Steps towards 'undigital' intelligent image processing: *Real-valued* image coding of photoquantimetric pictures into the JLM file format for the compression of Portable Lightspace Maps

Steve Mann[1], Corey Manders[1], Billal Belmellat[1], Mohit Kansal[1] and Daniel Chen[2]
[1]EyeTap Personal Imaging (ePi) Lab
Dept. of Electrical and Computer Engineering, University of Toronto
10 Kings College Rd, Toronto, ON, Canada M5S 3G4
E-mail: [mann, corey, billal, mohit] @ eyetap.org
[2] Human Media Lab (HML)
School of Computing and Information Science, Queen's University
Kingston, ON, Canada K7L 3N6
E-mail: chend@cs.queensu.ca

Abstract:

Recent advances in intelligent signal processing have made it possible to capture high dynamic range images which are better represented as an array of real numbers rather than the current convention of an array of integers. This paper proposes a solution to address the need for real, rather than just integer, image coding and file formats. Additionally, we propose that the **real-valued** data be linear in photoquantity (the quantity of light received by the camera) to avoid the image misrepresentation that occurs when a camera's non-linear dynamic range compressor and a display's dynamic range non-linear expander do not match. We present two novel image formats that achieve this: the Portable Lightspace Map (PLM) and its compressed version the JPEG Lightspace Map (JLM), that builds upon the JPEG compression scheme. The results of various compression levels for *real-valued* data and their corresponding file sizes are reported.

1. Introduction

The explosive growth of the digital cameras in the consumer electronics industry has caused an increase in spatial resolution by a factor of ten linearly (approximately 100 times the number of pixels) over the past ten years, and spatial resolution increases roughly exponentially with time, following roughly Moore's law, e.g. roughly 100 times the number of pixels every ten years. The digital camera industry and consequently the general consumer, has been focused on the "megapixel" race, or spatial resolution, while the importance of tonal resolution has been largely ignored. In fact, over the years, tonal resolution has remained relatively constant, at one byte (256 levels), or at most, 12 bits per pixel per color channel. Existing file formats for image encoding are capable of only this fixed tonal resolution, which means that much of the tonal information in a scene is lost.

In this paper we discuss the need for real valued representation as it applies to image formats. In addition, to preserve the tonal representation of an image, we propose that the **real-valued** data be linear in photoquantity (the quantity of light received by the camera) to avoid the misrepresentation of an image that occurs when a camera's non-linear compressor and a display's non-linear expander do not match. Our goal is to achieve an 'undigital' image representation which is independent of the capturing or displaying medium's particular properties, and also minimizes greatly quantization error through real valued representation. Our solution is presented through the two novel file formats, the Portable Lightspace Map (PLM) and its compressed version, the JPEG Lightspace Map.

2. Image range compression

Most cameras do not provide an output that varies linearly with light input. Instead, most cameras contain a unique non-linear dynamic range compressor, as illustrated in Fig. 1 which varies widely in its response function according to the particular camera system.

Historically, the dynamic range compressor in video cameras arose because it was found that televisions did not produce a linear response to a video signal. In particular, it was found that early cathode ray screens provided a light output approximately equal to voltage raised to the exponent of 2.5. Rather than build a circuit into every television to compensate for this nonlinearity, a partial compensation (exponent of 1/2.22) was introduced into the television camera at much lesser cost since there were far more televisions than television cameras in those days.

Coincidentally, the logarithmic response of human visual perception is approximately the same as the inverse of the response of a television tube (e.g. human visual response turns out to be approximately the same as the response of the television camera) [4]. For this reason, processing done on typical video signals will be on a perceptually relevant tone scale. Moreover, any quantization on such a video signal (e.g. quantization into 8 bits) will be close to ideal in the sense that each step of the quantizer will have associated with it a roughly equal perceptual change in perceptual units.

