Radiometric Self-Alignment of Image Sequences (CVPR '04)

S. J. Kim

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27517
sjkim@cs.unc.edu

M. Pollefeys

Department of Computer Science
University of North Carolina
Chapel Hill, NC 27517
marc@cs.unc.edu

Abstract

Color values in an image are related to image irradiance by a nonlinear function called radiometric response function. Since this function depends on the aperture and shutter speed, image intensity of a same object may vary during the acquisition of an image sequence due to auto exposure feature of the camera. While this is desirable to make optimal use of the limited dynamic range of most cameras, this causes problems for a number of applications in computer vision. In this paper, we propose a method for estimating radiometric response function and apply it to radiometrically align images so that the color values are consistent for all images of a sequence. Our approach computes the response function, exposure and white balance changes between images (up to some ambiguity) for moving camera without any prior knowledge about exposures. We show the performance of our algorithm by estimating the response function from synthetic images and also from real world data, using it to radiometrically align the images.

1 Introduction

Even with the important progress in the area of digital cameras and camcorders, cameras still can only capture limited dynamic range of a scene. Most cameras compress the dynamic range of the scene, introducing nonlinearity between recorded brightness intensity and the scene radiance. This mapping is called the radiometric response function. This nonlinearity of the radiometric response function may cause problems in many computer vision algorithms where observed intensity values are assumed to directly reflect the scene radiance. Recovering the radiometric response function is important especially to those algorithms that explicitly use scene radiance measurements such as color constancy, construction of high dynamic range images, photometric stereo, shape from shading, recovery of BRDF from images, and estimation of reflectance and illumination from shape and brightness [5]. It is also a requirement to be able to apply texture images recorded with different exposures on a 3D model.

1.1 Previous Works

A number of algorithms for estimating the radiometric response function have been introduced [3, 4, 5, 8, 9, 10, 11]. Mann and Picard [9] estimated the response curve assuming that the response is a gamma curve and they know the exposure ratios between images. Debevec and Malik [3] introduced a non-parametric method for the recovery by imposing smoothness constraint. The exact exposure values with which the pictures are taken are necessary for their method. Mitsunaga and Nayar [10] assumed the response curve to be a polynomial and estimated it iteratively with rough exposure ratio estimates. All these methods require a number of differently exposed images of static scene with fixed camera.

Grossberg and Nayar [4] explained ambiguities associated with the problem and introduced a response curve estimation method by recovering brightness transfer function from histograms. While their method does not require a fully static scene, a fixed camera is still necessary. Tsin, Ramesh, and Kanade [11] introduced a non-parametric method which estimates both the response and exposure with a statistical model of the measurement errors. Mann [8] proposed another algorithm which also estimates both the response and exposure by iterative method.

Even though majority of mentioned methods work fine in most cases, there are common disadvantages in most of these algorithms. One is that they require prior knowledge of exposure ratios which is not easy to know beforehand. The other is that images have to be taken with a fixed camera. Mann [8] addressed the problem of estimating the response with non-static camera, but it covers only pure rotation and zoom. Table 1 shows the summary of various radiometric response function recovery methods.

Method	Model	Exposures	Camera	Scene
[9]	P	R	Fixed	Static
[3]	NP	R	Fixed	Static
[10]	P	R	Fixed	Static
[4]	P	R	Fixed	Non-Static
[11]	NP	NR	Fixed	Static
[8]	NP	NR	Rotation Allowed	Static
Our Method	P	NR	Free Movement	Non-static

Table 1: Comparison of various methods (P:Parametric, NP:Non-parametric, R:Required, NR:Not Required)

