Comparametric Equations with Practical Applications in Quantigraphic Image Processing

Steve Mann

***Abstract*—It is argued that, hidden within the flow of signals from typical cameras, through image processing, to display media, is a homomorphic filter. While homomorphic filtering is often desirable, there are some occasions where it is not. Thus, cancellation of this implicit homomorphic filter is proposed, through the introduction of an antihomomorphic filter. This concept gives rise to the principle of quantigraphic image processing, wherein it is argued that most cameras can be modeled as an array of idealized light meters each linearly responsive to a semi-monotonic function of the quantity of light received, integrated over a fixed spectral response profile. This quantity is neither radiometric nor photometric, but, rather, depends only on the spectral response of the sensor elements in the camera. A particular class of functional equations, called comparametric equations, is introduced as a basis for quantigraphic image processing. Comparametric equations are fundamental to the analysis and processing of multiple images differing only in exposure. The well-known "gamma correction" of an image is presented as a simple example of a comparametric equation, for which it is shown that the underlying quantigraphic function does not pass through the origin. For this reason it is argued that exposure adjustment by gamma correction is inherently flawed, and alternatives are provided. These alternatives, when applied to a plurality of images that differ only in exposure, give rise to a new kind of processing in the "amplitude domain" (as opposed to the time domain or the frequency domain). While the theoretical framework presented in this paper originated within the field of wearable cybernetics (wearable photographic apparatus) in the 1970s and early 1980s, it is applicable to the processing of images from nearly all types of modern cameras, wearable or otherwise. This paper is a much revised draft of a 1992 peer-reviewed but unpublished report by the author, entitled "Lightspace and the Wyckoff principle."**

***Index Terms*—Comparametric equation, comparametric plot, image processing, lightspace, personal imaging, photography, quantigraphic imaging, wearable cybernetics, Wyckoff principle.**

I. INTRODUCTION

THE theory of quantigraphic image processing, with comparametric equations, arose out of the field of wearable cybernetics, within the context of so-called mediated reality (MR) [1] and personal imaging [2]. However, it has potentially much more widespread applications in image processing than just the wearable photographic personal assistant for which it was developed. Accordingly, a general formulation that does not necessarily involve a wearable photographic system will be given.

Manuscript received March 31, 1999; revised March 1, 2000. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Frater.

The author is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ont., M5S 3G4, Canada.

Publisher Item Identifier S 1057-7149(00)06143-1.

II. WYCKOFF PRINCIPLE AND THE RANGE OF LIGHT

The quantity of light falling on an image sensor array, or the like, is a real valued function $q(x, y)$ of two real variables x and y. An image is typically a degraded measurement of this function, where degredations may be divided into two categories, those that act on the domain (x, y) and those that act on the range q. Sampling, aliasing, and blurring act on the domain, while noise (including quantization noise) and the nonlinear response function of the camera act on the range q.

Registering and combining multiple pictures of the same subject matter will often result in an improved image of greater definition. There are four classes of such improvement:

1) increased spatial resolution (domain resolution);
2) increased spatial extent (domain extent);
3) increased tonal fidelity (range resolution);
4) increased dynamic range (range extent).

A. What is Good for the Domain is Good for the Range

The notion of producing a better picture by combining multiple input pictures has been well-studied with regards to the *domain* (x, y) of these pictures. Horn and Schunk, for example, provide means of determining optical flow [3], and many researchers have then used this result to spatially *register* multiple images in order to provide a single image of increased spatial resolution and increased spatial extent. Subpixel registration methods such as those proposed by [4] and [5] attempt to increase *domain resolution*. These methods depend on slight (subpixel) shift from one image to the next. Image compositing (mosaicking) methods such as those proposed by [6]–[8] attempt to increase *domain extent*. These methods depend on large shifts from one image to the next.

Methods that are aimed at increasing *domain resolution* and *domain extent* tend to also improve tonal fidelity, to a limited extent, by virtue of a signal averaging and noise reducing effect. However, we shall see in what follows, a generalization of the concept of signal averaging called quantigraphic signal averaging. This generalized signal averaging allows images of different exposure to be combined to further improve upon tonal fidelity (*range resolution*), beyond improvements possible by traditional signal averaging. Moreover, the proposed methodology drastically increases dynamic range (*range extent*). Just as spatial shifts in the domain (x, y) improve the image, we will also see how exposure shifts (shifts in the range, q) can, with the proposed methodology, result in even greater improvents to the image.

B. Extending Dynamic Range and Improvement of Range Resolution by Combining Differently Exposed Pictures of the Same Subject Matter

The principles of quantigraphic image processing and the notion of using differently exposed pictures of the same subject matter to make a picture composite of extended dynamic range was inspired by the pioneering work of Charles Wyckoff who invented so-called "extended response film" [9], [10].