Most still cameras also provide built-in dynamic range compression. For example, the Nikon D2h camera captures internally in 12 bits (per pixel per color) and then applies dynamic range compression, and finally

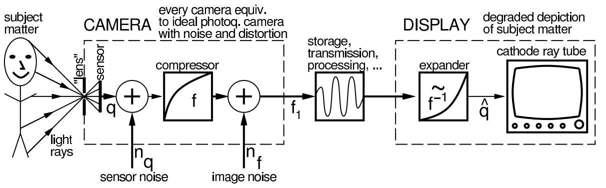

Figure 1. **Typical camera and display :** Light from subject matter passes through lens (typically approximated by simple algebraic projective geometry, e.g. an idealized "pinhole") and is quantified in units "q" by a sensor array where noise n_q is also added, to produce an output which is compressed in dynamic range by a typically unknown function f. Further noise n_f is introduced by the camera electronics, including quantization noise if the camera is a digital camera and compression noise if the camera produces a compressed output such as a jpeg image, giving rise to an output image $f_1(x,y)$. The apparatus that converts light rays into $f_1(x,y)$ is labelled CAMERA. The image f_1 is transmitted or recorded and played back into a DISPLAY system where the dynamic range is expanded again. Most cathode ray tubes exhibit a nonlinear response to voltage, and this nonlinear response is the expander. The block labelled "expander" is therefore not usually a separate device. Typical print media also exhibit a nonlinear response that embodies an implicit "expander".

outputs the range–compressed images in 8 bits (per pixel per color).

Some of the newer cameras, such as the Nikon D2h camera also allow output of images in a non–range–compressed "CCD-RAW" 12-bit (per pixel per color) format. However such CCD-RAW outputs are typically unstandardised and proprietary.

2.1 Range compressor expander mismatch

When range compressors were built into video cameras for the purpose of capturing data to be reproduced remotely, the display devices were all televisions with largely the same response to a video signal. Today, one may capture images with no notion of what the image may eventually be displayed on. The archived images may be displayed on analog or digital televisions, video projectors, electric eyeglasses (such as eyetap devices), print, volumetric displays, etc., just to name a few of the current possibilities. In the future, the number of options will certainly grow. Most importantly, the range expander in each of these devices will most likely vary. This means that to accurately display the image, with correct tonal representation, careful calibration is needed. In many cases, the calibration may be useless. This is simply because the range compression used by particular camera companies is proprietary, prohibiting accurate representation by arbitrary display devices. In essence, range compression results in a large probability of the compressor and expander not matching, resulting in improper tonal representation of images.

2.2 Range compression made one byte integers acceptable

One of the reasons for the slow progress of any standardized non–range–compressed file format is that the range compression helps to make one byte integers sufficient to represent the data in image compression such as JPEG file formats.

However, a number of new developments have:

- made it possible to capture "undigital" pixel data, i.e. as an array of REAL *8 (64 bit);
- made it practical to process the data (the proliferation of 64 bit computer architectures);
- made such capture desirable, and such data useful, as for example, in imaging where it is desired to combine multiple exposures of the same subject matter. In this case, for applications where linearity (homogeneity and superposition) is desired, it is preferable that the image data also be linear in the quantity of light.

Typically, because of the arithmetic involved in such processing, it is desired to have REAL *8 (double precision, i.e. 8 byte floats) rather than merely REAL *4 (single precision).

Thus, in raw form, there is an eightfold increase from INTEGER *1 (single byte integers) to REAL *8. Thus there is an even greater need for image compression when capturing "undigital" pictures.

2.3 "Undigital" pictures that are also photoquantimetrically linear

We propose to store images with no range compression whatsoever. Ideally, the values encoded in the file are simply measurements of the quantity (in the quantimetric sense [3]) of light present at each pixel element of the sensor over a particular period of time.