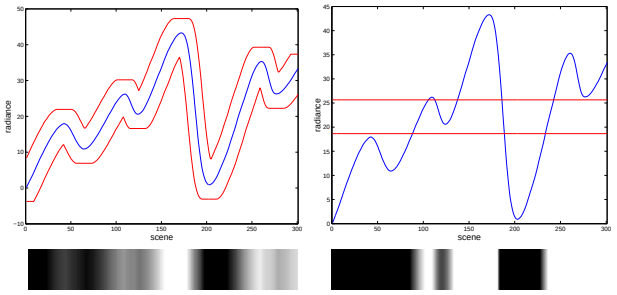

Figure 1: Advantage of Auto Exposure : Synthetic high dynamic range signal observed by 1D moving camera with (left) and without(right) auto exposure. (Top) Signal and the actual dynamic range (Bottom) Resulting mosaic images

1.2 Goal of the paper

In this paper, we propose a radiometric response function estimation algorithm that does not require prior knowledge of exposures and allows free movement of camera. Our primary interest of application is in radiometric alignment of image sequences. For example, assume there is a scene where brightly lit area and a dim area coexist. Since the dynamic range of the scene exceeds that of a camera, we have to expose the camera according to the area of interest. If such a scene was recorded with a video camera, there would be intensity variation between images of the same object due to the auto exposure function of the camera. We would like to correct these intensity variations due to different exposures so that we can take advantage of the auto exposure functionality of cameras. Auto exposure allows us to capture global high dynamic range from local low dynamic range and gives flexibility as to not having to worry about finding the right 8-bit range (Fig.1). Our approach is essentially different from other texture correction methods such as the method in [1] where color transform was adapted for correcting color discontinuity and the method in [2] where a common lighting between textures were derived to relight textures.

To compute response function, we propose a series of methods to estimate brightness transfer function between images so that input images for our method does not have to be static. With the information of brightness transfer between images, we estimate response function by modifying the Emperical Model of Response (EMoR) introduced in [5]. For the purpose of radiometric alignment in which the response function can be estimated up to exponential ambiguity, our method does not require any prior knowledge of exposure values with which the images were taken. Even if the response function has to be found without the ambiguity, our method requires far less prior information than other methods.

2 Our Algorithm

We first start with defining the behavior of the radiometric response function.

Basic relationship between the image irradiance and the image intensity can be stated as follows :

$$I_{pq} = f(E_p K_q) \tag{1}$$

where E_p is the image irradiance for a specific pixel, K_q is the exposure with which the picture was taken, I_{pq} is the observed image intensity, and the f is the radiometric response function. Because for cameras increasing irradiance will result in increasing (or constant) image intensity, the response function is (semi-)monotonic and can be inverted. Taking the inverse and the logarithm, Eq. (1) can be rewritten as follows:

$$\ln f^{-1}(I_{pq}) = \ln E_p + \ln K_q$$

or, with $g(I) = \ln f^{-1}(I)$ and $k_q = \ln K_q$,

$$g(I_{pq}) = \ln E_p + k_q \tag{2}$$

Assume that we have two images of the same scene taken with different exposure (K_1, K_2). Then the following relationship between the image intensities is obtained:

$$g(I_{p2}) - g(I_{p1}) - (k_2 - k_1) = 0 \tag{3}$$

Note that this equation is not only valid for a static camera, but also for a moving camera as long as the pixels in Eq. (3) are corresponding pixels and the absurd surface patch is lambertian.

2.1 Correspondence

One of the more challenging problems we have to deal with is the computation of correspondences. Ideally, only a limited number of points are required to estimate the radiometric response curve and the exposure ratio. However, because

Figure 2: For every pixel of the left image the same pixel in the right image contains the color value found at the corresponding pixel in another image of the sequence.

of a certain number of limitations it is best to estimate correspondences for a larger number of points. First, we want corresponding points to cover as much intensity values as possible (and this for each R, G and B channel separately). Then, because we want to deal with a moving camera, we have to deal with the fact that not all pixels correspond to Lambertian surfaces so that we can not always expect the radiance to be constant over varying viewing directions (which was not a problem for static or purely rotating cameras). In addition, matching between images recorded with different exposure settings will in itself be hard and we have to expect a significant number of wrong matches. Therefore, it is important to obtain as much redundancy as possible so that a robust approach can later be used to estimate the desired camera properties. The approach we follow in this paper consists of first estimating the epipolar geometry for each pair of consecutive images (for video keyframes would be selected so that the estimation of the epipolar geometry would be stable) using tracked or matched features, followed by stereo matching. To avoid problems with intensity changes it is important to use zero-mean normalized cross-correlation. A possible alternative might consists of using optical flow, although this is also complicated by intensity variations. While we do not explicitly deal with independent motions in the scene, our stereo algorithm combined with our robust joint histogram estimation approach will handle those as outliers. In Fig. 2, an example is shown of the correspondences that can be obtained automatically.

