CAMERA RESPONSE FUNCTION RECOVERY FROM DIFFERENT ILLUMINATIONS OF IDENTICAL SUBJECT MATTER

Corey Manders, Chris Aimone, Steve Mann

University of Toronto
Dept. of Electrical and Computer Engineering
10 King's College Rd.
Toronto, Canada

ABSTRACT

A robust method of camera response function estimation, using pictures taken by differently illuminating the same subject matter, is presented. The method solves for the response function directly using superposition constraints imposed by different combinations of two (or more) lights to illuminate the same subject matter. Previous methods of computing camera response functions typically used differing exposures of identical subject matter, leading to uniqueness problems (underconstrained due to comparametric periodicity or fractal ambiguity). The method used in this paper overcomes this problem. Finally, we compare the method of the paper to previous methods and find the new method outperforms the previous work.

1. INTRODUCTION: A SIMPLE CAMERA MODEL

While the geometric calibration of cameras is widely practiced and understood [1][2], often much less attention is given to the camera response function (how the camera responds to light). In digital cameras, the camera response function maps the actual quantity of light impinging on each element of the sensor array to the pixel values that the camera outputs.

Linearity (which is typically not exhibited by most camera response functions) implies the following two conditions:

1. Homogeneity: A function is said to exhibit homogeneity if and only if $f(ax) = af(x)$, for all scalar a.
2. Superposition: A function is said to exhibit superposition if and only if $f(x + y) = f(x) + f(y)$.

The two are often written together, as: $f(ax + by) = af(x) + bf(y)$. However, for the purposes of this paper we wish to consider homogeneity and superposition separately.

In image processing, homogeneity arises when we compare differently exposed pictures of the same subject matter. Superposition arises when we superimpose (superpose) pictures taken from differently illuminated instances of the same subject matter, using a simple *law of composition* such as addition (i.e. using the property that light is additive).

A variety of techniques have been proposed to recover camera response functions, such as using charts of known reflectance, and using different exposures of the same subject matter [3][4][5][6]. The method proposed in this paper differs from other methods in that it does not require the use of charts, nor a camera that is capable of adjusting its exposure. The method is very easy to use, produces very accurate results and requires only that the camera has an exposure lock feature. Furthermore, the technique is useful in situations where differently illuminated, rather than differently exposed pictures are available. For example, a camera in a typical building or dwelling may observe an at least partially static scene during which time various lights are turned on and off throughout the day.

The comparagram, as defined in [3][7][4][8] has been widely used as a tool for the comparison of multiple differently exposed pictures of the same subject matter. With enough data, a direct nonparametric solution for the camera response function can be obtained, otherwise, a semi-parametric method such as Candocia's piecewise linear comparametric method will often provide better results[3]. A drawback of completely nonparametric methods is that comparametric periodicity (periodicity in the amplitude domain, i.e. amplitude "ripples", also known as fractal ambiguity[9] and comperiodicity[10]), plague the result unless more than two input images are used with exposure differences that are inharmonic (in the amplitude domain).

The method we propose in this paper uses the notion of superposition rather than homogeneity to solve for the camera response function. In this method the linear constraint of superposition disambiguates comparametric periodicity.

The following technique is used: in a dark environment, set up two distinct light sources. Take three pictures, one with each light on individually (p_a, p_b), and one with the two lights on together (p_c). From this data we solve for the camera response function f by using the following constraints: For the i^{th} pixel position in each of the three images: $p_a[i] = f(q_a)$, $p_b[i] = f(q_b)$, and $p_c = f(q_a + q_b)$. Where the quantity q is known as the *photographic quantity* or *photoquanity*[3][8]. Note that the photoquantity is neither radiance, irradiance, luminance, nor illuminance, rather, it is a unit of light, unique to the spectral response of a particular camera. The results obtained through this method were more accurate than using homogeneity (e.g. comparagrams) or typically available (coarsely quantized) charts.

In this paper we also propose an alternative construct which we refer to as the superposigram. Just as the comparagram provides an insightful analysis of homogeneity, the superposigram provides an insightful analysis of superposition.

2. THE CAMERA RESPONSE FUNCTION

The camera response function f may in general be modeled by cascading two non-linear functions as shown in figure 1. In this diagram, the *photographic quantity* (*photoquantity*) of light measured by the sensor, is mapped into pixel space by a non-linear dynamic range compression function and a uniform quantizer. We call the first function the Range Compression Function because most camera response functions, such as the familiar gamma mapping, are convex. In this paper we assume that this function is monotonic and convex[11]. The quantizer in turn maps the range compressed photoquantities into discrete pixel values.