This description implies representation as greyscale images, but the idea is easily expandable to red, green and blue sensitivities or other multibanded, multispectral and color images. In terms of redisplaying the archived images, the display device simply needs to output the red, green and blue intensities collected by the sensor and archived in the file. There is no calibration needed (once the output of the device has been calibrated) and consequently an accurate representation of light is portrayed by the display device. The dynamic range compressor and expander cannot affect the image because they have been removed from the process.

However, the main benefit, beyond merely a true and accurate display, is that processing done on the quantimetric data is really linear processing. Thus, for example, deblurring done on the data so-represented is deblurring in lightspace, rather than homomorphic filtering. (It has been shown [2] that so-called linear filters, when used on images, are not at all linear, and are in fact incorrect homomorphic filters.)

2.4 Processing of traditional image formats

When video signals are processed using linear filters, there is an implicit homomorphic filtering operation on the photoquantity (a measure of the quantity of light present on a sensor array element [3]). As should be evident from Fig. 1, operations of storage, transmission, and image processing take place between approximately reciprocal nonlinear functions of dynamic range compression and dynamic range expansion.

Many users of image processing methodology are unaware of this fact, because there is a common misconception that cameras produce a linear output, and that displays respond linearly. Also, it is perceived that the nonlinearities in cameras and displays arise from defects and poor quality circuits, when in actual fact these nonlinearities are fortuitously present in display media and deliberately present in most cameras. Thus the effect of processing signals such as f_1 in Fig. 1 with linear filtering is, whether one is aware of it or not, homomorphic filtering. Tom Stockham advocated a kind of homomorphic filtering operation in which the logarithm of the input image was taken, followed by linear filtering (e.g. linear space invariant filters), followed by taking the antilogarithm [5].

In essence, what Stockham did not appear to realize, is that such homomorphic filtering is already manifest in simply doing ordinary linear filtering on ordinary picture signals (whether from video, film, or otherwise). In particular, the compressor gives an image $f_1 = f(q) = q^{1/2.22} = q^{0.45}$ (ignoring noise n_q and n_f) which has the approximate effect of $f_1 = f(q) = log(q+1)$ (i.e. roughly the same shape of curve, and roughly the same effect, example: to brighten the mid–tones of the image prior to processing). Similarly, a typical video display has the effect of undoing (approximately) this compression, e.g. darkening the mid–tones of the image after processing with $\hat{q} = \widetilde{f^{-1}}(f_1) = f_1^{2.5}$. Thus in some sense what Stockham did, without really realizing it, was to apply dynamic range compression to already range compressed images, then do linear filtering, then apply dynamic range expansion to images being fed to already expansive display media.

2.5 Correcting the nonlinear camera response problem

There exist certain kinds of image processing for which it is preferable to operate linearly on the photoquantity q. Such operations include sharpening of an image to undo the effect of the point spread function (PSF) blur of a lens, or to increase the camera's gain retroactively. We may also add two or more differently illuminated images of the same subject matter if the processing is done in photoquantities. What is needed in these forms of photoquantigraphic image processing is an *anti–homomorphic filter*. The manner in which an anti–homomorphic filter is inserted into the image processing path is shown in Fig. 2.

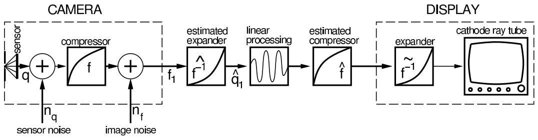

Figure 2. **The anti–homomorphic filter :** Two new elements $\hat{f}^{-1}$ and $\hat{f}$ have been inserted, as compared to Fig. 1. These are *estimates* of the the inverse and forward nonlinear response function of the camera. Estimates are required because the exact nonlinear response of a camera is generally not part of the camera specifications. (Many camera vendors do not even disclose this information if asked.) Because of noise in the signal f_1, and also because of noise in the estimate of the camera nonlinearity f, what we have at the output of $\hat{f}^{-1}$ is not q, but, rather, an estimate, $\tilde{q}$. This signal is processed using linear filtering, and then the processed result is passed through the estimated camera response function, $\hat{f}$, which returns it to a compressed tone scale suitable for viewing on a typical television, computer, or the like, or for further processing.