2.2 Joint Histogram and Brightness Transfer Function

For a pair of images, all the information relevant to our problem is contained in the pair of intensity values of corresponding points. As suggested by Mann [8] these can all be collected in a two-variable joint histogram which he calls *comparagram*. For a pair of intensity values (I_{p1}, I_{p2}), the corresponding value in the joint histogram $J(I_{p1}, I_{p2})$ indicates how many pixels the intensity value changes from I_{p1} to I_{p2}.

As noted in [4], a function should ideally relate the intensity values between the two images. From Eq. (3), one immediately obtains

$$I_{p2} = T(I_{p1}) := g^{-1}(g(I_{p1}) + \Delta k)) \ . \tag{4}$$

with $\Delta k = k_2 - k_1$. The function T is called the *brightness transfer function* (BTF). It was shown in [4] that under reasonable assumptions for g, T is monotonically increasing, $T(0) = 0$ and if $\Delta k > 0$, then $I \leq T(I)$. Inversely, if $k < 0$ then $I \geq T(I)$. Ideally, making abstraction of noise and discretisation, if $I_{p2} \neq T(I_{p1})$, then we should have $J(I_{p1}, I_{p2}) = 0$. However, real joint histograms are quite different due to image noise, mismatches, view-dependent effect and a non-uniform histogram (Fig. 3). In the presence of large number of outliers, least square solutions for response function as have been used by others are not viable. We propose to use the following function as an approximation for the likelihood of the BTF passing through a pixel of the joint histogram.

$$P(T(I_1) = I_2|\bar{J}) = (G(0,\sigma) * \bar{J})(I_1, I_2) + P_o \tag{5}$$

where $G(0,\sigma)*$ represent the convolution with a zero-mean Gaussian with standard deviation σ to take image noise into account and P_0 is a term that represents the probability for $T(I_1) = I_2$ independent of the joint histogram. This term is necessary to be able to deal with the possibility of having the BTF pass through zeros in the joint histogram which could be necessary if for some intensity values no correct correspondence was obtained. Based on these assumptions the most probable solution is the BTF that maximizes

$$\ln P(T|\bar{J}) = \iint J_T(I_1, I_2) ln P(I_1 = T(I_2)|\bar{J}) dI_1 dI_2 \tag{6}$$

with $J_T(I_1, I_2)$ a function that is one where $I_2 = T(I_1)$ and zero otherwise. Using dynamic programming it is possible to compute the BTF that maximizes Eq. (6) under the constraints discussed above, i.e. semi-monotonicity, $T(0) = 0$, $T(255) = 255$ and $T(I) \geq I$ or $T(I) \leq I$ for all I.

2.3 Emperical Model of Response

With computed BTFs, we now estimate the radiometric response function using the low parameter Emperical Model of Response(EMoR) introduced by Grossberg and Nayar in [5]. They combined theoretical space of response function and database of real world camera response functions (DoRF) to create the EMoR which is a Mth order approximation :

$$f(E) = f_0(E) + \sum_{n=1}^{M} c_n h_n(E), \tag{7}$$

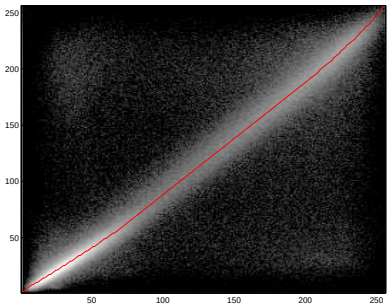

Figure 3: An example of a joint histogram with estimated BTF overlaid on it.

where h_ns are basis functions found by using PCA to DoRF and f_0 is the mean function.

