High Dynamic Range Tone Mapping Based On Per-Pixel Exposure Mapping

Jason Huang
Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario
Email: jason.huang@mail.utoronto.ca

Valmiki Rampersad
Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario
Email: valmiki.rampersad@mail.utoronto.ca

Steve Mann
Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario
Email: hydraulophone@gmail.com

***Abstract*—The needs of a realtime vision aid require that it functions immediately, not merely for production of a picture or video record to be viewed later. Therefore the vision system must offer a high dynamic range (often hundreds of millions to one) that functions in real time. In compliment with the existing efficient and real-time HDR compositing algorithms, we propose a novel method for compressing High Dynamic Range (HDR) images by Per-Pixel Exposure Mapping (PPEM). Unlike any existing methods, PPEM only varies exposure to achieve tone mapping. It takes advantage of the camera response to enable exposure synthesis which we call Wyckoff set expansion. The method evaluates the synthetic exposures in a recursive pairwise process to generate a tone mapped HDR image. The results can be approximated using a look-up table, which can be used for real-time HDR applications.**

I. Wearable Cameras as a Sousveillance Device

A more recently coined word is the word "sousveillance", which is an etymologically correct opposite formed by replacing the prefix "sur", in "surveillance", with its opposite, "sous" [2], [3], [4], [5]. (See last 3 entries of Table I.) Consider the following English words and their French counterparts, with particular emphasis on the **difference between veillance and surveillance**:

Whereas surveillance cameras (cameras on property such as land or buildings) enjoy the benefit of optimal well-thought-out placement in relation to light sources in the environment (and many surveillance cameras even incoroprate their own light sources), sousveillance cameras (cameras on people) cannot be planned to such a degree.

In contrast to surveillance (cameras affixed to property), sousveillance (cameras borne by people) is a relatively new and and technologically challenging practice.

There are two kinds of sousveillance cameras: hand-held, and wearable [2], [4], [5], [3], [6], [7]. Of these, the wearable camera is the most challenging, technologically. Hand-held cameras are usually used intermittently, to photograph particular subject matter of interest. And the photographer has the advantage of being able to move around and consciously position the camera for best picture (e.g. orienting the camera so that the sun is behind the camera and not shining directly into it, for example).

TABLE I: The Veillances (SurVeillance and SousVeillance)

English	French
to see	voir
to look (at)	regarder
to watch	**veiller**
watching (monitoring)	**veillance**
watching over (oversight)	**surveillance**
to oversee (to watch from above)	**surveiller**
over (from above)	**sur**
under (from below)	**sous**
"undersight" (to watch from below)	**sousveillance**

Wearable computing, AR (Augmented or Augmediated Reality), the personal seeing aid (digital eyeglass), and the VMP (visual memory prosthetic) have emerged as new uses for cameras and computer vision [1]. Wearable cameras and related technologies give rise to sousveillance that is quite different from surveillance. They are often used continuously, in such applications as the Digital Eye Glass (computational seeing aid), and the Visual Memory Prosthetic [8], which can, for example, be used by a blind person to record their entire day and then have a sighted person explain their lives to them using *described memories for the visually impaired*.

A. Motivation

Unlike the recording tasks that require well-thought-out placement or post recording editing, a digital eye glass or VMP must work under all lighting conditions, even when looking directly into the sun, as might happen in day-to-day life, especially if worn by a non-sighted person unable to aim the camera optimally. This is critical for tasks including object or face recognition with the lack of lightning control. The most common difficulty in day-to-day recording of imagery, as well as automated object recognition, is the dynamic range limitations of the most cameras. Ultra miniature cameras suitable for being worn usually lack the dymamic range of their larger counterparts used in surveillance, or used for hand-held photography. Therefore the single most pressing need for sousveillance technologies is a means of managing dynamic range.

The needs of a realtime vision aid require that it functions immediately, not merely for production of a picture or video

978-1-4799-0929-2/13/$31.00

record to be viewed later. Therefore the vision system must offer a high dynamic range (HDR) (often hundreds of millions to one) that functions in real time. In addition, the processing requires preservation of details captured in the extended dynamic range but not compromising the users in critical situations by having them constantly adjusting many system settings.

A realtime seeing aid in extended dynamic range with simplified control thus becomes the key to improve the sousveillance system over the standard use of wearable cameras. Efficient and real-time HDR compositing methods have been developed by [9], [10]. However, the composited HDR images have a wider range than the conventional media, such as monitors, projectors or in print. The HDR images therefore cannot be displayed without being compressed or using specially designed HDR media [11]. The process of 'compressing' a HDR image to view the full range on Low Dynamic Range (LDR) media is commonly called spatial-tonal mapping or tone mapping. There exist simple tone mapping operators, such as the simple log operator, that is fast to apply to an HDR image, but produces an image which is low in contrast. There are better tone mapping operators that produce full range high contrast images from HDR images [12], [13], [14], [15], but may not be fast enough for real-time HDR tone mapping. In addition, their approach is oriented towards post-production use (such as image editing), which usually presents several settings for quality adjustments, not necessarily suitable for realtime seeing aid systems. Also, there are pixel blending techniques [16] that produce tone mapped outputs, however such techniques, do not correctly account for how a camera responds to varying amounts of light. In this work we approach the problem of tone mapping in a novel way, which accounts for the camera's response to light and is capable of running in real-time.

II. Exposure Synthesis As HDR Tone Mapping

In this section, we first briefly review the key concepts of the HDR compositing based on the works from Mann and Ali [17], [10], then we discuss how their work lead to our proposed tone mapping algorithm.

A. Wyckoff Set For Images

The term is first introduced by Mann [17] which describes the method for the digital image capturing devices to capture a scene with extended dynamic range, which is synonymous to the extended response film invented by Charles Wyckoff [18]. To extend the dynamic range, we can vary the exposure setting of camera for the image sensor to capture different ranges of photo-quantigraphic levels. By accumulating the details captured by each image of specific exposure setting we begin to encapsulate a more complete visual detail of the scene (even more than what human eyes are capable of). An example of Wyckoff set is shown in Fig. 2.