Previous work has dealt with the insertion of an anti–homomorphic filter in the image processing chain. However, in the case of using a camera in which the raw 12-bit data is available, processing using the raw data (NEF files), may proceed as shown in Fig. 3.

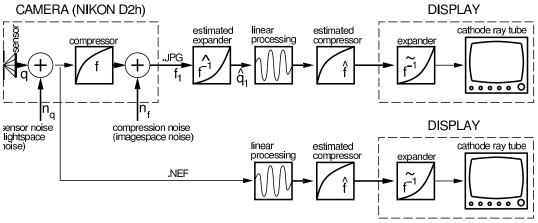

Figure 3. A modified method of photoquantimetric image processing (shown in figure 2), in which the raw data is available, and consequently no anti–homomorphic filter is necessary. Moreover, a comparison (e.g. comparametric analysis) between compressed and raw data is possible.

3. Applications to file formats

As mentioned in the previous section, an unfortunate consequence of using dynamic range compression is that often the compressors and expanders don't match. Even if they do match, there is still the problem that working with images typically makes linear filtering impossible, since we usually do not know the proprietary camera response function that a camera vendor has used.

Typically, camera sensors are laid out in a Bayer pattern[1] using red, green, and blue colour filters in order to produce colour images. The sensitivity to red, green, and blue wavelengths varies in different cameras. This leads to differing pixel values in the dynamically compressed image. Furthermore, displays (monitors, projectors, and printed matter) vary in terms of their representation of red, green and blue. These differences lead to inaccuracies in representation, in addition to the aforementioned inability to do true and accurate linear filtering. Often, this is noticeable in digital televisions playing digital video. For example, in playback of DVD recordings the red channel is often incorrectly displayed, and obviously out of balance with the green and blue channels. This contradicts the intuition that digital media is necessarily better.

3.1 Being "undigital"

The difference between discrete time signal processing and digital signal processing is whether or not the elements in a discrete lattice of samples are quantized. Thus, "digital" implies quantization (finite word length), and thus quantization noise, as well as the aforementioned artifacts of such noise having a homomorphic effect. Additionally, spatial resolutions of modern cameras have increased dramatically, and will likely continue to increase exponentially. As a result, spatial sampling will soon surpass (and has in some cases already surpassed) the Nyquist rate, dictacted by the rest of the system, such as lens, optics, etc., so that spatially, the image is sufficiently "undigital" that the tonal quantization becomes the main limiting factor.

Accordingly, what is desired is an "undigital" file format that can better represent analog (continuous real-valued) values, to which a REAL *8 representation provides a sufficiently good approximation, i.e. sufficiently "undigital" as to, when combined with linearity, permit true and accurate linear image processing. With the advent of 64 bit (8 byte) computer architectures, "undigital image processing" is now practical.

If the response of a camera is known, or the image is recorded in lightspace as photoquantities, knowing the response of the display device permits images to be displayed accurately. With computers moving to 64-bit processors (like the Apple Mac G5, AMD's Athlon 64 and Itanium) double precision values may be dealt with natively allowing photoquantities to be easily handled by computers. Rather than using pixel based storage such as portable pixmaps (PPMs) or jpeg[1] images or PNGs, we have proposed photoquantimetric file formats, specifically PLMs (Portable Lightspace Maps) and JLMs (Jpeg Lightspace Maps). Such file formats will allow for the efficient storage of accurately representable images, as well as the capability for true linear processing.

3.2 The PLM file format

We began by extending the Portable aNyMap (PNM) file format from its original PGM (portable grey map, i.e. INTEGER *1 greyscale) and PPM (portable pixmap, i.e. INTEGER *1 color), to include a REAL *8 image format, as well as a REAL *8 linearly quantimetric format.