In this paper, we are interested in log space to separate the exposure term from the irradiance(Eq. (2)). Eq. (7) becomes :

$$g(I) = g_0(I) + \sum_{n=1}^{M} c_n h'_n(I), \tag{8}$$

where $g(I) = \ln f^{-1}(I)$ and h'_ns are basis functions for log inverse response function of the database. The h'_ns are found by applying PCA to the log space of DoRF. One thing to notice is that elements of the first column and the first row of the covariance matrix of DoRF in log space are -∞ since data are normalized form zero to one. So, we remove the first column and the first row from the matrix for PCA. Therefore, intensity range of basis functions(h_n') are 1 to 255 instead of starting from 0. Fig. 4 shows first four basis functions of the log space of DoRF and the cumulative energy occupied by first 15 basis. First three eigenvalues explain more than 99.6%, which suggest that the EMoR model represents the log space of response function very well.

2.4 Radiometric Response Function Estimation

We estimate the response function and log exposure differences between images by using the computed BTFs and combining Eq. (3) and Eq. (8).

$$g_0(T_{ij}(I)) - g_0(I) + \sum_{n=1}^{M} c_n(h'_n(T_{ij}(I)) - h'_n(I)) - k_{ji} = 0 \tag{9}$$

where $k_{ji} = k_j - k_i$ and $T_{ij}()$ is the brightness transfer function from an image at log exposure k_i to an image at log exposure k_j.

Adopting the simplifying assumption that the effect of white-balance corresponds to changing the exposure independently for each color band, the unknowns of the equations are coefficients c_n's and exposure differences k_{ji}'s for

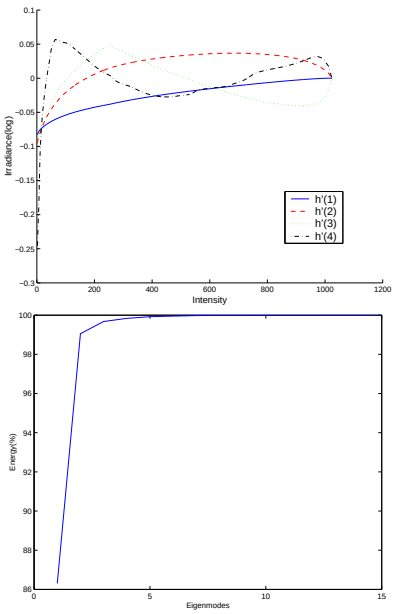

Figure 4: (Left) First four basis of DoRF(log space), (Right) Cumulative energy occupied by first 15 basis

each different color channels(R,G,B).

For each image pair ($1 \le i \le N-1$, N : number of images) and each color channel($l \in \{R,G,B\}$ or $\{1,2,3\}$), we build following matrices $A_i^l(255 \times (M+3 \times (N-1)))$ and $b_i^l(255 \times 1)$ (Eq. (10)(11)).

$$A_i^l(y,x) = h_x(T_{i,i+1}^l(y)) - h_x(y); 1 \le y \le 255,\ 1 \le x \le M$$

$$A_i^l(y,x) = -1; 1 \le y \le 255,\ x = M + (N-1) \times (l-1) + i$$

$$A_i^l(y,x) = 0; \text{elsewhere} \tag{10}$$

$$b_i^l(y) = g_0(y) - g_0(T_{i,i+1}^l(y)); \quad 1 \le y \le 255 \tag{11}$$

Since the response function will typically have a steep slope near I_{max} and I_{min}, we should expect that the response function will be less smooth and will fit the data more poorly near these extremes[3]. To reflect this to our algorithm, each column of A_i^l and b_i^l are weighted with a Gaussian mask with the mean being 128 and the standard deviation is chosen empirically.