Two known techniques to capture the Wyckoff set are:

1) using one device taking multiple images in sequence

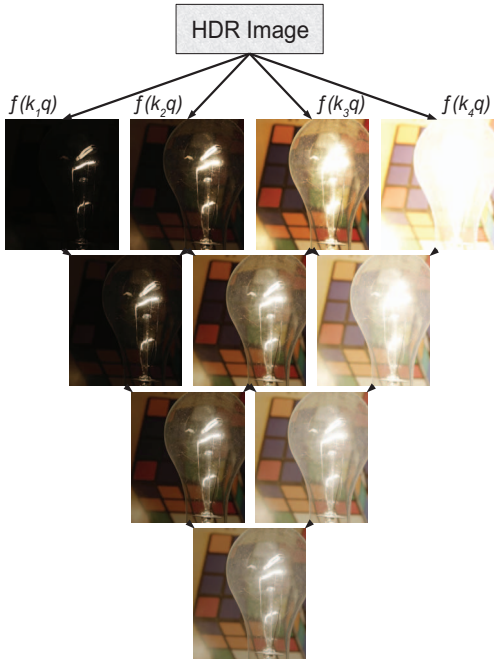

Fig. 1: This figure shows an example of the complete tree for the PPEM algorithm operating on a generated Wyckoff set of 4 images. The last image in the tree is the tone mapped HDR image.

2) using multiple devices taking a single image from different perspectives simultaneously.

In our work, we capture the scene with multiple images in sequence to expand the dynamic range. For example, we collect a Wyckoff set of five images with the exposure increasing to allow double amount of photo-quantigraphic level sensed by the sensor than the previous. The advantage of sequential capture is to reduce the number of capture devices required and thus demanding a lower bandwidth on image processing and transmission overall. It is also more configurable on how great the dynamic range by varying the number of image captures or the change in exposure amount. The disadvantage with sequential capture is the issue with objects in motion in the scene. They are considered as temporal noise because we assume the scene is static. For clarification, the new tone mapping to be introduced in this work neglects the problem associated with the noises (temporal noise due to motion in a Wyckoff set) as the alignment of the objects displaced over time should be dealt as a separate problem.

B. Prior Work: Comparametric HDR Compositing

Denote f and f^{-1} as the camera response and the inverse camera response respectively. A pixel, p, is the camera response of quantigraphic measure, q, at a selected exposure level, k_i, can be expressed as:

$$p = f(k_i q) \tag{1}$$

Inversely, the estimate $\hat{q}_i$ observed based on p and k_i:

$$\hat{q}_i = f^{-1}(p) k_i^{-1} \tag{2}$$

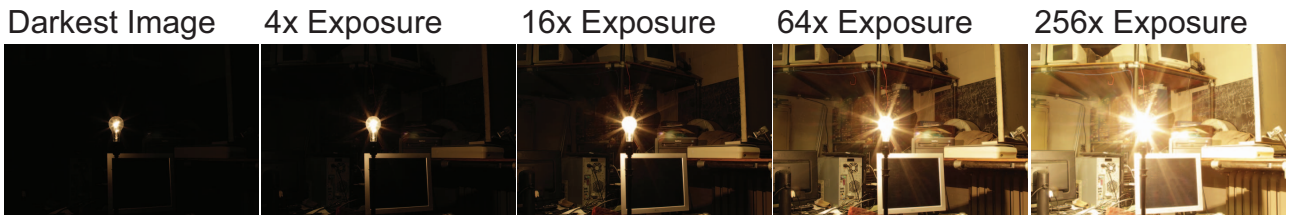

Fig. 2: Wyckoff set of five images. The set was captured with the exposure increasing by four times after each image (from left to right).

where the hat notation denotes the 'estimate' of the associated variable. Thus the final (or joint) estimate $\hat{q}$ based on a set of q_i with different exposures:

$$\hat{q} = f_{joint}(\hat{q}_1, .., \hat{q}_i, .., \hat{q}_N) \tag{3}$$

There are various methods to compute $\hat{q}$ based on different models of f_{joint} proposed by [10], [17], [19], [20]. Here, we have only applied the weighted sum method:

$$\hat{q} = \frac{\sum w_i \cdot \hat{q}_i}{\sum w_i} \tag{4}$$

based on [17] on a Wyckoff set [18] of size N, for $N \geq 1$.

C. Comparagram and Noise Model

A comparagram is a cross-histogram of two differently exposed images [9]. In a noise free world, the incident light on the camera sensor can be captured as continuous signals with infinite precision. Therefore, the comparagram of two differently exposed images is assumed to be a monotonic curve (comparagraph). However in reality, due to noises, there exists inconsistency in the correspondence of p_i to p_j or vice versa. Therefore, the mapping between pixels of different exposures on the comparagram spread around the comparagraph. The observed pixel values no longer have a one-to-one correspondence between two differently exposed images of the same subject matter, affected by quantization noise and other types of noise [9], [21]. For example, at a lower exposure setting, the details of lighter objects (light bulb, highly reflective materials or outdoor environment in a sunny day) are revealed; whereas at a higher exposure setting the darker objects (shadow, texture of darker colour materials, or night scene) become visible. The thickness (spread) of the curve can be interpreted as the standard deviation of the statistical distribution of pixel value mapping from one exposure to another [10]. The model of the pixel output from camera was summarized in [9]:

$$p_i = f(k_i q + n_{q_i}) + n_{p_i} \tag{5}$$

is affected by the quantigraphic noises n_{q_i} and camera noise n_{p_i}. The observation difference summarized in [10] is modelled as a random variable:

$$X_i = p_i - f(k_i q) \tag{6}$$

where X_i is a zero mean Normal distribution. Its standard deviation σ_{X_i} is approximated by the Inter-Quartile-Range (IQR) of the 'thickness' of the comparagraph per pixel value [10].

Originally, Ali and Mann [10] proposed a such noise model to more accurately estimate q by taking the spread or the 'thickness' observed on the comparagram into account. This leads to the idea of observing the relative uncertainty between the two pixels captured at different exposures. In this work, we see this measured uncertainty as an indicator to how poorly the camera was configured in exposure setting. Then we argue that the problem of tone mapping can be seen as an exposure value selection process per $\hat{q}$ based on the observed uncertainty. This results our proposed tone mapping algorithm that aims to scale $\hat{q}$ with a synthetic exposure value then compresses the range using the original camera response of the device for low dynamic range display media.