Thus we have added four new types, P7, P8, P9, and PA to the existing P1-P6 types already defined:

- *P1 and P4* ASCII text and binary bitmap images (standard)
- *P2 and P5* ASCII text and binary greyscale images (standard)
- *P3 and P6* ASCII text and binary rgb colour images (standard)
- *P7 and PA* ASCII text and binary dynamically decompressed greyscale images (new)
- *P8 and PB* ASCII text and binary dynamically decompressed rgb colour images (new)

We also then developed a 64-bit file compression format similar to jpeg for dynamically decompressed high dynamic range images of type P7 to PA.

3.3 The need for compression of "undigital" images

Typically, images are very large when stored in our PLM file formats (i.e. 8 times larger). For example, a 640x480 colour PLM is approximately 7 megabytes. In contrast, most images today use just 24-bits per pixel which accounts for the standard 16 million colors (255^3). In this sense, PLMs are represented in 192-bit color. In its raw form, the 192-bit color image is, in many situations, not practical, but when compressed, will often be no larger than an ordinary JPEG image, because much of the oversize is informatically redundant. Therefore, the JLM (Jpeg Lightspace Map) could be used in many forms of media so that a simple and intuitive standard could be used for all imaging purposes.

4. Overview of algorithm

As mentioned, care was taken to follow the widely accepted and used JPEG algorithm. This was done because of the algorithm's simplicity and proven reliability. The major difference between a JLM and JPEG is in the use of double precision numbers for JLM images. The open source library set out by the joint photographic experts group is based on INTEGER *1, and hence is heavily optimized towards that one byte size. Whereas a standard PLM file allows each pixel to carry a quantity of light (q) that can range from 0 to the maximum allowable double value. This flexibility easily handles any desired precision required today and that of any image in the foreseeable future. The main steps in the compression process are outlined in Fig. 4. As described, the compression is based on the use of quantizing (or elimination of less important data) and that of run length encoding (a simplified version of Huffman coding).

4.1 Discussion of key parts

- *Quantization* The construction of quantization tables are key to the success of any jpeg-type compression routine. In the case of the JPEG Lightspace Map a base equation utilizing a dynamic quality factor was used to

---

[1] See http://www.jpeg.org

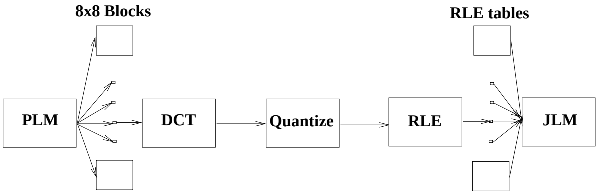

Figure 4. **Schematic of JPEG compression:** A simple block diagram depicting the process. The image is first broken up into 8x8 blocks as to minimize the strain on the DCT process. From there each block is DCT encoded. After applying quantization to these matrices, run length encoding following a zigzag pattern is written to a table in the final file. The summation of all the 8x8 blocks makes up the final JLM file.

determine the exact quantization values. Higher values in a table result in good compression and poor quality and lower values result in less compression in higher quality. A compromise between the two was found as to be discussed in the experiment section.

- *Run Length Encoding* If the quantization tables used in the quantization process worked well, the resulting 8x8 blocks should be plentiful with zeros. The run length encoding that follows a zigzag pattern (to capitalize on the tendencies of the DCT), uses simple encoding logic to prevent the storage of less important information that has been reduced to zero from being stored into 8-bytes. The exact structure and developed standard will be discussed later in this paper. The "throwing out" of information is simply the removal of the elements of the DCT with a low-bin count and/or high frequency. This can have adverse effects in images with many sharp lines and text, as what is presumed to be noise in this instance could be texture.

5. Results

An example of the compression is shown in Fig. 5. It depicts a typical lightspace image and how the different quality factors affect the appearance of the image.