To deal with discretization problem, we also compute BTFs in the opposite direction(i+1 to i), build matrices $A_i^{l'}$ and $b_i^{l'}$ which is similar to A_i^l and b_i^l except that $T_{i,i+1}$ is changed to $T_{i+1,i}$ and $A_i(y,x) = -1$ is changed to $A_i(y,x) = 1$ in (Eq. (10)).

After all the matrices are built, we solve for the coefficients of the model and the exposure differences(x) at once in least squares sense($Ax = b$) by combining all the matrices to form A and b as in Eq. (12) where each A^l and b^l are formed by combining A_i^l and b_i^l for all image pairs.

$$A = \begin{bmatrix} A^R \\ A^{R'} \\ A^G \\ A^{G'} \\ A^B \\ A^{B'} \end{bmatrix} \quad b = \begin{bmatrix} b^R \\ b^{R'} \\ b^G \\ b^G \\ b^B \\ b^{B'} \end{bmatrix} \tag{12}$$

$$x = \left[c_1, \ldots, c_M, k_{12}^R, \ldots, k_{12}^G, \ldots, k_{12}^B, \ldots, k_{n-1n}^B\right]^T \tag{13}$$

The least square solution x of $Ax = b$ at this point will suffer from the exponential ambiguity. Exponential ambiguity means that if g and k are solution to the Eq. (3) then so are αg and αk. Simply put, there can be many response functions and exposure differences that satisfy $Ax = b$ as long as they have the same scale factor. As stated in [4], it is impossible to recover g and k simultaneously from BTF alone, without making assumptions on g or k.

To resolve the problem of this ambiguity, we gave constraints to the equation by setting the value of initial exposure differences k_{12}^R, k_{12}^G, and k_{12}^B with a number. This serves as fixing the scale of the log inverse response function. For many applications including high dynamic range image construction and the texture alignment application which is the primary interest of this paper, choice of three values is not critical which is an advantage over other methods which require exact or rough estimate exposure values. In our case, initial values were chosen so that $g(128) = ln(0.5)$. If we needed the accurate response function, we would need to know three exposure differences but it is still far less than other methods.

3 Experiments

3.1 Static Camera - Synthetic Data

Even though our algorithm has been developed for application to moving cameras, we first experimented with static data to validate the performance of our method in estimating the response function. Also in this case, our modification should provide additional robustness. We constructed synthetic images given an image, a log inverse response function from the DoRF, and exposures with Eq. (3)(Fig. 5). A total of 20 sets of images were generated from 20 different response function.

Response functions were then estimated from these sets. An example of an inverse response function used (gamma curve,γ = 2.2) and its estimate are shown in Fig. 6. Note the effect of different initial exposure values on the estimation. Since we know the exposure values in this case, we can extract the response curve without exponential ambiguity as can be seen from the figure.

Figure 5: (Top) Examples of synthetic images (Bottom) Images aligned to the last image using the estimated response function

Another experiment for evaluating our algorithm was to compute the average intensity difference between aligned images. We can align images in the same set in regards to intensity with the estimated response function by using Eq. (3) (Fig. 5). The intensity difference between these aligned images was calculated to evaluate the algorithm. Our algorithm resulted in RMS intensity difference(ϵ) of 0.77 per pixel in our 20 image sets :

$$\epsilon = \sqrt{\frac{1}{n_x \times n_y} \sum_x^{n_x} \sum_y^{n_y} \sum_c^3 (I_{org}(x,y,c) - I_{est}(x,y,c))^2}$$

3.2 Moving Camera - Real Data

To test the algorithm with non-static images, a sequence of images of a tree was taken with a digital camera(Sony DSC-F717). Total of 19 images were taken, with the first image having the highest exposure since it is in a dark region (shadow). The first row of Fig. 8 shows few images of the sequence and the color discrepancy is easily seen. As mentioned in Sect. 2.2, joint histogram in our case may be quite noisy as shown in Fig. 3 which is one of the joint histograms in this tree sequence. Estimating the response function by least squares like most of the previous methods do not work well in this case due to outliers. We could not get a good estimate of the response from this sequence using the original EMoR method([5]). However by using our method, we were able to get a good estimate of the response function of this scene as shown in Fig. 7.