III. Proposed Exposure Synthesis For HDR Tone Mapping

With an estimate of $\hat{q}$ and a known camera response f, one can approximate the camera output [22]:

$$p_i \approx \hat{p}_i = f(k_i \hat{q}) \tag{7}$$

One use of such approximation is to reconstruct the captured Wyckoff set. This is achieved by setting $k_i = 2^{EV_i}$ to match the exposure value per captured image. The variable EV_i denotes the exposure value, commonly referred as 'stop' value [23]. Another use is to synthesize images (exposure synthesis) that were taken at an arbitrary EV within the maximum and minimum exosure values of the Wyckoff set. This allows the users to fine tune the exposure setting. However, regardless of the choice of EV, by applying a constant EV to all $\hat{q}$ in an image, the output still suffers from detail loss on the extremely bright and dark regions of the scene as both appear together in the same view. This can be resolved if the exposure value is synthesized per pixel, based on the amount of incident light. We term this method Per-Pixel-Exposure-Mapping (PPEM). The algorithm ultimately generates a 2-dimensional matrix of the same size as the HDR image. Each element $\hat{k}(x, y)$ is a scalar that modifies $\hat{q}(x, y)$ at coordinate x and y of the corresponding HDR image. Per pixel, the tone mapping is achieved by:

$$p_{mapped} = f(\hat{k}\hat{q}). \tag{8}$$

where the coordinates are omitted, since the algorithm works on each pixel independently.

A. Recursive Pairwise Per-Pixel-Exposure-Mapping

Denote i and $j \in \mathbb{N}$ as the indices of exposure values in the Wyckoff set, where $1 \leq i < j \leq \mathrm{N}$ and assume $k_i < k_j$. Consider p_i and p_j in a Wyckoff set, with a difference in exposure values:

$$\Delta EV_{i,j} = \log_2(k_i) - \log_2(k_j) \tag{9}$$

where $\log_2(k_i) = EV_i$. We wish to obtain a new exposure value $k_{i,j}$ such that $f(k_{i,j}\hat{q})$ avoids clipping at the extreme values. In particular, we find $k_{i,j}$ as the exponential of the weighted sum of a pair of EV_i and EV_j. Denote the weighted sum function as W:

$$EV_{i,j} = W(EV_i, EV_j) = \frac{C_i EV_i + C_j EV_j}{C_i + C_j} \tag{10}$$

where C_i is the certainty of EV_i. For Wyckoff set of size N=3:

$$EV_{1,2,3} = W(EV_{1,2}, EV_{2,3}) \tag{11}$$

Generalizing for N≥3 becomes a recursion based on Eq. 10:

$$EV_{1..N} = W(EV_{1..N-1}, EV_{2..N}) \tag{12}$$

Generalized for N inputs, $2^{EV_{1..N}}$ is the equivalent of $\hat{k}$ discussed previously. The weighted sum is a method to under value the 'weaker' candidates in a sample set. However, the weighted sum method may yield poor results in cases where the difference of EV_i and EV_j is large and they score low exposure certainty values. This may result a jump in the exposure map and produce spatial artifacts (See Fig.3). To resolve this problem, we reduce ΔEV between the pair (p_i, p_j) by Wyckoff set expansion, to be discussed in Section IV-B.

B. Desgining Exposure Certainty Function

The exposure certainty is designed based on the distribution of the pixel correspondence (thickness of the comparagraph) between two differently exposed images, as discussed in Section II-B. Let us first gain some intuition regarding the design of the certainty function. Consider a pixel $p_i = f(k_i q)$. Assuming q is fixed, one can vary k to simulate the change of exposure. However, due to the discontinuity of camera response f that is unable to recover pixel values greater than p_{max} or pixel values lower than p_{min}, this implies that there is also a finite range of k_i such that a change of value in k_i is effective in changing p_i given q. Beyond the clamping points, one can only obtain the maximum estimate $f^{-1}(p_{max}) = \hat{q}_{max}$ for k_i too high or the minimum estimate $f^{-1}(p_{min}) = \hat{q}_{min}$ for k_i too low. This also censors the original k_i value if we wish to estimate the value with only the observed p_{max} or p_{min}. Specifically, the problem of censored k_i value becomes obvious by taking the ratio of estimate $(f^{-1}(p_i))$ over the supposed quantity exposed to the sensor $(k_i q)$:

$$Y_i = \frac{f^{-1}(p_i)}{k_i q}, \quad Y_i \in [0, \infty) \tag{13}$$

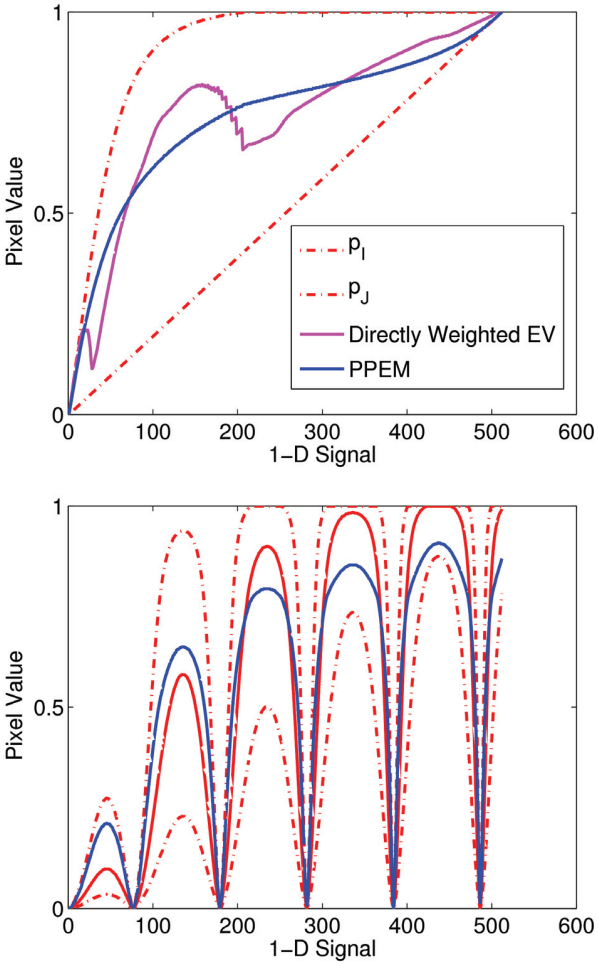

Fig. 3: The figure demonstrates the response of the tone mapping algorithm on a 1-D $\hat{q}$ signal. The signals $p_i = f(2^{-1.5}\hat{q})$ and $p_j = f(2^{1.5}\hat{q})$ are a Wyckoff pair of 3 ΔEV. From the top figure, the method of weighted sum without Wyckoff set expansion produces multiple abrupt transitions on a supposed smooth gradient. This is resolved with Wyckoff set expansion and compute the results according to Eq.12. The bottom figure demonstrates the tone mapped result on a raised sinusoid function of $\hat{q}$ that is amplitude modulated by an exponentially increasing envelope. The tone mapping results a curve to follow p_j closely when the overall Wyckoff set produce low pixel values; similarly, the curve sticks closely to p_i when the overal Wyckoff set are clipping at the maximum pixel value.