The graph shown in Fig. 6 gives us a sense of the amount of compression we obtain when working with a typical lightspace image. It was found after experimenting with a variety images that a quality factor of (7) on a scale from (1-12) produced a compressed image with comparable quality to the original.

When using sharp lined images and text the JLM compression routine tended to blur these edges and reduce quality notably. It is obvious that user discretion is needed to acknowledge a higher quality factor should be used on sharp images, while more scenic and flowing images can do with a smaller factor.

5.1 JLM file syntax

It was attempted to keep the syntax as simple as possible so manipulation and reading is quick and easy.

Header :

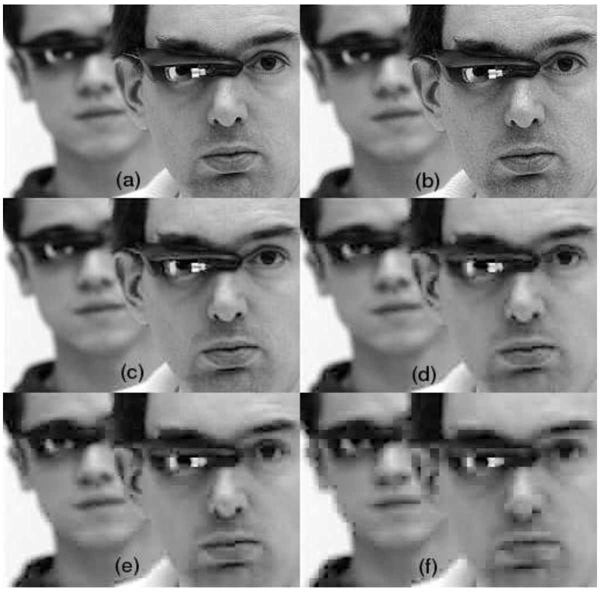

Figure 5. **Varying levels of compression :** Shown are six different levels of compression on the same image. The original image (a) is compressed to form the other five images. It should be quite clear the increasing compression (or decreasing quality factor) moves left to right and downwards across the figure. We can start to see a notable difference in the images at (d) which is at a quality factor of (5). As the compression increases we can see that a blocky effect appears. Also, early on in the compression process detailed sections of the eyetap worn by Professor Mann (the one in the foreground) show blurring. This is an example of the utility having difficulty with sharp lines and sudden color change. However, if you look at Professor Mann's forehead, where there is little color change, we can see little difference between the original and further compressed images.

- *image identifier* JLM for JPEG Lightspace Map
- *height and width* Picture dimensions
- *quantization tables* Equation used to obtain (involving array indicies)
- *quality factor used*

Huffman table (for each 8x8 block) :

- *number of non-zero numbers in matrix* (unsigned char)
- *number of zeros preceding non-zero value* (unsigned char)
- *non-zero value* (double)

6. "Undigital photography" as a form of visual art

The proposed image representations are also useful for the production of visual art, such as simulation of multiple exposures, or accurate simulation of other photographic phenomena.

For example, the proposed image representations were used to create various multiple exposure art forms,

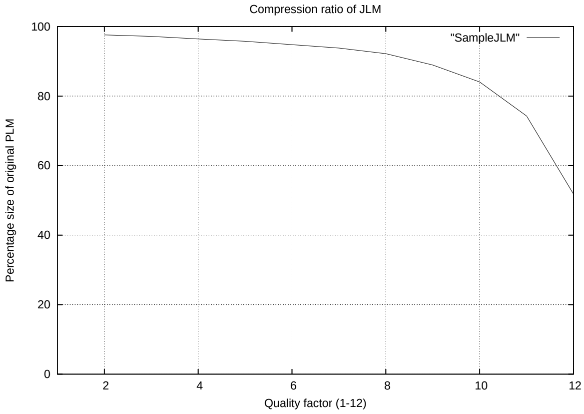

Figure 6. **Percentage Compressed vs. Quality Factor :** We can see that after a certain threshold (ten in this instance) the rate of compression is small. To capitalize on this one should not apply a smaller quality factor as the compression benefits will be quite small in comparison.

and cement these together into a single image. Such combining differently illuminated exposures of the same subject matter, is illustrated in Fig. 7.