After estimating the response function, we aligned the textures of images in the sequence with the computed response function and exposure differences. The second

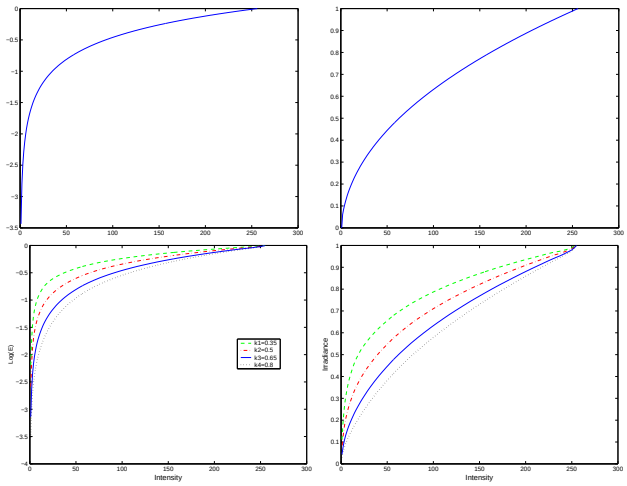

Figure 6: (Top) An Inverse response function($ln(f^{-1})$(left) and f^{-1}(right)) (Bottom) Effect of initial K value on estimation($ln(f^{-1})$(left) and f^{-1}(right)). If initial K value is known(0.65 in this case), we can estimate the response function without ambiguity

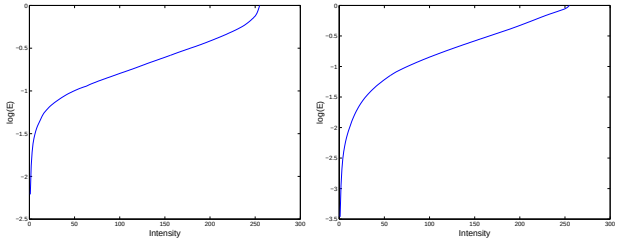

Figure 7: Estimated log inverse response function from the tree sequence(left,Sony DSC-F717) and the face sequence(right, Cannon GL2)

through the third row of Fig. 8 show examples where images in the set are radiometrically aligned to the first(brightest) and the last image(darkest) respectively. We can easily see improved color constancy from the example.

Finally we applied our method to normalize texture images recorded with different exposure on a 3D model. Fig. 9 shows some sample images from the sequence. Note the difference in the color of the face due to exposure changes. The 3D model which is texture mapped with these images shows color discrepancy as can be seen on the bottom left in Fig. 9. We computed response function(Fig. 7) from this sequence which contained a total of 12 images and normalized each image in the sequence to the brightest image. The resulting textured 3D model is shown on the bottom right in Fig. 9. Some of the remaining discrepancies (mostly visible on the nose)are due to view-dependent highlights which are not compensated for by our approach.

4 Conclusion

Radiometric response function defines relationship between image irradiance and image intensity. In this paper, we have proposed a method for computing radiometric response function of cameras. One advantage of our response function estimation method over previous methods is that non-static camera can be used in non-static scenes. We compute the response function up to exponential ambiguity where prior knowledge of exposures are not required, which is another advantage of our algorithm. Even if we compute the function without the ambiguity, we only need one exposure difference of an image pair out of the whole sequence which is far less than most algorithms require.

Our response function estimation was proven accurate which was shown by experiments with static images. If the initial exposure difference is known, we can estimate the function without exponential ambiguity. We also showed response function estimation from non-static images and normalized images with the estimated response function which was the primary application of interest of the paper. Normalized tree sequence and the normalized face model showed much improvement in color constancy.

In this paper, color changes were described by the response function and exposure changes in each channel independently. In the future, we would like to extend our method to allow for cross-talk between the channels to deal with the correlation between color channels([1]). We also want to find out the effect of vignetting on the images and adopt our method to it accordingly. Finally, we also plan to further explore high dynamic range texture and video generations([3], [7], [6]).