which only results a constant value of 1 if no noise or clamping is present in the device when capturing the Wyckoff set. By applying the $\log_2$ operator on Y_i, we have:

$$Z_i = \log_2(Y_i) = \log_2(f^{-1}(p_i)) - \log_2(k_i q) \tag{14}$$

For the moment, if we assume that q can be perfectly recovered and only the estimate of k_i is affected by noises and finite dynamic range then:

$$Y_i = \frac{\hat{k}_i q}{k_i q} = \frac{\hat{k}_i}{k_i} \tag{15}$$

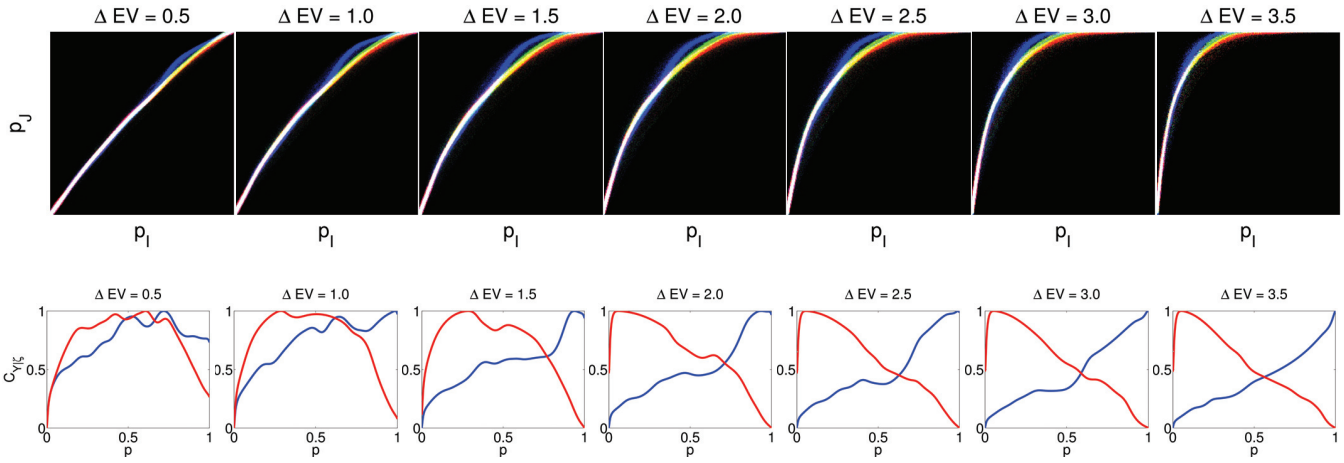

Fig. 4: An example of comparagrams (top row) and the associated certainty fucntion pairs (bottom row). The comparagrams were constructed based on the average of ten Wyckoff sets; each set captured 8 images at an incremental 0.5 ΔEV. This sampling method is discussed in [22]. The certainty functions were determined according to Section II. The certainty functions of the lower exposure input ($C_{Y_i|\zeta}$) and the higher exposure input ($C_{Y_j|\zeta}$) are in blue and red respectively. At a lower ΔEV, the certainty function pairs are similar across the mid-range pixel values but exhibiting contrast at the extreme values. The contrast of certainties becomes obvious from the comparagrams of higher ΔEV. For example, at 3.5 ΔEV the pair form an 'X' shape, indicating inputs of lower exposure score higher certainty towards brighter pixels than inputs of higher exposure. In our setup, certainty functions calibrated from comparagrams of ΔEV = $\{0.5, 1.0, 1.5, 2.0\}$ are used. The rest are discarded due to noise issues, discussed in Section IV-B.

and

$$Z_i = \log_2(\hat{k}_i) - \log_2(k_i) \tag{16}$$

By looking at Z_i alone, it informs about the spread of estimated k_i in EV scale. Which provides a clue on how this exposure value increases the dynamic range for a pixel. For example, if a pixel $p_i = p_{max}$ is sampled with EV_i, then setting the camera to a higher exposure value $EV_j > EV_i$ at best captures $p_j = p_{max}$. Assuming that no other noise is present and having p_i as the reference, the sampled Z_j would result $\Delta EV_{i,j}$ which infers no increase in dynamic range. Clearly, this is an example of Wyckoff set of two pixels and both are saturated at p_{max}. For the clamping case (the worst case), Z_j indicates the loss of dynamic range in observation against the actual exposure settings.

Such provides some intuition with quantifiable observation regarding the effect of bad exposure setting (e.g. the cause of value clamping at the maximum or the minimum of the sensor reading). Note that this is not an intention of observing the inconsistency in camera shutter or the image noise due to camera ISO setting. Rather, this is an effort to have one variable that summarizes over all kinds of noises present in the image captures that also includes the effect of poor exposure setting problem. Near the clamping region of the camera reading, we no longer observe the correct ΔEV from the two pixels of different exposures that were configured with a $\Delta EV > 0$. Instead, at those extreme ends, the ΔEV we observe is approximately zero.

The resulting observational difference Z_i is greater than zero for Y_i is greater than one. We see the connection between Z_i and X_i as $Z_i$can be expressed a function g of X_i from Eq.6:

$$Z_i = g(X_i) = \log_2(f^{-1}(X_i + f(k_i q))) - \log_2(k_i q) \tag{17}$$

An interesting observation here is that for $f = \log_2$, $Z_i = X_i$. However, in general, the camera response is device specific. It can be generally modelled by a family of exponential or logarithmic based functions [9], not necessarily a pure logarithm.

