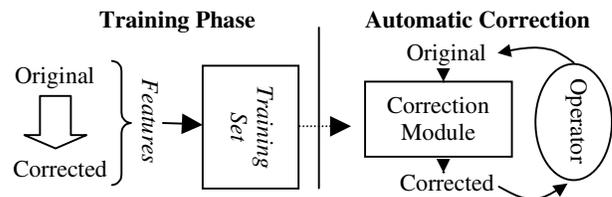


Figure 3. Overall system framework. Depending on the training scenario, different features are extracted during training and used in the automatic correction phase.

Training Phase

For a given category (e.g., Impressionist paintings), there can be three different scenarios corresponding to the training phase, depending on the amount of information available. In *scenario I*, the individual correction operations (functionality and parameters) performed by the operator are known to the learning system. The input to the learning module, then, for each corrected image, consists of a set of features for the *original* image (e.g., histogram), and an ordered list of operation-parameter pairs (e.g., $S = \{ \{O_p, P_p\}, \{O_2, P_2\}, \dots, \{O_n, P_n\} \}$, where O_i is an operation and P_i corresponds to operation O_i 's parameters}).

In *scenario II*, the system will have access to the *original* and *corrected* images, and the system's designer will have *general* knowledge regarding the operations that were used in the correction process. Based on this general knowledge, the system is designed to extract features that reflect differences (caused by specific corrections) between the *original* and *corrected* images. For example, if it is known that contrast is usually corrected by the operator, a set of features, $F = \{f_1, f_2, \dots, f_n\}$, that specifically measure change in contrast can be extracted. The input to the training module, then, for each image, will consist of a set of features for the *original* image alone (e.g., histogram), and a set of features extracted from the differences between the *original* and *corrected* images. This scenario, described in more detail in the Data Analysis section, corresponds to the setup of our study at the N.G.A..

In *scenario III*, there is no information about the operations. Like in *II*, the system must analyze the differences between *original* and *corrected* images, and extract features to be used in a training set. The features extracted, however, will be not be based on any information about the possible operations being performed.

Learning and Automatic Correction

One possibility for the learning component is to use unsupervised learning, where training examples (e.g., *original* images from the previous section) are clustered, based on their feature (e.g., histogram), according to a similarity metric (e.g., Euclidean distance). In the automatic correction process, the image to be corrected is placed in the appropriate cluster (based on similarity with the training images), and corrections are applied based on the corrections that were performed to the training images in the same cluster. If this approach is taken, to perform clustering it is necessary to extract, from the new image, the same features that were extracted from the *original* images in the training stage (e.g., histogram). Note that those features could also be used to learn/automatically determine a given image's category (e.g., Impressionist painting).

Once the clustering has been performed, the type of correction that is made depends on the training scenario. In *scenario I*, for example, it would be possible to apply a sequence of operations, with known parameters, to the new image. The actual operations and parameters could be determined as a function of the corrections made to training images in the same cluster.

In *scenarios II*, and *III*, there is no access to operation lists within a cluster. During training, the features extracted from pairs of *original-corrected* images have to be designed to measure specific aspects of the image correction process, and determine the actual correction that could have produced the changes. In the analysis presented in this paper, for example, we focus on analyzing changes in contrast. From the *original/corrected* images, the choice was made to extract a feature that measured the mapping of L^* values (CIE Lighthness) from the *original* image to the *corrected* image. The results, as will be explained in the Data Analysis section, can be applied directly to correct the contrast of new images.

Analysis of Correction Process

At the *N.G.A.* the expert operator has been using an off-the-shelf software program for a number of years; the list of operation-parameter pairs of *scenario I* is not available. To have access to detailed operation information it would have been possible to build new software with functionality similar to that of the software used by the expert. Given his experience of several years with the current setup, however, this was not a viable option, particularly because it is not possible to implement *exactly* the same operations of the software currently being used (detailed information about the current implementation is not available). It would have been necessary for the expert to learn the effect of (even slightly) different operations, possibly affecting the correction process. Another possibility would have been to ask a different person to train the new system, but that was not adequate in terms of our study because one of the goals was to gain some insight into the image correction process as performed by an expert operator.