References

[1] A. Agathos, R.B. Fisher, ”Colour Texture Fusion of Multiple Range Images”, Proc. Fourth International Conference on 3-D Digital Imaging and Modeling, Alberta, Cananda, October 2003

[2] E. Beauchesne, S. Roy, “Automatic Relighting of Overlapping Textures of a 3D Model”, Proc. Computer Vision and Pattern Recognition (CVPR-03), Wisconsin, June 2003.

[3] P. Debevec, J. Malik, “Recovering high dynamic range radiance maps from photographs”, Computer Graphics, Proc. SIGGRAPH’97, pp. 369–378, 1997.

[4] M. Grossberg and S. Nayar, ”What can be Known about the Radiometric Response Function from Images ?”, Proc. ECCV’02. Vol. 4, pp. 189-205, 2002.

[5] M. Grossberg and S. Nayar, ”What is the Space of Camera Response Functions?”, Proc. Computer Vision and Pattern Recognition (CVPR-03), Wisconsin, June 2003.

[6] M. Grossberg and S. Nayar, ”High Dynamic Range from Multiple Images: Which Exposures to Combine?”, Proc. ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV), Nice, France, October 2003.

[7] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski ”High Dynamic Range Video”, Proc. ACM SIGGRAPH ’03, 2003

[8] S. Mann. “Comparametric equations with practical applications in quantigraphic image processing”, IEEE Trans. Image Proc., 9(8):1389–1406, August 2000.

[9] S. Mann and R. Picard, “On being ’undigital’ with digital cameras: Extending Dynamic Range by Combining Differently Exposed Pictures”, Proc. IS&T 46th annual conference, pp. 422-428, May 1995.

[10] T. Mitsunaga and S. Nayar, “Radiometric Self-Calibration”, Proc. CVPR’99. Vol.2, pp. 374–380, 1999.

[11] Y. Tsin, V. Ramesh, T. Kanade, “Statistical calibration of the ccd imaging process”, Proc. ICCV’01, Vol. 1, pp. 480–487, 2001.

Figure 8: (First Row) Few sample images of the sequence, (Second) Images aligned to the brightest image, (Third Row) Images aligned to the darkest image. *Reviewers, please refer to the supplementary video provided.

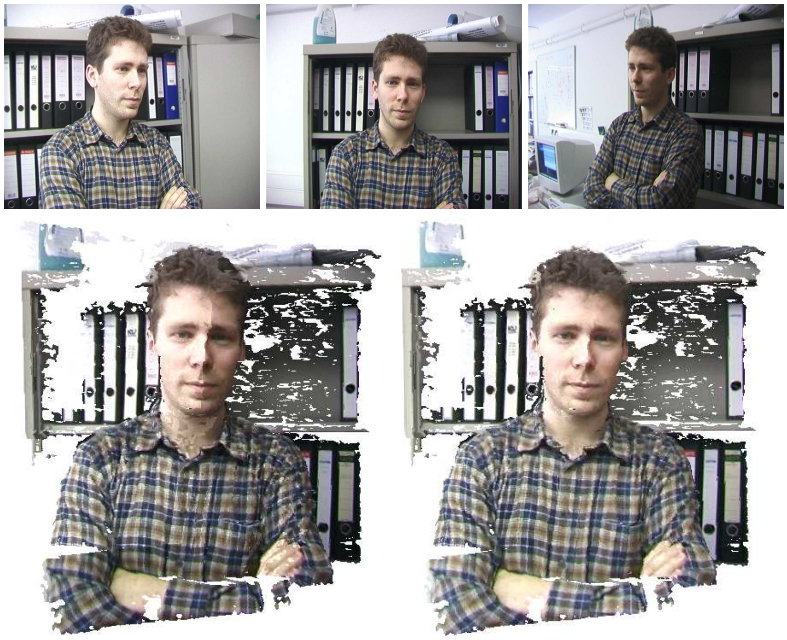

Figure 9: (Top) Few sample images from the sequence aligned to the (Bottom) 3D model with texture before(left) and after(right) normalization