The significance of bridging Z_i and X_i is to show their similarity in the uncertainty observation between pixels of different exposures that could be applied in either HDR compositing or assessment on the quality of exposure setting. Unlike Ali and Mann's work which solves an optimization problem to maximize the probability of the likelihood of $\hat{q}$, the inverse problem "compression of $\hat{q}$ for display" cannot be solved in the same way. Instead, we rely on the heuristics of attaching a virtual exposure value that shifts the $\hat{q}$ for compression. Specifically, we weight the exposure value of each pixel in the Wyckoff set by a function of exposure uncertainty. In particular, the uncertainty of the exposure value is computed based on the variance observed in Y_i.

Empirically, Y_i can be approximated as a Skew-Normal[24] distribution. The skew becomes more obvious towards the clamping values which we may refer to as a piling effect (the pixel values are clustered around the maximum and the minimum sensor readings when both images are nearly over or under exposed). However, even with the Skew-Normal fitting, the piling effect at the maximum and minimum values require a separate estimate on top of a single Normal-like distribution. In Ali and Mann's work [10], the IQR is applied to discard the sample noises (including ones that fall in the piling

effect region) and approximate the X_i as a standard Nor distribution. Therefore, as a workaround, we instead treat as random variables with standard Normal distribution. W random variable Y_i, we approximate its entropy [25] refer to the Shannon's entropy in the study of informat theory) as the uncertainty of the exposure value. This strat is synonymous to diminishing the significance of the expos value by the severity of the observed spread. Assuming th $\hat{q}_i$ estimated from a single pixel has a certain dynamic ran

$$\text{dynamic range of } \hat{q}_i = \log_2(\frac{q_{max}}{q_{min}}) \quad ($$

By capturing an additional pixel with half of exposure, obtain $\hat{q}_j$ that is at most two times the $\hat{q}_i$. After the H compositing, the expected increase in dynamic range of pixel is at most but not necessarily by one bit. The nc that reduces the effective increase in dynamic range be summarized by its entropy. The entropy quantifies s uncertainty in bits which matches the scale of the expos value. According to [26], based on Shannon's definition entropy:

$$H^{SN}(Y_i) \leq H^N(Y_i) = -\log_2(2\pi\sigma^2_{Y_i}) \quad ($$

where $H^{SN}(Y_i)$ denotes the entropy of Y_i and $H^N($ denotes the entropy for Y_i modelled as a standard normal, N $(\mu_{Y_i}, \sigma_{Y_i})$. The inequality by Eq.19 provides an upperbo measure of entropy of Y_i. Further, $\sigma^2_{Y_i}$ can be approxima based on $\sigma^2_{X_i}$, by using Delta's Method [27]:

$$\sigma^2_{Y_i} \approx \left(\frac{\delta g(E[X_i])}{\delta X_i}\right)^2 \sigma^2_{X_i} \quad ($$

where σ_{X_i} is approximated using IQR [10]. The estima statistical properties are extracted from the comparagrams, moments of Y_i and X_i are therefore conditional to $\Delta EV_{i,j}$ q. For clarification, we only consider EV_i as a relative quan to EV_j instead of a relative quantity to the EV sample sp of the Wyckoff set. For consistency, between a pair $(p_i,$ and assuming the relation $EV_i < EV_j$, we enforce EV_i $-\frac{1}{2}\Delta EV_{i,j}$ and $EV_j = +\frac{1}{2}\Delta EV_{i,j}$ throughout the remain of the paper. For a more concise use of symbols, we denote vector $\zeta = [q, \Delta EV_{i,j}]$ to simplify the conditional variab We now define the certainty function to be the inverse of entropy upperbound:

$$C_{Y_i|\zeta} \equiv H^N(Y_i|\zeta)^{-1} \quad ($$

for $H^N(Y_i|\zeta) \in \mathbb{R}^+$ and $C_{Y_i|\zeta} \in [0, 1]$ where linear norm ization step is applied on $C_{Y_i|\zeta}$. This unbiases the maxim and minimum values possible between $C_{Y_i|\zeta}$ and $C_{Y_j|\zeta}$.

IV. Implementation

A. Calibration Of Certainty Function

$C_{Y_i|\zeta}$ is a function of p_i based on the approximation Eq.20. While f remains the same and $\hat{q}$ is held constant varies if and only if k_i changes. This occurs as $\text{EV}_{a..b-1}$ $\text{EV}_{a+1..b}$ are updated per iteration of Eq.10. The correspo ing ΔEV also deviates from its initial value. Consequer

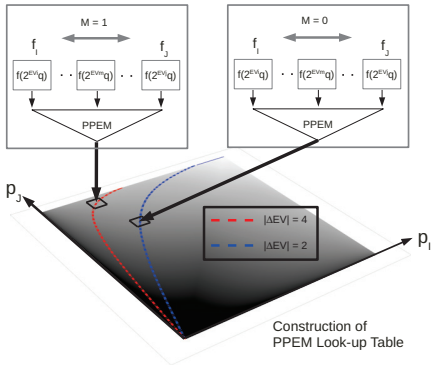

Fig. 5: Visualization of the look-up table (LUT) construction. The ΔEV of all pairs of (p_i, p_j) on the red and blue curves are 4 and 2 respectively. We emphasize the two curves here to show that the construction of LUT requires different amount of input syntheses for Wyckoff set expansion. The symbol M indicates the number of virtual pixels are inserted between p_i and p_j. In our example, we have certainty functions calibrated from comparagrams of $\Delta EV = \{0.5, 1.0, 1.5, 2.0\}$. Without known certainty functions for 4 ΔEV, the pixel pairs along the red curve requires an expansion. The expansion inserts one p_m (EV_i = -2, EV_m = 0, EV_j = +2) to initialize the input set. The pixel pairs of 2 ΔEV can refer to the calibrated certainty functions directly without the expansion.

$C_{Y_i|\zeta}$ needs to be re-computed and normalized accordingly. This increases the overal computing complexity as it requires new measures from the comparagam that matches the updated ΔEV. To avoid such redundancy, we devise a calibration strategy based on multiple comparagrams of different ΔEV. Consider a Wyckoff set of size N, captured with a constant incremental ΔEV. We obtain up to $N-1$ pairs of certainty functions based on comparagrams that were constructed with image pairs of $\Delta EV, 2\Delta EV, ..., (N-1)\Delta EV$. The certainty values of an arbitrary ΔEV is interpolated based on the known certainty functions. Alternatively, these can refer to the certainty function calibrated to the nearest ΔEV.