The expert was interviewed on several occasions, to obtain some knowledge about the image correction process, described next and outlined in 0.

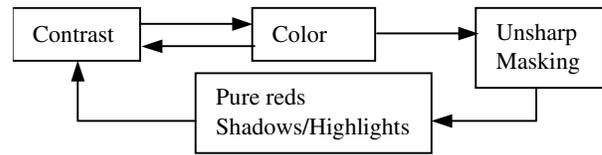


Figure 4. Image correction process.

Contrast: start by slightly increasing contrast. The color balance changes made in the following step affect luminosity, so large initial changes in contrast can create an image whose contrast is too high to begin with and that cannot be easily corrected.

Color Balance: add or subtract colors in the shadows, mid tones, or highlights. This is done until the colors match “fairly well,” after which the operator goes back to adjust the contrast slightly. Several iterations are performed, going back and forth between *color balance* and *contrast*.

Unsharp Masking: performed with typical values ranging between 30%-50%, 2.0-4.0 radius, and 0 threshold.

Local Retouching: isolated areas of the image are selected for adjustment. Usually pure reds look “too orange”, so the hue/saturation controls are used. Additionally, the shadows are sometimes separated and the levels control (used for contrast, in step one) is used to make them lighter. The same method is used to darken highlights if needed. Once the isolated areas have been corrected, the expert goes back to make new global adjustments, often performing the second pass a few hours later.

The descriptions provided by the expert are important in that they highlight several issues: (1) the process is highly subjective in each of the *correction factors* described earlier (*what, when, where, and how* corrections are made), (2) in spite of the subjectivity, the operator uses a fairly stable workflow in the correction process.

Data Analysis

The goal of the experiments was to analyze the *original* images (those scanned by the operator) and compare them with the images *corrected* by the operator. Two sets of images were used (see Annex A), one of 19 Dutch/Flemish portraits and one of 19 Impressionist paintings. We plotted the L^* histograms for each image (Fig. 5), and the L^* mapping of values of each *original* image to the L^* values of the corresponding *corrected* image (0).

As Fig. 5 shows, the differences in histograms between the two categories are quite significant. This suggests, as was discussed in the Automatic Correction section, that histograms could be used in the training stage to automatically classify images into different categories. Additionally, the histograms within a category (e.g., Impressionist paintings) could be used as features to cluster visually similar images together. Although this is difficult to

see in the figure, in the experiments, it was possible to find clusters in terms of visual appearance (in the same category) and their corresponding histogram clusters. We also found that the mapping of L^* to change contrast varied, and that the variation was related to variations in the histogram. In other words, the data suggests that new images can be clustered according to their histogram, and that their contrast can be changed using L^* mapping functions that are used for images with similar histograms. In 0, notice that clusters in the L^* mapping functions are quite evident, and that the mapping differs between the collections, when we plot the mean function for each of the collections (0, right).

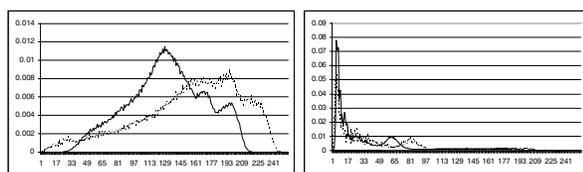


Figure 5. L^* histogram plots for original (solid lines) and corrected (dotted lines) images of Impressionist paintings (left) and Dutch/Flemish paintings (right).