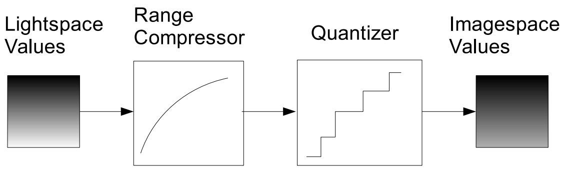

Fig. 1: The completed camera model as used in the paper. Lightspace values collected by the camera sensor elements are compressed with a non-linear function, and then quantized to yield pixel values.

By assumption the Range Compression Function is monotonic. Thus: photoquantities in the range $[q_1, q_2), q_1 < q_2$ will result in some pixel value p_1; photoquantities in the range $[q_2, q_3), q_2 < q_3$, will result in pixel value p_2; with $p_1 < p_2$. We then simplify our analysis by approximating the Range Compression Function as being linear between quantization points, and by assuming that the probability distribution of the measured photoquantities is uniform in this range. Therefore, given pixel value p_x, the maximum likelihood estimate of the original photoquantity $\bar{q}_x$ is given by:

$$f^{-1}(p_x) = q_x = \frac{q_x + q_{x+1}}{2}, \tag{1}$$

where q_x and q_{x+1} are lightspace quantization points for pixel value p_x and α is an arbitrary nonzero scale factor.

3. SOLVING FOR THE INVERSE CAMERA RESPONSE FUNCTION

Since the property of superposition holds with photoquantities, we can form the following equation:

$$f^{-1}(f(q_a)) + f^{-1}(f(q_b)) = \hat{q}_c. \tag{2}$$

By equation (1), $\hat{q}_c = \bar{q}_a + \bar{q}_b$ and by our assumptions, $\hat{q}_c$ is the maximum likelihood estimate of $q_a + q_b$ given our prior knowledge of $f(q_a)$ and $f(q_b)$. We can now form the superposition equation:

$$f^{-1}(f(\bar{q}_a)) + f^{-1}(f(\bar{q}_b)) = f^{-1}(f(\bar{q}_a + \bar{q}_b)) + \bar{\epsilon}_Q, \tag{3}$$

where $\bar{\epsilon}_Q$ is the mean error due to quantization of $\hat{q}_c$. Under the assumptions stated, we can solve for f^{-1}, i.e. the mapping from pixel value p_x to maximum likelihood (ML) photoquantities $\bar{q}_x$, by minimizing the the following equation:

$$e = \sum_{\forall n} \left(f^{-1}(p_a[n]) + f^{-1}(p_b[n]) - f^{-1}(p_c[n]) \right)^2, \tag{4}$$

where $p_a[n], p_b[n], p_c[n]$ are the n^{th} pixels of three images taken of a scene with constant exposure and three illumination permutations of two light sources in an otherwise dark environment. Pixel values p_a and p_b are from images of the scene with each of the two light sources turned on independently. Pixel value p_c is from the image of the scene with both light sources turned on together, as shown in Fig 2.

Fig. 2: One of the many datasets used in the computation of the superposigram. Leftmost: Picture of Deconism Gallery with only the upper lights turned on. Middle: Picture with only the lower lights turned on. Rightmost: Picture with both the upper and lower lights turned on together.

Since digital cameras output a finite range of discrete pixel values, care must be taken when applying the assumptions made at the ends of the camera's range where clipping occurs. In the remainder of the paper, we will assume that the camera outputs pixel values in the range $[0, 255]$ with clipping occurring at 0 and 255. This is not always the case, but the modification to the analysis under other conditions is very simple.

Using the proposed model, pixel values 0 and 255 are produced by the range of photoquantities $[0, q_1)$ and $[q_{255}, \infty)$ respectively. Since the range $[q_{255}, \infty)$ is infinite in size, the assumption that the distribution of photoquantities that produce a pixel value of 255 will approach a uniform distribution over a finite number of images will obviously not hold. A similar argument applies for pixel value 0. We therefore do not try to solve for $f^{-1}(0)$ or $f^{-1}(255)$, or equivalently, to solve for $\bar{q}_0$ and $\bar{q}_{255}$. Instead we can solve for the quantization points q_1 and q_{255}. This allows us to conclude that if we measure a pixel value of 0 with the camera, a quantity of light below q_1 was measured. Similarly, a pixel value of 255 represents a photoquantity greater than q_{255}. The method of solving for these thresholds is presented later in the paper.