B. Wyckoff Set Expansion

Given $\hat{q}$, we can generate p_i with some arbitrary k_i. This enables exposure synthesis and generation of virtual pixel pairs that differ by an arbitrary ΔEV. Between a pair of pixel values p_i and p_j, we can insert M number of virtual pixels, $f(2^{EV_m}\hat{q})$, such that:

$$M = \lceil \frac{|\Delta EV_{i,j}|}{\Delta EV} \rceil - 1, \quad EV_i < EV_m < EV_j \qquad (22)$$

This is termed the Wyckoff set expansion. It leads to reduction of the total number of comparagrams needed for calibration stage. In practice, users may require a different exposure bracketing for viewing adjustments. For example, if $EV_{max} - EV_{min}$ of the new setting is greater than the largest ΔEV used for certainty calibration, the expansion inserts M extra pixel values such that:

$$\Delta EV_M = \frac{EV_{max} - EV_{min}}{M} \qquad (23)$$

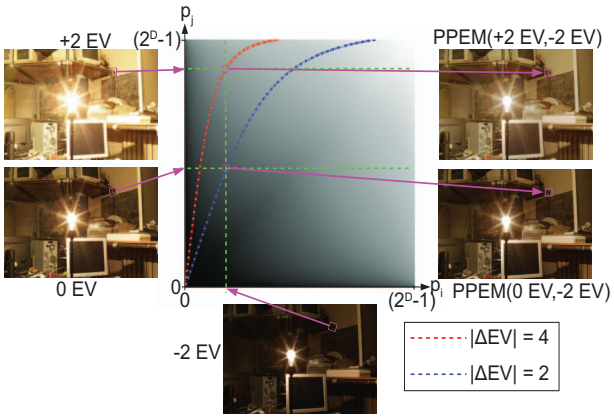

Fig. 6: Example of look-up table (LUT) use. The ΔEV of all pairs of (p_i, p_j) on the red and blue curves are 4 and 2 respectively. We show that the LUT can be used for image pairs that have different $\Delta EV's$. The axes of the LUT correspond to indices of the input pair of images. The horizonal-axis is for the lower exposed image of the pair and the vertical-axis is for the higher exposed image of the pair. The output of LUT is a pixel value which can be converted directly to a LUT index for the proceeding iterations of PPEM.

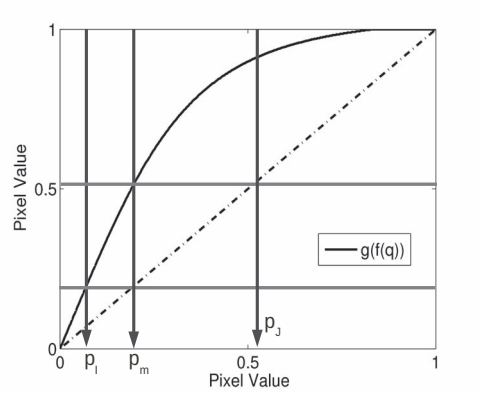

Fig. 7: An example of comparagraph $g(f(q))$. Consider p_i and p_j are the darkest and the brightest camera outputs in the associated Wyckoff set for tone mapping. Further, the process expands the set by inserting an extra pixel p_m. We can 'trace' the comparagraph to obtain p_i, p_m and p_j iteratively. This process can be accelerated with a LUT of 2^D elements to store the discretized $g(f(q))$ values. Given $\lfloor p_m(2^D-1)+0.5 \rfloor$ as an input index, the LUT returns the output in index form $\lfloor p_{m+1}(2^D-1)+0.5 \rfloor$ for the proceeding look-ups.

The expansion forces pixel pairs to differ at the lower ΔEV. This avoids tone mapping a pair of pixels in high contrast (very dark p_i and very bright p_j). Without the expansion, a smooth gradient region of image that contains high contrast pixel pairs may induce colour bands, which is a type of spatial artifacts. Empirically, with our setting, the artifacts appear for any pixel pair of greater than 2 ΔEV. Therefore, certainty functions calibrated for 2.5, 3.0 and 3.5 ΔEV are discarded.

C. Accelerating PPEM With Look-up Tables

Consider the original Wyckoff set of size N. The recursion (Eq.12) has an $O(N^2)$ runtime complexity per pixel. With an expanded Wyckoff set, the runtime increases non-linearly to $O\left((MN)^2\right)$. To reduce the overall computation cost, we resort to a pre-computed PPEM look-up table (LUT) to accelerate the process. Denote D as bit depth, we construct a 2-dimensional LUT of 2^D entries each dimension. The LUT takes an input pair of pixel values p_i and p_j each quantized to one of the 2^D possible values in the range $[0, 1]$. For each combination of input pair, the LUT stores $f(2^{EV_{i..j}}\hat{q})$ based on the PPEM algorithm described in Section III, incorporating the expansion from IV-B. The precision of the result depends on the choice of D. For a larger D, the LUT covers more combinations of input pairs and consequently includes results from a greater range of $\Delta EV_{i,j}$. The LUT-based approach allows a flexible implementation of the algorithm to meet constraints on computation time and memory usage. The computation complexity can be reduced to $O(1)$ when performing look-up on the darkest and the brightest samples (p_1, p_N) of $\Delta EV_{1,N}$.

D. Alternative Look-up Scheme

An $O(1)$ direct look-up in a case where $\Delta EV_{1,N} \gg D$ may lead to undesireable results. Alternatively, we can reduce $\Delta EV_{1,N}$ of the pixel pairs using the concept of Wyckoff set expansion. This generates S number of virtual pixels as the input set, where $S \geq 2$ and

$$\Delta EV_s = \frac{\Delta EV_{1,N}}{S-1} \tag{24}$$

A recursive look-up is applied on the virtual input set, identical to the methodology described in Section III-A. For example, with an original Wyckoff set of $|\Delta EV_{1,N}|$ = 10 and choose S = 3, we first generate virtual pixels p_1, p_2 and p_3 at -5EV, 0EV, and +5EV respectively. In the first iteration, PPEM look-ups are performed on the pairs (p_1, p_2) and (p_2, p_3) independently. In the second iteration, PPEM look-up is performed on the previously obtained $f(2^{EV_{1,2}}\hat{q})$ and $f(2^{EV_{2,3}}\hat{q})$ and result $f(2^{EV_{1,2,3}}\hat{q})$ as the final output.

