CAMERA RESPONSE FUNCTION RECOVERY FROM AUTO-EXPOSURE CAMERAS

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 applicable to auto-exposure cameras is presented. The method uses the superposition property of light to solve for the response function directly using superposition constraints imposed by using different combinations of two (or more) lights that illuminate the same subject matter in varying proportions. An iterative optimization is utilised to simultaneously recover the exposure ratios of the images and the camera response function. Previously published multiple exposure methods that simultaneously estimate exposure ratio and response function suffer from a fundamental ambiguity. The use of the proposed superposition constraints solves this problem. We introduce a simple method for combining both superposition and homogeneity (from multiple exposure techniques) to accurately recover the response function from auto-exposure cameras.

Index Terms— Cameras, Image sensors, Calibration, Image color analysis, Lighting, Image registration

1. INTRODUCTION

Camera response functions map the actual quantity of light impinging on each element of a sensor array (or each region of a film plane) 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: **Linearity**: $f(ax + by) = af(x) + bf(y)$.

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 the camera response function using different exposures of the same subject matter [1][2][3][4]. These methods work well in situations where one has some control over the exposure of the camera. In situations where the exposure ratios between the images is not known, or cannot be set (as is the case with many automatic exposure cameras that have no manual override), the lack of a unique solution for the exposure ratio and response function makes a simple and accurate recovery difficult. Grossberg and Nayar [5] call this "The exponential ambiguity", and prove that it is the only ambiguity in recovering the response function and exposure ratios together.

In [5], Grossberg and Nayar present a theoretical method for recovering the exposure ratio without solving for the response function as a means of breaking this ambiguity. Their method estimates the exposure ratio from the derivative the camera response function at pixel value zero. Unfortunately, due to the methods heavy dependence on high SNR at small intensity values, it is difficult to apply in practice.

The exponential ambiguity can be dealt with by making assumptions about the form of the response function. For example, Mitsunaga and Nayar [6] used a polynomial form for the response function. Barros and Candocia [2] indirectly make assumptions about the shape of the response function by assuming that the comparametric function of the response function is piecewise linear. Mitsunaga and Nayar were then able to obtain the polynomial approximated response function as well as the exposure ratios, by alternately solving for the response function and the exposure ratios in an iterative scheme.

Tsin et al [7] solve for the camera response function from multiple exposure data while allowing for small deviations from the nominal exposure due to effects like automatic white balance. Their optimization avoids trivial solutions caused by the exponential ambiguity by punishing large deviations in exposure. Note that this technique would not work for completely unknown exposures.

Shafique and Shah [8] proposed a method that uses varying illumination rather than multiple exposures, however they make the strong assumption that the camera response function is a gamma curve.

The method presented in this paper is easy to use, assumes only that the camera response function is semi-monotonic and does not require knowledge of or control of the exposure level (i.e. it works with manual *or* automatic exposure cameras). It avoids the exponential ambiguity by using the superposition constraints created using two separately controllable lights to illuminate a scene. The method is an extension of the technique presented in [9] that uses superposition constraints to break the comperiodic ambiguity [1] The method presented in [9] requires that the camera's exposure sensitivity does not change as the scene lighting changes and thus does not work with automatic exposure cameras in which there is no explicit control over camera exposure setting.

[1]Given fixed exposure ratios, different camera response functions give rise to the same camera response function. Also known as the "self similar ambiguity" in [6][5].

The acquisition of a set of images suitable for use with our method is easy using controllable lights. However the technique is also useful in many situations where direct access to the illumination sources is impossible. 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 (i.e. as when a surveillance camera observes a scene in which lights get switched on and off by cleaning staff or by automatic timers in various permutations). In some situations, the method can also work due to flickering of various lights (i.e. lights on different phases of an electric supply, or by the fact some lights like tungsten lamps lag behind fluorescent lamp flicker).