In accordance with this development, we define f^{-1} as the mapping from pixel values $(1, 2, 3...254)$ to the maximum likely photoquantities $(\bar{q}_1, \bar{q}_2, \bar{q}_3...\bar{q}_{254})$. We can now write equation 4 more simply as:

$$e = \sum_{\substack{\forall n, P_a, P_b \\ P_c \neq 0, 255}} \left(\bar{q}_{p_a[n]} + \bar{q}_{p_b[n]} - \bar{q}_{p_c[n]} \right)^2 \tag{5}$$

Equation (5) can be efficiently minimized using a singular value decomposition (SVD). To do this, we represent f^{-1} as a vector $\vec{f}^{-1} = [\bar{q}_1, \bar{q}_2, \bar{q}_3...\bar{q}_{254}]^T$ and we form a constraint matrix A such that the n^{th} row of the matrix corresponds to the the n^{th} pixel in images p_a, p_b and p_c. Each row has a 1 in columns a and b, -1 in column c and zeros in all other columns. In the n^{th} row, a, b and c correspond to pixel values $p_a[n]$, $p_b[n]$ and $p_c[n]$ respectively. The least squares solution of the homogeneous equation: $A\vec{f}^{-1} = 0$ is then obtained by obtaining the SVD of $A = U\Sigma V$ and using the column of V corresponding to the smallest singular value in Σ.

Solving for f^{-1} by this method assumes that the error: $\epsilon = q_a + \bar{q}_b - \bar{q}_c$ has zero mean. Without noise, clipping at 255 can create a problem by biasing the distribution of the measured pixel values. With camera noise, this bias becomes very significant in pixel ranges near both clipping points: 0 and 255. Also, as with all least squares methods, outlier points can significantly perturb the solution.

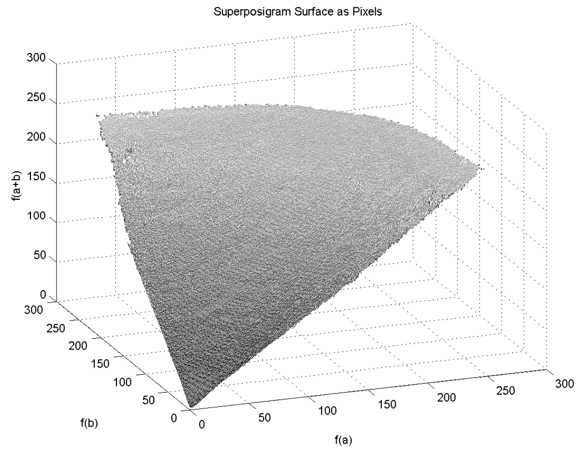

Fig. 3: Surface of the superposigraph (slenderized superposigram, where the slenderization occurs along the z-axis using maximum likelihood). The surface is given by $f(a)+f(b)=\bar{f}(a+b)$ with $f(a)$ in the x-axis, $f(b)$ in the y-axis, $\bar{f}(a+b)$ in the z-axis.

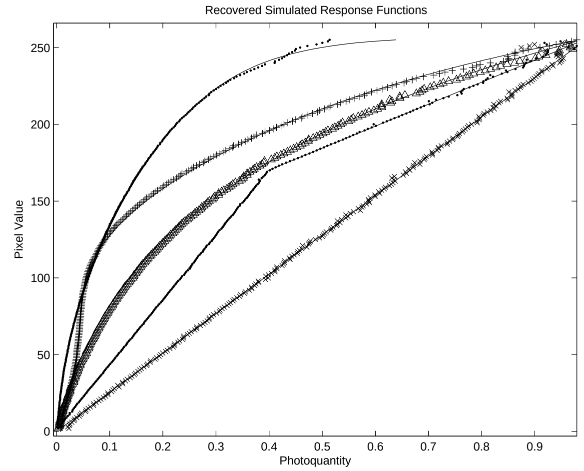

Fig. 4: Plots of five different recovered response functions, together with ground truth, using synthetically generated data. The proposed algorithm was then used to recover 256 discrete points (plotted as various shaped points). Note that the recovered points fall virtually on top of the original functions (plotted as solid lines).

With these considerations, the method is improved by robustly estimating $f(\bar{q}_c)$ by generating a histogram of the measured pixel values of c for each additive combination of a and b. By assuming that the normalized histogram is a reasonable approximation of the actual probability distribution of c, we can use the peak of this histogram $\hat{c}_{a+b}$ as our best estimate of $f^{-1}(f(\bar{q}_a+\bar{q}_b))$. Our minimization problem thus becomes:

$$e=\sum_{\forall pairs\{x,y\}} N_{\{x,y\}}\left(\bar{q}_x+\bar{q}_y-\bar{q}_{\hat{c}_{x+y}}\right)^2 \quad (6)$$

Where $N_{\{x,y\}}$ is the number of instances of $f^{-1}(a)+f^{-1}(b)=f^{-1}(c)$ in the dataset.