Most everyday scenes have a far greater dynamic range than can be recorded on a photographic film or electronic imaging apparatus. However, a set of pictures, that are identical except for their exposure, collectively show us much more dynamic range than any single picture from that set, and also allow the camera's response function to be estimated, to within a single constant scalar unknown [6], [11], [12].

A set of functions

$$f_i(\mathbf{x}) = f(k_i q(\mathbf{x})) \tag{1}$$

where k_i are scalar constants, is known as a Wyckoff set [6], [12]. A Wyckoff set of functions, $f_i(\mathbf{x})$ describes a set of images differing only in exposure, when $\mathbf{x} = (x, y)$ is the continuous spatial coordinate of the focal plane of an electronic imaging array (or piece of film), q is the quantity of light falling on the array (or film), and f is the unknown nonlinearity of the camera's (or combined film's and scanner's) response function. Generally, f is assumed to be a pointwise function, e.g., invariant to $\mathbf{x}$.

C. Photoquantity q

The quantity, q, in (1), is called the *photoquantigraphic* quantity [13], or just the photoquantity (or photoq) for short. This quantity is neither radiometric (e.g. neither *radiance* nor *irradiance*) nor photometric (e.g. neither *luminance* nor *illuminance*). Most notably, since the camera will not necessarily have the same spectral response as the human eye, or, in particular, that of the photopic spectral luminous efficiency function as determined by the CIE and standardized in 1924, q is neither brightness, lightness, luminance, nor illuminance. Instead, quantigraphic imaging measures the quantity of light integrated over the spectral response of the particular camera system

$$q = \int_0^\infty q_s(\lambda) s(\lambda)\, d\lambda \tag{2}$$

where $q_s(\lambda)$ is the actual light falling on the image sensor and s is the spectral sensitivity of an element of the sensor array. It is assumed that the spectral sensitivity does not vary across the sensor array.

D. Camera as an Array of Lightmeters

The quantity q reads in units that are quantifiable (e.g. linearized or logarithmic), in much the same way that a photographic light meter measures in quantifiable (linear or logarithmic) units. However, just as the photographic light meter imparts to the measurement its own spectral response (e.g., a light meter using a selenium cell will impart the spectral response of selenium cells to the measurement) quantigraphic imaging accepts that there will be a particular spectral response of the camera, which will define the quantigraphic unit q. Each camera will typically have its own quantigraphic unit. In this way, the camera may be regarded as an array of lightmeters, each being responsive to the quantigral

$$q(x, y) = \int_0^\infty q_{ss}(x, y, \lambda) s(\lambda)\, d\lambda \tag{3}$$

where q_{ss} is the spatially varying spectral distribution of light falling on the image sensor.

Thus, varying numbers of photons of lesser or greater energy (frequency times Planck's constant) are absorbed by a given element of the sensor array, and, over the temporal quantigration time of a single frame in the video sequence (or the exposure time of a still image) result in the photoquantity given by (3).

In the case of a color camera, or other color processes, $q(x, y)$ is simply a vector quantity. Color images may arise from as little as two channels, as in the old bichromatic (orange and blue) motion pictures, but more typically arise from three channels, or sometimes more as in the four color offset printing, or even the high quality Hexachrome printing process. A typical color camera might, for example, include three channels, e.g., $[q_r(x, y), q_g(x, y), q_b(x, y)]$, where each component is derived from a separate spectral sensitivity function. Alternatively, another space such as YIQ, YUV, or the like, may be used, in which, for example, the Y (luminance) channel has full resolution and the U and V channels have reduced (e.g., half in each linear dimension giving rise to one quarter the number of pixels) spatial resolution and reduced quantizational definition. In this paper, the theory will be developed and explained for greyscale images, where it is understood that most images are color images, for which the procedures are applied either to the separate color channels, or by way of a multichannel quantigrahic analysis. Thus in both cases (greyscale or color) the continuous spectral information $q_s(\lambda)$ is lost through conversion to a single number q or to typically three numbers, q_r, q_g, q_b. Although it is easiest to apply the theory of this paper to color systems having distinct spectral bands, there is no reason why it cannot also be applied to more complicated polychromatic, possibly tensor, quantigrals.

Ordinarily cameras give rise to noise, e.g., there is noise from the sensor elements and further noise within the camera (or equivalently noise due to film grain and subsequent scanning of a film, etc.). Thus a goal of quantigraphic imaging is to attempt to estimate the photoquantity q, in the presence of noise. Since $q_s(\lambda)$ is destroyed, the best we can do is to estimate q. Thus q is the fundamental or "atomic" unit of quantigraphic image processing.