Due to the fact that it is difficult to capture pictures that have pixel values that are well distributed over the range of the camera, often only a partial recovery of the camera response function is possible using a single set of three images (i.e. with one light, the other light, and both lights, as in Fig 1) of differently illuminated images. A secondary method is thus presented to complete the recovery by using the partially recovered response function to add the homogeneity constraints provided by one or more other sets of images taken at arbitrary exposure values. This secondary method uses the constraint covariance matrix as a natural structure to aggregate the camera response function constraints obtained using both homogeneity and superposition.

1.1. Solving for the Inverse Camera Response Function By Superposition

The standard superposimetric [9] procedure is used: in a dark environment, two distinct light sources are set up. Three pictures are taken, 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 *photoquantity* [1][10].

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

$$f^{-1}(f(q_a)) + f^{-1}(f(q_b)) = f^{-1}(f(q_c)) + \epsilon_Q, \qquad (1)$$

where ϵ_Q is the mean error due to quantization and other noise processes. We can thus solve for f^{-1}, i.e. the mapping from pixel value p_x to quantized photoquantities $\bar{q}_x$ that are optimal in the least-squares sense by minimizing the following equation:

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

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 1.

Since most cameras output a finite range of 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. As a result of clipping at the ends of the camera's range, we do not try to solve for $f^{-1}(0)$ or $f^{-1}(255)$. Instead we can solve for the quantization points $\hat{q}_{0,1}$ and $\hat{q}_{254,255}$, where $\hat{q}_{a,a+1}$ is the quantization point separating pixel value a and $a+1$. This allows us to conclude that if we measure a pixel value of 0 with the camera, a quantity of light below $\hat{q}_{0,1}$ was measured. Similarly, a pixel value of 255 represents a photoquantity greater than $\hat{q}_{254,255}$. A method for solving for these thresholds is not presented, but can be accomplished by using a multi-exposure technique after one has accurately determined a portion of the camera response function.

Fig. 1: One of the image sets used. Leftmost: Picture with light A turned on. Middle: Picture with light B turned on. Rightmost: Picture with light A and B turned on together; note the double shadows.

We thus define f^{-1} as the mapping from pixel values $(1, 2, 3...254)$ to the quantized photoquantities $(\bar{q}_1, \bar{q}_2, \bar{q}_3...\bar{q}_{254})$. We can now write equation 2 more simply as:

$$e = \sum_{\substack{\forall n, P_a, P_b \\ P_c \neq 0, 255}} \left(\bar{q}_{p_a[n]} + \bar{q}_{p_b[n]} - \bar{q}_{p_c[n]} \right)^2 \qquad (3)$$

Equation (3) 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 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 $A = U\Sigma V^T$ 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: $\epsilon = q_a + \bar{q}_b - \bar{q}_c$ is Gaussian. 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.

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:

$$e = \sum_{\forall pairs\{x,y\}} N_{\{x,y\}} \left(\bar{q}_x + \bar{q}_y - \bar{q}_{\hat{c}_{x+y}} \right)^2 \qquad (4)$$

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

For a digital camera with 256 pixel levels, this collection of histograms can be expressed in a $256 \times 256 \times 256$ 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. This representation is effective since we can easily compile information from multiple

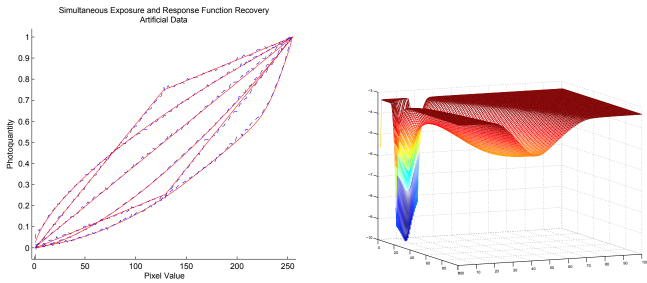

Fig. 2: The left figure shows a variety of response functions that were used to test the algorithm. Images with a predetermined exposure ratios were generated randomly with gaussian noise ($\sigma = 10$ pixel levels) added to simulate camera noise. The proposed algorithm was then used to recover the response function and the exposure ratios with the results shown on the figure in dashed lines. Note that the recovered points fall virtually on top of the original functions (plotted as solid lines). The right figure shows an example 3-D error surface parameterised over $k_1 = (0, 5)$, $k_2 = (0, 5)$. The figure shows the local minimum associated with the correct exposure ratio as well as the $k_1 + k_2 = 1$ singularity.