We call the collection of histograms for each additive combination of pixels a Superposigram (histogram of superpositions). For a digital camera with 256 pixel levels, its associated Superposigram will be a $256\times256\times256$ array, with the first two dimensions being the pixel values in image P_a and P_b respectively, and the third dimension containing the number of occurrences of each pixel value for each $\{a,b\}$ combination. The Superposigram representation is effective since we can easily compile information from multiple image sets by simply adding the Superposigrams produced by each set, thereby increasing the accuracy of our estimate of $f(\bar{q}_c)$. In our experiments, we have also smoothed each histogram with a Gaussian kernel to improve the estimate of the peak location in conditions where $\{a,b\}$ combinations are poorly represented.

It is revealing to plot the histogram peaks in the Superposigram as a surface in three dimensions (x,y,z). Here x and y represent the $\{p_a,p_b\}$ combination and z represents $f(\bar{q}_c)$ (the peak of the histogram). We call this *slenderized* version of the superposigram, a superposigraph. Since most cameras exhibit a nonlinear response, we expect this surface to be curved. See figure 3 for the Superposigram surface of a Nikon D1 digital camera.

4. TESTING THE NEW METHOD

To show that the method presented in this paper performs reliably, random synthetic lightspace data was generated. To this data, a C^1 continuous function was applied to the lightspace data, and the result quantized into 256 imagespace/pixel values. Following this procedure, Gaussian noise of standard deviation 10 was added to the pixel values. The singular value decomposition method was then used to recover the function using the described constraint matrix where the equations were generated using the slenderized superposigram. The results of this procedure on the synthetic data are shown in figure 4. As can be seen in the figure, the high quantity of noise added has significantly affected the recovered response functions, however it demonstrates the stability of the algorithm. When more reasonable noise levels are used, such as a standard deviation of 3, the results are very accurate.

The algorithm was then tested on actual image data captured by a Nikon D1 digital camera. The superposigraph shown in figure 3 was generated from the images presented in figure 2. To create more data which would span the entire range of the camera response function, several exposures $f(q_a)$, $f(q_b)$ and $f(q_a+q_b)$ were used. Though it is possible to generate good response functions from one set of three images, using more image sets did in fact produce a more reliable response function. From the superposigraph that was generated from the multiple image sets (we used 10 in total), the response function for camera was determined.

To visualize how well the recovered camera estimated response function did at linearization, the inverse camera response function may be applied to the superposigraph. The linearized lightspace surface arising from $f^{-1}(f(q_a))+f^{-1}(f(q_b))=\hat{f}^{-1}(f(q_a+q_b))$, is shown in figure 5 [1].

5. CONFIRMING THE CORRECTNESS OF THE CAMERA RESPONSE FUNCTION

To test the accuracy of the recovered camera response functions, two tests were devised. The following sections describe these tests, after which our results are presented.

[1]If it were not for the nonlinear nature of the response function, f, the superposigram would be a plane plus noise in three dimensional space. With the recovered camera response function, we can linearize the images by converting them from imagespace, $f(q)$, into lightspace, q.

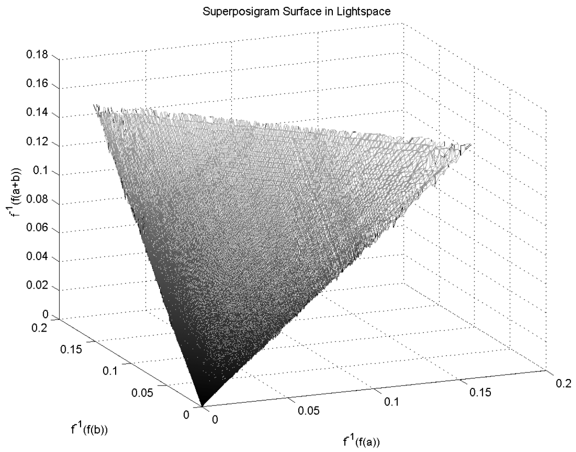

Fig. 5: We can visualize the efficacy of our method by looking at how well the inverse response function planarized the superposigram. The surface of $f^{-1}(f(a)) + f^{-1}(f(b)) = \hat{f}^{-1}(f(a+b))$ with $f^{-1}(f(a))$ in the x-axis, $f^{-1}(f(b))$ in the y-axis, and $\hat{f}^{-1}(f(a+b))$ in the z-axis.