Depending on the model of the camera response, the calculations for the initial Wyckoff set expansion can be expensive. This is especially inefficient for high resolution HDR images. However, this can be accelerated using pre-computed LUT for approximating each p_m of constant incrementing exposures. We exploit the technique used for unrolling the comparagraph introduced by [9]. Given a comparagraph g we can quickly generate the initial Wyckoff set by computing recursively $g(p_m) = p_{m+1}$. This is true because

$$g(f(\hat{q})) = f(2^{\Delta EV_M}\hat{q}) \tag{25}$$

Since g is a function of ΔEV, we need to re-compute a new one-dimensional LUT of comparagraph when S changes (refer to Eq.24). The new LUT of 2^D elements stores the discretized:

$$\hat{p} = \lfloor g(f(\hat{q}))(2^D-1)+0.5 \rfloor \tag{26}$$

V. Results

The proposed PPEM algorithm has been implemented in Matlab, C and CUDA programming languages to generate the

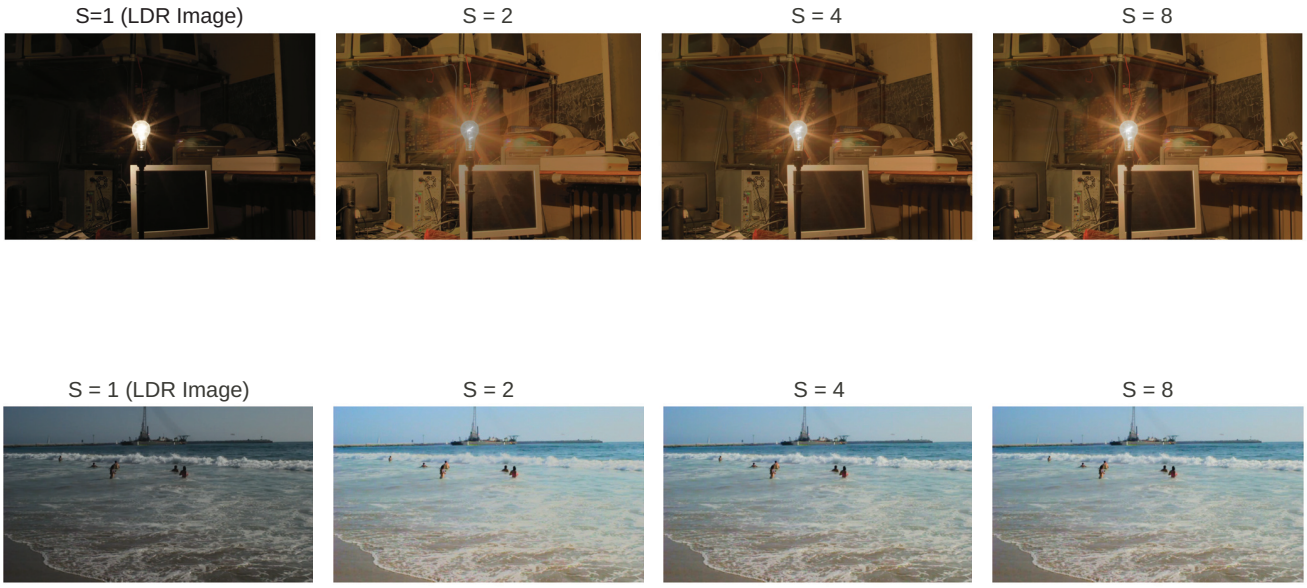

Fig. 9: Output samples based on based on a set of 3 source images taken at the beach in a sunny day. The images were taken at an incremental exposure value of ΔEV = 4. Visible tonal distortion (whitened ocean colour) appears in the case of S = 2 but not for $S \geq 4$.

results and performance benchmarks in this paper. The camera response, certainty function and comparagrams were calibrated for Cannon 60D. The comparagrams were constructed based on images of $\Delta EV = \{0.5, 1, 1.5, 2.0\}$, averaged from ten different Wyckoff sets of eight images. A LUT with a dimension of 512x512 for each of RGB color channels is constructed by the methods described in Section III and IV-A.

In Fig.8 and Fig.9, we show examples of tone mapped results with different S values. For S=1, the exposure map remains constant across the image, which is equivalent to an LDR image taken by a virtual camera. The quality of tone mapped results improve as S increases; the improvements for $S \geq 3$ become less visible. The performance of the algorithm is measured in frames per seconds (FPS), shown in TableII. The algorithm is suitable for parallel processing due to per-pixel computations without information sharing between the neighbourhood pixels. For a mobile platform unit, the algorithm is able to achieve more than 60 FPS for real-time processing on 720p videos.

		S=2	S=4	S=6	S=8
Frames Per Second (FPS)	CPU	6.25	2.56	1.32	0.69
	GPU	164.2	71.8	34.3	16.1

TABLE II: Performance of PPEM using look-up table. The proposed PPEM algorithm ran on a laptop using Intel Core i7-2670QM CPU at 3.0 GHz with Nvidia GT 425M GPU. The benchmark were performed on video frames with 1280 x 720 resolution.

Qualitatively, we show that PPEM is able to deliver comparable results against the two well known tone mapping methods (see Fig.10). We choose Reinhard's [13] and Durand's [15] based on the tone mapping evaluation conducted by [28]. In the evaluation, Reinhard's method ranks top on the overall quality score whereas Durand's ranks nearly at the bottom. The light tube in the scene is of interest. The PPEM method is able to reveal the details just as well as Reinhard's, without additional spatial filtering on the edge and detail enhancements. Unlike, Reinhard or Durand's, the proposed PPEM method only requires the user to choose S as the size of the input Wyckoff set for tone mapping. The parameter S effectively increases the precision to react to the increase in dynamic range of the captured Wyckoff set, in trade-off with extra computation time.

VI. Conclusion

We have demonstrated a way to tone map an HDR image by keeping the camera response as the range compressor and varying only the exposure per pixel. Qualitatively, the tone mapped results are comparable to the other well known methods. This method can be accelerated to achieve real-time performance by using a multi-dimensional look-up table, which incorporates the Wyckoff set expansion. The look-up table approach leads to a flexible implementaion that is capable of handling different ΔEV. It also reduces the computation complexity to approach constant runtime per pixel. These are the advantages the proposed PPEM tone mapping attribute. In compliment with the efficient and real-time HDR compositing algorithms by [9], [10], we can achieve real-time tone mapped HDR on Wyckoff video sources. This enables the construction of real-time seeing aid with the extended dynamic range.