image sets by simply adding the collection of histograms produced by each set, thereby increasing the accuracy of our estimate of $f(\bar{q}_c)$. In order to improve the estimate of the peak location, the histograms are each smoothed with a Gaussian kernel. This procedure is especially important when $\{a, b\}$ combinations are poorly represented.

2. SIMULTANEOUS ESTIMATION OF EXPOSURE RATIOS AND RESPONSE FUNCTION

To simultaneously solve for the exposure ratios and the camera response function, we form a matrix equation for the constraints on the response function paramaterized by k_1, k_2.

$$e = \sum_{\forall pairs\{x,y\}} N_{\{x,y\}} \left(k_1\bar{q}_x + k_2\bar{q}_y - \bar{q}_{\hat{c}_{x+y}}\right)^2 \tag{5}$$

Where k_1 is the exposure ratio images p_a and p_c, k_2 is the exposure ratio of images p_b and p_c. $N_{\{x,y\}}$ is the number of instances of $f^{-1}(a) + f^{-1}(b) = f^{-1}(c)$ in the data set.

We minimize over (k_1, k_2) and $\vec{f}^{-1} = [\bar{q}_1, \bar{q}_2, \bar{q}_3 ... \bar{q}_{254}]^T$.

By computing:

$$C_A\{i,:\} = [0^1, ..., 0^{\bar{q}_x-1}, N^{\bar{q}_x}_{\{x,y\}}, 0^{\bar{q}_x+1}, ..., 0^{255}] \tag{6}$$

$$C_B\{i,:\} = [0^1, ..., 0^{\bar{q}_y-1}, N^{\bar{q}_y}_{\{x,y\}}, 0^{\bar{q}_y+1}, ..., 0^{255}] \tag{7}$$

$$C_I\{i,:\} = [0^1, ..., 0^{\bar{q}_x-1}, -N^{\bar{q}_{\hat{c}_{x+y}}}_{\{x,y\}}, 0^{\bar{q}_x+1}, ..., 0^{255}] \tag{8}$$

$$C_{Mat} = {k_1}^2 C_A^T C_A + {k_2}^2 C_B^T C_B + C_I^T C_I + k_1 k_2 (C_A^T C_B + C_B^T C_A) + k_1(C_A^T C_I + C_I^T C_A) + k_2(C_B^T C_I + C_I^T C_B) \tag{9}$$

If we know k_1 and k_2, the least squares solution for f^{-1} can be easily obtained by using the singular value decomposition (SVD): $C_{Mat} = U\Lambda V^T$. f^{-1} will be the column of V corresponding to the smallest singular value in Λ.

To find k_1 and k_2 we perform a gradient descent linesearch optimization in the 2D space (k_1, k_2) using the smallest singular value of C_{Mat} as the error function:

$$error(k_1, k_2) = \Lambda(255, 255); \tag{10}$$

The optimization is made computationally efficient by pre-computing $C_A^T C_A, C_B^T C_B, C_I^T C_I, C_A^T C_B, C_B^T C_A, C_A^T C_I, C_I^T C_A, C_B^T C_I$ and $C_I^T C_B$, leaving only scalar multiplication and addition of 254x254 square matrices and an SVD of the result.