E. Accidentally Discovered Compander

Most cameras do not provide an output that varies linearly with light input. Instead, most cameras contain a dynamic range compressor, as illustrated in Fig. 1. Historically, the dynamic range compressor in video cameras arose because it was found that televisions did not produce a linear response to the 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

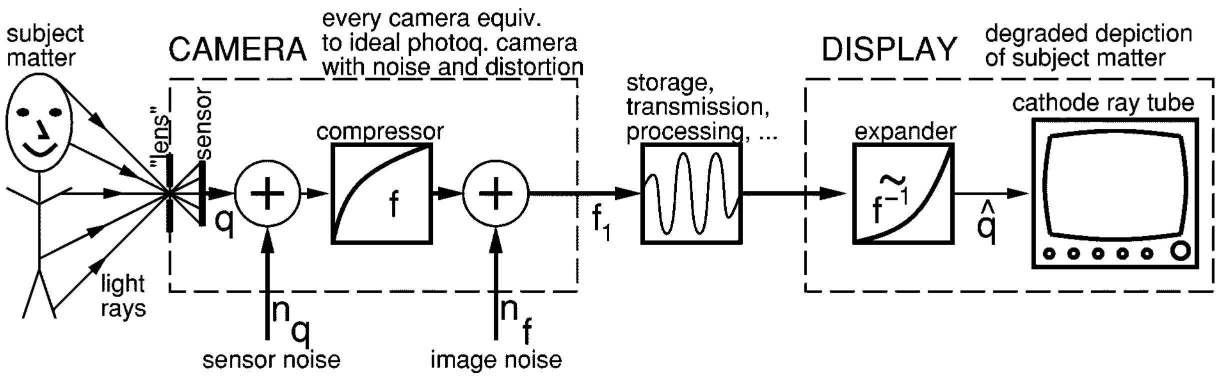

Fig. 1. *Typical camera and display:* light from subject matter passes through lens (typically approximated with simple algebraic projective geometry, e.g. an idealized "pinhole") and is quantified in units "q" by a sensor array where noise n_q is also added, to produce an output which is compressed 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 labeled 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 labeled "expander" is generally a side effect of the display, and is not usually a separate device. It is depicted as a separate device simply for clarity. Typical print media also exhibit a nonlinear response that embodies an implicit "expander."

television to compensate for this nonlinearity, a partial compensation (exponent of 1/2.22) was introduced into the television camera at much lesser total cost since there were far more televisions than television cameras in those days before widespread deployment of video surveillance cameras and the like. Indeed, the original model of television is suggested by the names of some of the early players: ABC (American Broadcasting Corporation); NBC (National Broadcasting Corporation); etc.. Names like this suggest that they envisioned a national infrastructure in which there would be one or two television cameras and millions of television receivers.

Through a very fortunate and amazing coincidence, 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) [14], [15]. 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.

Fig. 2 shows plots of the compressor (and expander) used in video systems together with the corresponding logarithm $\log(q+1)$, and antilogarithm $\exp(q)-1$, plots of the human visual system and its inverse. (The plots have been normalized so that the scales match.)

With images in print media, there is a similarly expansive effect in which the ink from the dots bleeds and spreads out on the printed paper, such that the mid tones darken in the print. For this reason printed matter has a nonlinear response curve similar in shape to that of a cathode ray tube (e.g., the nonlinearity expands the dynamic range of the printed image). Thus cameras designed to capture images for display on video screens have approximately the same kind of built-in dynamic range compression suitable for print media as well.

It is interesting to compare this naturally occurring (and somewhat accidental) development in video and print media with the deliberate introduction of companders (compressors

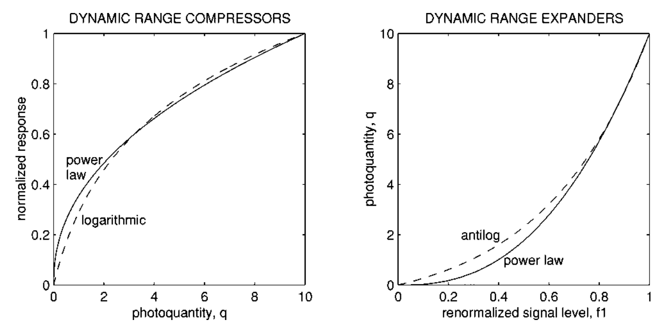

Fig. 2. The power law dynamic range compression implemented inside most cameras has approximately the same shape of curve as the logarithmic function, over the range of signals typically used in video and still photography. Similarly, the power law response of typical cathode ray tubes, as well as that of typical print media, is quite similar to the antilog function. Therefore, the act of doing conventional linear filtering operations on images obtained from typical video cameras, or from still cameras taking pictures intended for typical print media, is, in effect, homomorphic filtering with an approximately logarithmic nonlinearity.

and expanders) in audio. Both the accidentally occurring compression and expansion of picture signals and the deliberate use of logarithmic (or mu-law) compression and expansion of audio signals serve to allow 8 bits to be used to often encode these signals in a satisfactory manner. (Without dynamic range compression, 12 to 16 bits would be needed to obtain satisfactory reproduction.)