7. Conclusion and future work

In this paper, we have proposed two file formats for standardized storage, transmission, and processing of quantimetrically linear image data, using REAL rather than INTEGER quantities. The first of these, the Portable Lightspace Map (PLM), is an extension of the PNM file format, but the file sizes are quite large. The second, the Jpeg Lightspace Map (JLM) is a compressed version of the PLM that uses a data compression methodology that builds upon JPEG compression. Both of the two proposed file formats are intended to be used independently, without dynamic range compression or expansion, and are therefore truly universal image formats that permit quantimetrically linear image processing. The formats are independent of the particular camera that takes the picture and the particular device that displays them, and hence allow the user to reclaim the original photoquantities with ease. Also, as the formats are based in lightspace they provide an intuitive and easy to use foundation for intelligent image processing. With the creation of the JLM, lightspace file sizes are manageable for the everyday user and as a result these formats can now more easily be adopted by the computing community. Current work on the project is focused on optimizing and integrating the JLM compression utility to perform on 64-bit based processors and run symbiotically with other lightspace utilities already in use.

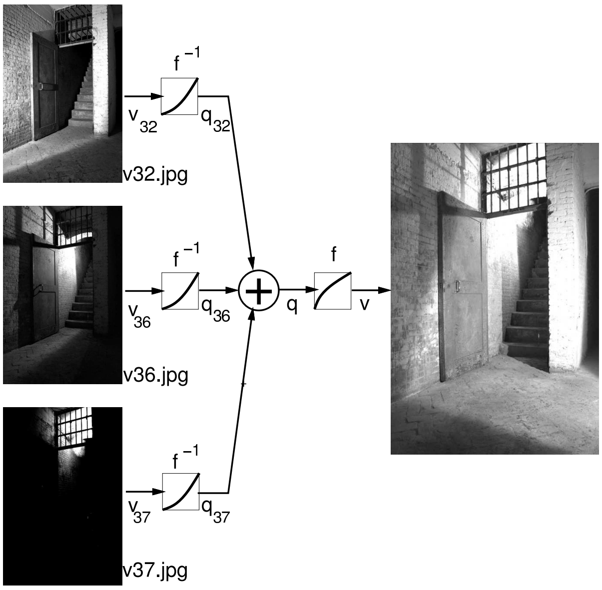

Figure 7. Various exposures to different sources of illumination are combined quantimetrically. In, for example, v32.jpg, the open basement door under cell block "A" on Alcatraz Island, is exposed to light from a flash lamp held to the left. The flash lamp is then moved to the right, to illuminate the scene from the right, in exposure v36.jpg. Finally, exposure v37.jpg captures light coming from upstairs, beyond the jail bars above the door. Each exposure is made by illuminating the space from one of various viewpoints. These pictures are then converted into lightspace by applying an estimate of the inverse of the camera's photographic response function[2]. The resulting photographic quantities are added together, and the combined exposure is then converted from light space back into a picture.

References

[1] B. Bayer. Color imaging array, 1976. U.S. Patent 3,971,065.

[2] S. Mann. Comparametric equations with practical applications in quantigraphic image processing. *IEEE Trans. Image Proc.*, 9(8):1389–1406, August 2000. ISSN 1057-7149.

[3] S. Mann. *Intelligent Image Processing.* John Wiley and Sons, November 2 2001. ISBN: 0-471-40637-6.

[4] C. Poynton. *A technical introduction to digital video.* John Wiley & Sons, 1996.

[5] T. G. Stockham, Jr. Image processing in the context of a visual model. *Proc. IEEE*, 60(7):828–842, July 1972.