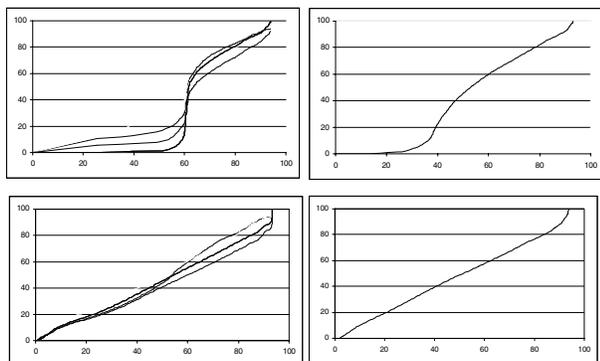


Figure 6. Functions that map L^* values from original to corrected images, for Impressionist paintings (top) and Dutch/Flemish paintings (bottom). Figures on the left show four separate clusters in each category, while figures on the right show the mean for each category.

Conclusions and Future Work

We proposed a framework to learn corrections to digital images, made by an expert, in a museum setting. We analyzed actual data from two sets of images corrected by an expert from the *National Gallery of Art* in Washington, D.C. (USA). In particular, we examined a set of images in two different categories (Impressionist, and Dutch/Flemish paintings). Our study showed that the two collections show differences in the histograms, and that the contrast corrections made were different between the two collections. This supports our hypothesis that although corrections are image-dependent, it is possible to find similar correction patterns in images that are visually similar

(i.e., in the same category). Since similar corrections were found for images in the same category, this study also suggests the possibility of constructing systems that learn correction patterns that are category specific, and apply them to images depending on their category. Our future work includes the implementation of the learning framework proposed, as well as a more extensive analysis of images and image features.

Acknowledgements

The authors wish to express their gratitude to Neal Johnson, at the National Gallery of Art for his generous help, and to Ric Snyder for helping provide access to the N.G.A.'s transparencies. Additionally, the authors would like to thank Jim Christensen of IBM for providing some of the software used in the project, and Gordon Braudaway of IBM for providing insight during many discussions and for also providing some of the software used in this project.

References

1. D. Adkins, et. al., "Color Mapping Using Neural Networks," *IS&T/SID Color Imaging Conference*, Scottsdale, AZ, 1993.
2. Boldrin et. al., "Learning Color Appearance Models," *IS&T 5th Color Imaging Conference*, Scottsdale, AZ, Nov. 1997.
3. P.G. Engeldrum, "Image Quality Modeling: Where Are We?," *IS&T PICS Conference*, Savannah, Georgia, 1999.
4. B. Funt, et. al., "Learning Color Constancy," *4th IS&T/SID Color Imaging Conference*, Scottsdale, AZ, 1995.
5. B. O. Hultgren, "The Role of a Virtual Observer in an Automatic, Image Processing System," *IS&T PICS Conference*, Portland, Oregon, May 1998.
6. P. Laihanen, et. al., "Automatic Colour Correction," *2nd IS&T/SID Color Imaging Conference*, Scottsdale, AZ, 1994.
7. F. Mintzer, "Developing Digital Libraries of Cultural Content for Internet Access," *IEEE Comm. Mag.*, Jan. 1999.
8. F.C. Mintzer, et. al., "Toward On-line, Worldwide Access to Vatican Library Materials," *IBM J.R&D.*, 4(2):139-162, March 1996.

Appendix

At the *N.G.A.*, the paintings were photographed by professional photographers in special studios. The resulting color transparencies (typically 8x10 inches) were scanned at the *N.G.A.* using the IBM Pro Imaging System. The images were then corrected by the expert, and published on the Internet. The original scans for the images published on the Internet were not available at the *N.G.A.*. Therefore, we replicated the IBM scanner setup at the IBM TJ Watson Research Center, and obtained copies of the original 8x10 transparencies. The transparencies were then scanned to replicate the *original* digital images. We used a collection of 38 images from the *N.G.A.*, 19 Impressionist paintings (19th/20 Century), and 19 Dutch/Flemish portraits (17th C).

Biography

Alejandro Jaimes received his Computing Systems Engineering degree from Universidad de los Andes, Bogota,

Colombia in 1994, the MS in Computer Science from Columbia University in 1997, and the M.Phil. degree in 2000. He is currently pursuing a PhD in Electrical Engineering at Columbia University in New York City.