Most still cameras also provide dynamic range compression built into the camera. For example, the Kodak DCS-420 and DCS-460 cameras capture internally in 12 bits (per pixel per color) and then apply dynamic range compression, and finally output the range-compressed images in 8 bits (per pixel per color).

F. Why Stockham was Wrong

When video signals are processed, using linear filters, there is an implicit homomorphic filtering operation on the photoquantity. As should be evident from Fig. 1, operations of storage,

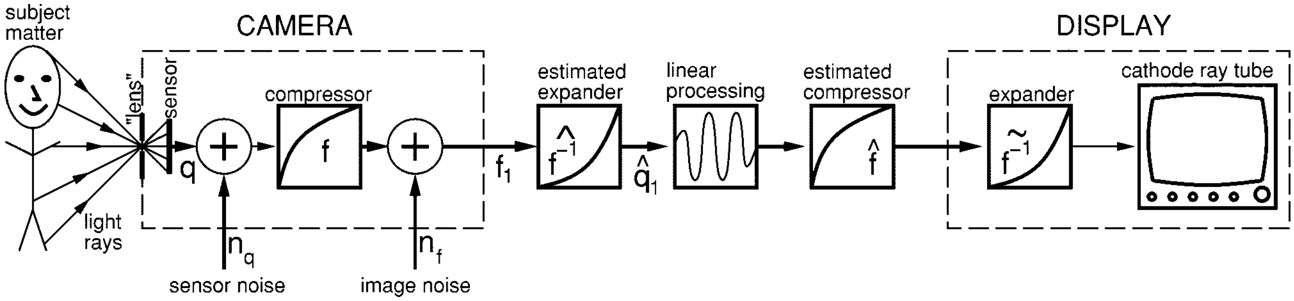

Fig. 3. *The anti-homomorphic filter:* Two new elements $\hat{f}^{-1}$ and $\hat{f}$ have been inserted, as compared to Fig. 1. These are *estimates* of the the inverse and forward nonlinear response function of the camera. Estimates are required because the exact nonlinear response of a camera is generally 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.

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. In fact there is a common misconception that 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.

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 [16].

In essence, what Stockham didn't 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 $f_1 = f(q) = \log(q+1)$ (e.g., roughly the same shape of curve, and roughly the same effect, e.g., to brighten the mid-tones of the image prior to processing), as shown in Fig. 2. 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 $\hat{q} = \tilde{f}^{-1}(f_1) = f_1^{2.5}$.

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.

G. On the Value of Doing the Exact Opposite of What Stockham Advocated

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. It is interesting to note that many textbooks and papers that describe image restoration (e.g. deblurring an image) fail to take into account the inherent nonlinearity deliberately built into most cameras.

What is needed to do this deblurring and other kinds of quantigraphic 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. 3.

Consider an image acquired through an imperfect lens that imparts a blurring to the image. The lens blurs the actual spatiospectral (spatially varying and spectrally varying) quantity of light $q_{ss}(x, y, \lambda)$, which is the quantity of light falling on the sensor array just prior to being *measured* by the sensor array

$$\tilde{q}_{ss}(x, y, \lambda) = \iint B(x-u, y-v) q_{ss}(u, v, \lambda)\, du\, dv. \tag{4}$$

This blurred spatiospectral quantity of light $\tilde{q}_{ss}(x, y, \lambda)$ is then photoquantified by the sensor array

$$\begin{aligned} q(x, y) &= \int_0^\infty \tilde{q}_{ss}(x, y, \lambda) s(\lambda)\, d\lambda \\ &= \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty B(x-u, y-v) q_{ss}(u, v, \lambda) \\ &\quad \cdot s(\lambda)\, du\, dv\, d\lambda \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty B(x-u, y-v) \\ &\quad \cdot \left(\int_0^\infty q_{ss}(u, v, \lambda) s(\lambda)\, d\lambda \right) du\, dv \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty B(x-u, y-v) q(u, v)\, du\, dv \end{aligned} \tag{5}$$

which is just the blurred photoquantity q.

Thus the antihomomorphic filter of Fig. 3 can be used to better undo the effect of lens blur than traditional linear filtering which simply applies linear operations to the signal f_1 and therefore operates homomorphically rather than linearly on the photoquantity q.

Thus we see that in many practical situations, there is an articulable basis for doing exactly the opposite of what Stockham advocated (e.g., expanding the dynamic range of the image before processing and compressing it afterward as opposed to what Stockham advocated which was to compress the dynamic range before processing and expand it afterward).