References

[1] S. Mann, "Wearable computing," in *Encyclopedia of Human-Computer Interaction*, M. Soegaard and R. F. Dam, Eds. The Interaction Design Foundation. Available online at http://www.interaction-design.org/encyclopedia/wearable_computing.html, 2012.

[2] S. Mann, J. Nolan, and B. Wellman, "Sousveillance: Inventing and using wearable computing devices for data collection in surveillance environments." *Surveillance & Society*, vol. 1, no. 3, pp. 331–355, 2002.

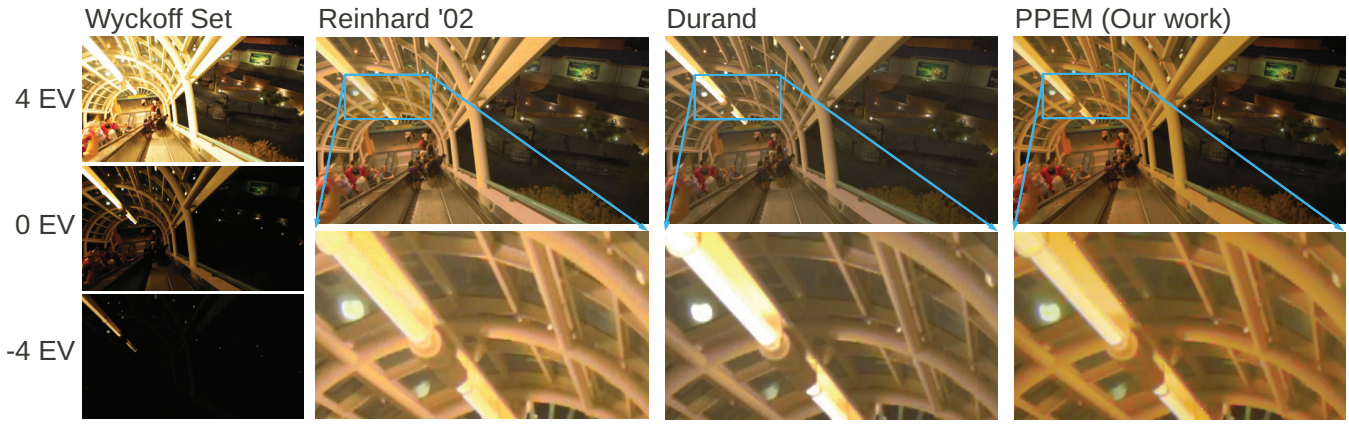

Fig. 10: Tone mapping comparisons on a Wyckoff set of three images $EV = \{-4, 0, 4\}$. The images are extracted from a video sequence recorded by the author. The light tube in the scene is of interest. The PPEM method is able to reveal the details just as well as Reinhard's, without additional spatial filtering on the edge and detail enhancements.

[3] G. Fletcher, M. Griffiths, and M. Kutar, "A day in the digital life: a preliminary sousveillance study," *SSRN, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1923629*, September 7, 2011.

[4] K. Michael and M. Michael, "Sousveillance and point of view technologies in law enforcement: An overview," 2012.

[5] J. Bradwell and K. Michael, "Security workshop brings' sousveillance'under the microscope," 2012.

[6] C. Reynolds, "Negative sousveillance," *First International Conference of the International Association for Computing and Philosophy (IACAP11)*, pp. 306 – 309, July 4 - 6, 2011, Aarhus, Denmark.

[7] V. Bakir, *Sousveillance, media and strategic political communication: Iraq, USA, UK*. Continuum, 2010.

[8] S. Mann, "**Wearable, tetherless computer–mediated reality**: WearCam as a wearable face–recognizer, and other applications for the disabled," M.I.T. Media Lab Perceptual Computing Section; Also appears in **AAAI Fall Symposium on Developing Assistive Technology for People with Disabilities**, 9-11 November 1996, MIT; http://wearcam.org/vmp.htm, Cambridge, Massachusetts, TR 361, February 2 1996.

[9] ——, "Comparametric Equations, Quantigraphic Image Processing, and Comparagraphic Rendering," *Intelligent Image Processing*, pp. 103–178, 2002.

[10] M. A. Ali and S. Mann, "Comparametric image compositing: Computationally efficient high dynamic range imaging," in *To appear, Proc. Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*. IEEE, March 2012.

[11] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3, pp. 760–768, 2004.

[12] R. Mantiuk, K. Myszkowski, and H. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Transactions on Applied Perception (TAP)*, vol. 3, no. 3, pp. 286–308, 2006.

[13] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 267–276, 2002.

[14] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 249–256, 2002.

[15] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM transactions on graphics*, vol. 21, no. 3, pp. 257–266, 2002.

[16] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on*. IEEE, 2007, pp. 382–390.

[17] S. Mann, "Compositing multiple pictures of the same scene," in *Proceedings of the 46th Annual IS&T Conference*, vol. 2, 1993.

[18] C. Wyckoff, "An experimental extended response film," *SPIE Newslett*, pp. 16–20, 1962.

[19] M. Robertson, S. Borman, and R. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging*, vol. 12, p. 219, 2003.

[20] C. Manders and S. Mann, "Digital camera sensor noise estimation from different illuminations of identical subject matter," in *Information, Communications and Signal Processing, 2005 Fifth International Conference on*. IEEE, 2005, pp. 1292–1296.

[21] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H. Seidel, and H. Lensch, "Optimal hdr reconstruction with linear digital cameras," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 215–222.

[22] S. Mann and R. Picard, "Beingundigital'with digital cameras," 1994.

[23] S. Ray, "Camera exposure determination," *The Manual of Photography: Photographic and Digital Imaging*, 2000.

[24] A. Azzalini and A. Dalla Valle, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, no. 4, pp. 715–726, 1996.

[25] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley-interscience, 2012.

[26] R. ARELLANO-VALLE, J. CONTRERAS-REYES, and M. GENTON, "Shannon entropy and mutual information for multivariate skew-elliptical distributions," 2011.

[27] G. Oehlert, "A note on the delta method," *The American Statistician*, vol. 46, no. 1, pp. 27–29, 1992.

[28] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Image attributes and quality for evaluation of tone mapping operators," in *National Taiwan University*. Citeseer, 2006.