From experimentation using simulated data, it was found that this error surface was very well behaved, giving few pitfalls in the optimization. Figure 2 gives some samples of response function used in the simulation. The most noteworthy feature of the error surface that must be considered in all cases is the degenerate situation when $k_1 + k_2 = 1$. In this case, the second smallest singular value must be used for the error, to avoid obtaining the trivial constant solution to f^{-1}. A typical error surface is depicted in figure2. This shows both the $k_1 + k_2 = 1$ singularity as well as the well defined local minimum generated by the true solution for k_1 and k_2.

When there is no a priori knowledge of the exposure ratios, two starting points on either side of the $k_1 + k_2 = 1$ line must be used for the optimization. In most cases optimization works well without any tricks, however in some of our experiments it was necessary to put a penalty on negative k_1 and k_2.

In general because of the difficulty in capturing a single set of pictures that span the whole camera range, it is often difficult to achieve reliable results. This problem is more pronounced in cameras that have a high level of noise and low resolution. Typically the histograms of images have pixel values lumped in some part of the camera range.

Though the superposition method achieves very good results, when there is not enough data, the algorithm is difficult to stabilize[2] and can produce inaccurate results. For example, to add a smoothness term, can use the λ weighted constraints:

$$0 = \lambda(-\bar{q}_{n-2} + 16\bar{q}_{n-1} - 30\bar{q}_n + 16\bar{q}_{n+1} - \bar{q}_{n+2}) \tag{11}$$

The logical solution is to solve for only part of the response function. This partial response function can then be used to find the exposure difference between pairs of images, making it easy to add more constraints to the solution. By using the sample covariance matrix C_{Mat} as a structure to aggregate the data, it is easy use homogeneity and superposition constraints for the solution. The fixed size of the matrix ($254x254$), allows data to be continually added without increasing the memory requirements.

To solve for a partial response function, we can crop the Covariance matrix C_{Mat} to a pixel value range that is well represented in the images. $C_{part} = C_{Mat}(minVal : maxVal, minVal : maxVal)$, where $minVal$ and $maxVal$ are chosen from a well populated overlapping pixel value range from the superposition image triplet. For example, the triplet shown in figure1 used $minVal = 10$ and $maxVal = 140$ for the analysis. The results from the simultaneous recovery of exposure ratios and camera response function in this range is presented in figure3.

3. IMPROVING THE ESTIMATE USING HOMOGENEITY

By using the partially recovered response function (shown in the top right plot of figure 3), computing the exposure differences between

[2] usually with add-hoc regularization strategies

pairs of images of similar subject matter is a simple task. This enables us utilize homogeneity constraints without the problem of the exponential ambiguity.

By using the covariance matrix C_{Mat} to aggregate the constraints, we add the data obtained through homogeneity relations by: computing C_u and C_{Ku} for each corresponding pixel i in a pair of images taken of the same subject matter using different exposures.

$$C_u\{i,:\} = [0^1, ..., 0^{\bar{q}_u - 1}, 1^{\bar{q}_u}, 0^{\bar{q}_u + 1}, ..., 0^{255}] \tag{12}$$

$$C_{Ku}\{i,:\} = [0^1, ..., 0^{\bar{q}_{Ku} - 1}, 1^{\bar{q}_{Ku}}, 0^{\bar{q}_{Ku} + 1}, ..., 0^{255}] \tag{13}$$

By solving for the exposure difference κ between each pair of images using the recovered partial response function, we add the new data to the aggregate C_{Mat}:

$$C_{Mat} = C_{Mat} + (C_u - \kappa C_{Ku})^T (C_u - \kappa C_{Ku}) \tag{14}$$

Note that when we are adding to the C_{Mat} previously generated from the superposition data. Thus for a set images of the same subject matter we can add the constraint data from all possible pairs of images. Generally κ will be pair dependent. With an auto-exposure camera it is quite easy to generate a variety of exposures, for example: placing a small light in the center of the scene to "fool" the camera, i.e. since the camera slowly responds to this disturbance, a variety of exposures will occur. This effect also occurs naturally, for example when a surveillance camera monitors a scene and lights are switched on and off in a room, it takes time for the camera to adjust itself (automatic exposure) to the changes in light levels. Thus successive frames of video are used in order to obtain sets of images that differ only in exposure, to a particular lighting configuration. These changes in lighting, and in exposure (automatic gain control in the camera and possibly automatic white balance) work together to provide a rich space of superposition and homogeneity data. Results from applying this method to a Nikon D70 camera (used for its ability to provide ground truth data) are presented in figure 3. In this figure, the top left image describes the resulting C_{Mat} from the superposition constraints provided by the image sequence shown in figure 1. The top right plot shows the recovered partial response function from the C_{Mat} on the left. The bottom left image shows the aggregate C_{Mat} from the original superposition constraints together with the homogeneity constraints from a sequence of ten differently exposed images. The bottom right the plot shows the response function solution provided by the aggregate C_{Mat} overlaid on the partial response function solution. The plot also shows a comperiodic (wavy, ambiguous) solution obtained from using one image pair alone, illustrating success of the method in breaking the self-similar ambiguity.

4. CONCLUSION

A method of recovering an unknown response function of an automatic-exposure camera was presented. This method is particularly useful because many modern low cost cameras such as web cameras, as well as cameras found in camera phones and laptop computers, use automatic exposure mechanisms that provide no manual override or any means to manually adjust or even lock the exposure setting. Our method is based on the linearity properties of light: superposition and homogeneity, which can be used separately or together.

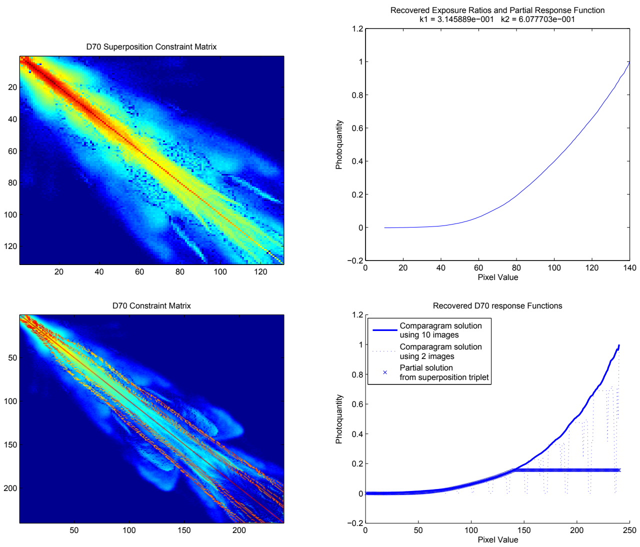

Fig. 3: Top Left: An images describing the resulting C_{Mat} from the superposigram constraints provided by the image sequence shown in Fig 1. Top Right: The recovered partial response function from the C_{Mat} on the left. Bottom Left: The aggregate C_{Mat} from the original superposition constraints together with the homogeneity constraints from a sequence of ten differently exposed images. Bottom Right: The plot shows the response function solution provided by the aggregate C_{Mat} overlaid on the partial response function solution. The plot also shows a the periodic solution obtained from using one image pair alone, illustrating success of the method in breaking the self-similar ambiguity.

5. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proc. IEEE Int. Conf. Image Processing*, 1999, vol. 3, pp. 159–163.

[5] Michael D. Grossberg and Shree K. Nayar, "Determining the camera response from images: What is knowable?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 11, pp. 1455–1467, 2003.

[6] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999, vol. 1, pp. 374–380.

[7] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of the ccd imaging process," in *in Proc. Int. Conf. Computer Vision*, 2001, vol. 1, pp. 480–487.

[8] K. Shafique and M. Shah, "Estimation of the radiometric response functions of a color camera from differently illuminated images," in *Proc. ICIP*, 2004, pp. 2339–2342.

[9] Corey Manders, Chris Aimone, and Steve Mann, "Camera response function recovery from different illuminations of identical subject matter.," in *Proc. ICIP*, 2004, pp. 2965–2968.

[10] 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.