5.1. Confirming the correctness of the camera response function by homogeneity

The first measure described is termed a homogeneity-test of the camera response function (regardless of how it was obtained). The homogeneity-test requires two differently (by a scalar factor of k) exposed pictures, $f(q)$ and $f(kq)$, of the same subject matter.

To conduct the test, the dark image $f(q)$ is lightened, and then tested to see how close it is (in the mean squared error sense) to $f(kq)$. The mean-squared difference is termed the *homogeneity error*. To lighten the dark image, it is first converted from imagespace f to lightspace, q, by computing $f^{-1}(f(q))$. Then the photoquantities q bre multiplied by a constant value, k. Finally, we convert it back to imagespace, by applying f. Alternatively we could apply f^{-1} to both images and multiply the first by k and compare them in lightspace (as photoquantities).

5.2. Confirming the correctness of the camera response function by superposition

Another test of a camera response function termed the *superposition-test*, requires three pictures $p_a = f(q_a), p_b = f(q_b)$ and $p_c = f(q_{a+b})$. The inverse response function is applied to p_a and p_b and the resulting photoquantities q_a and q_b are added. We now compare this sum (in either imagespace or lightspace) with p_c (or q_c). The resulting mean squared difference is the *superposition error*.

5.3. Comparing homogeneity and superposition errors in response functions found by each of various methods

The results of comparison of homogeneity and superposition errors in response functions found by various methods (including previous published work) are compared in Table 1.

6. CONCLUSION

In this paper we showed how an unknown nonlinear camera response function can be recovered using the superposition property of light. As with earlier work using comparagrams where improved results were obtained through slenderization of the comparagram, we found in this paper that improved results were obtained through slenderization of the superposigram. The new method was tested on various synthetic sequences with synthetic noise, to prove ground-truth, as well as on actual data, to show that it works in real world images as well.

Method used to determine the response function	Superposition Error	Homogeneity Error
Homogeneity with parametric solution (Previous Work [8][4])	8.8096	9.9827
Homogeneity, direct solution	8.6751	9.4011
Superposition, direct solution	8.5450	9.5361

Table 1: This table shows the per-pixel errors observed in using lookup tables arising from several methods of calculating f and f^{-1}. The leftmost column denotes the method used to determine the response function. The middle column denotes how well the resulting response function superimposes images, based on testing the candidate response function on pictures of subject matter taken under different lighting positions. The rightmost column denotes how well the resulting response function amplitude-scales images, and was determined based on using differently exposed pictures of the same subject matter. The entries in the rightmost two columns are mean squared error divided by the number of pixels in an image.

7. REFERENCES

[1] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, NJ, 1998.

[2] O. Faugeras, *Three-Dimensional Computer Vision - A Geometric Viewpoint*, M.I.T. Press, Cambridge, MA, 1993.

[3] F. M. Candocia, “A least squares approach for the joint domain and range registration of images,” *IEEE ICASSP*, vol. IV, pp. 3237–3240, May 13-17 2002, avail. at http://iul.eng.fiu.edu/candocia/Publications/Publications.htm.

[4] A. Barros and F. M. Candocia, “Image registration in range using a constrained piecewise linear model,” *IEEE ICASSP*, vol. IV, pp. 3345–3348, May 13-17 2002, avail. at http://iul.eng.fiu.edu/candocia/Publications/Publications.htm.

[5] S. Mann and R. Mann, “Quantigraphic imaging: Estimating the camera response and exposures from differently exposed images,” *CVPR*, pp. 842–849, December 11-13 2001.

[6] Martin Bichsel and Krystyna W. Ohnesorge, “How to measure a camera’s response curve from scratch,” Tech. Rep. ifi-93.19, 1, 1993.

[7] F. M. Candocia, “Synthesizing a panoramic scene with a common exposure via the simultaneous registration of images,” *FCRAR*, May 23-24 2002, avail. at http://iul.eng.fiu.edu/candocia/Publications/Publications.htm.

[8] S. Mann, “Comparametric equations with practical applications in quantigraphic image processing,” *IEEE Trans. Image Proc.*, vol. 9, no. 8, pp. 1389–1406, August 2000, ISSN 1057-7149.

[9] Michael D. Grossberg and Shree K. Nayar, “What can be known about the radiometric response function from images ?,” *Proc. of European Conference on Computer Vision (ECCV) Copenhagen, May*, 2002.

[10] C. Manders S. Mann and J. Fung, “Painting with looks: Photographic images from video using quantimetric processing,” *ACM Multimedia 2002*, pp. 117–126, 2002.

[11] Michael D. Grossberg and Shree K. Nayar, “What is the space of camera response functions,” *In Proc. IEEE Computer Vision and Pattern Recognition (CVPR), Wisconsin, June*, 2003.